www.sciencemag.org/cgi/content/full/336/6084/1030/DC1

Supplementary Materials for

Linking Crystallographic Model and Data Quality

P. Andrew Karplus and Kay Diederichs

*To whom correspondence should be addressed. E-mail: kay.diederichs@uni-konstanz.de

Published 24 May 2012, *Science* **336**, 1030 (2012)
DOI: 10.1126/science.1218231

This PDF file includes:

Materials and Methods

The EXP data set and its analyses

Wild type cysteine dioxygenase (CDO) was purified and crystallized as described previously (*9*). The crystals are of space group $P4_32_12$ with a=b=57.64 Å, c=122.41 Å. The data were collected from a single crystal soaked in 100 mM cysteine as before (*7*), with a variation being that the soak and flash freezing were done in an anaerobic chamber. The data set was collected at Beamline 5.0.1 at the Advanced Light Source (Lawrence Berkeley National Laboratory), and included 218 contiguous 1˚-frames. Data were processed with XDS (*23*, *24*)(version 12-Dec-2010), using default parameters. For obtaining the $CC_{1/2}$ values, a separate program HIRESCUT was written; we note that the output file of the program SCALA (*6*)includes the $CC_{1/2}$ statistic under the name CC_Imean. CC_{work} and CC_{free} were calculated using sftools (*23*). Data statistics for the ***EXP*** data set based on 1.8 Å (which would have been chosen based on current standard criteria) and 1.42 Å resolution cutoffs are presented in Table S1.

For refinement, a fully-automated protocol was carried out using Phenix.refine (*25*) (version 1.7.1) at each of the six resolution cutoffs tested. The protocol we used is validated in that its application to the PDB 3ELN data set yielded a model with R_{work}/R_{free}=0.148/0.172, having both lower R_{free} and less overfitting (i.e. a higher R_{work}) than the published refinement (*9*). All refinements began with the protein and solvent coordinates taken from the unliganded CDO structure (PDB entry 2B5H) with five active site waters (505, 621, 651, 668, and 758) replaced by the cysteine persulfenate atoms taken from PDB entry 3ELN. Hydrogen atom positions were constructed, and an isotropic refinement run was first carried out using Phenix.refine with identification and update of a solvent model. This resulting model is referred to as the isotropic refined model. This model was then submitted to anisotropic refinement with solvent update and real-space identification of the best-fitting sidechain rotamers. Anisotropic refinement at high-resolution cutoffs better than 2.0 Å resulted in values of R_{free} lower by 0.5% to 2%, compared to the isotropic refinement. For the 2 Å refinements, the isotropic refinement always gave the lowest R_{free} values. Table S2 gives the R_{work}/R_{free} values for the refinements carried out at the resolutions tested, as well as select values for R_{work}/R_{free} against data truncated at a lower resolution than that at which the refinement was carried out. Consistently, comparison of these R_{work}/R_{free} values shows improvement of R_{free} and reduction of the R_{free}-R_{work} gap for a model that was refined at higher resolution.

The simulated (SIM) data set and its analyses

Synthetic data frames were generated by the SIM_MX program (*20*), which simulates, using a set of input intensities, a diffraction experiment characterized by crystal and beam properties, geometry, background noise and counting statistics. The input intensities were calculated from the 3ELN model in spacegroup $P4_32_12$ (a=b=57.52Å, c=122.19Å), using anisotropic atoms, added hydrogen atoms and a solvent model; they provide for the simulated data set a perfect 'true' reference data set. The frames were processed using XDS (*23*, *24*), giving the ***SIM*** data set. Data statistics for the ***SIM*** data set based on 1.6 Å (which would have been chosen based on current standard criteria) and 1.42 Å resolution cutoffs are presented in Table S3. A second simulated data set – **SIMstrong** – was created using input intensities that were 15-fold larger. This provided a reference data set against which ***SIM*** could be compared that is analogous to comparing the ***EXP*** data with 3ELN.

Refinements of models against the ***SIM*** data were carried out exactly as were those against the ***EXP*** data, and the R_{work}/R_{free} values for the refinements are presented in Table S4. Figure S3 (panels A-E) provides for the ***SIM*** data set the results equivalent to what Figures 1 through 4 of the main text provide for the ***EXP*** data set.

One notable deviation of the results using the ***SIM*** data from those with the ***EXP*** data is that CC_{work} and CC_{free} are both much closer to CC* in the resolution range below about 2.5 Å. We believe this occurs because the intensities of the synthetic data set were computed from a single perfectly defined molecular model with flat bulk solvent, and so they can be very well fit by the single model used in the refinement. In contrast, the ***EXP*** data are derived from a real crystal which will include many features (such as unmodeled bulk solvent and complex molecular disorder involving backbone and side-chain alternative conformations and motions) not well accounted for by a single refined model.

Retrospective analyses of two published structures

To further confirm that the behaviours shown here for the EXP and SIM data sets are not related to any unique property of that crystal form, we selected two examples from the literature that (i) were recently published, (ii) were carried out by a highly experienced research group, (iii) had been analyzed at medium resolution, (iv) had the raw diffraction images publically available, and (v) had data extending beyond the published resolution cutoff that were relatively complete. Furthermore, to confirm that high redundancy is not required for the $CC_{1/2}$ statistic to be useful, we chose one of the examples to be a low symmetry space group having a data set with lower (3- to 4-fold) redundancy. The two data sets analyzed have different space groups (Table S5) and were simply the first two data sets we found that fulfilled the above criteria; both were reported in a single study (*27*). The raw diffraction images from the Center for Structural Genomics of Infectious Diseases archive (http://csgid.org/csgid/) were reprocessed using XDS. For controlled comparisons, we carried out refinements using a common protocol at the published resolution cutoff and at extended resolution cutoffs. For the refinements, all solvent molecules were removed from the PDB file, ligand CIF files were produced with Phenix.ready_set (*25*) and Phenix.refine was run with the options "ordered_solvent=true strategy=individual_sites+individual_adp+tls fix_rotamers=true." For PDB entry 3E4F, 16 TLS groups were defined by Phenix.find_tls_groups; for PDB entry 3N0S, one TLS group was used per chain. For each case, three resolution cutoffs were used: that from the original publication, and those having $CC_{1/2}$-values in the 0.4-0.5 and the 0.1-0.2 ranges. As reported in Table S5, in both cases the phenix refinement protocol is validated as it yields R-factors comparable to those published. Regarding the resolution extension, paired R-factor comparisons show that the models refined against data to the higher resolution limits are improved. Including the data out to near $CC_{1/2}$=0.1-0.2 lowers R_{free} and improves the R_{free}/R_{work} differential (R_{free}-R_{work}) at the published resolution cutoff by 0.25% for the 3E4F case and by 0.77% for the 3N0S case.

Options and commands for refinement and analysis

a) Phenix.ready_set and Phenix.elbow were used for adding hydrogens to the protein model, and to obtain CIF files for ligands.

b) For isotropic EXP or SIM refinement, Phenix.refine options were as follows:

"`ordered_solvent=true  xray_data.high_resolution=...`" .

For anisotropic EXP or SIM refinement, additional options were

"`ordered_solvent.new_solvent=anisotropic`
`adp.individual.anisotropic="not element H"  fix_rotamers=True`".

We did not deviate from other phenix.refine defaults. In particular, the number of phenix.refine macro cycles was not changed from its default of 3, and no specific optimization of weights was performed.

c) for obtaining R values at a lower resolution than that at which the model was refined, we used the phenix.refine options as follows:

"`main.number_of_macro_cycles=1  strategy=None  fix_rotamers=False`
`ordered_solvent=False  xray_data.high_resolution=...`".

d) sftools commands for calculating CC_{work}, CC_{free} at 1.42Å are as follows:

```
# first read mtz file with experimental data, this has a column "IOBS"
# second read mtz file written by phenix.refine; this has a column "F-model".
```

Then create F^2 values for the model:

```
calc col F-modelsq = col F-model col F-model *
```

Then calculate CC_{work}:

```
select only col R-free-flags > 0
correl col IOBS F-modelsq SHELLS 14 RESOLUTION 999 1.42
```

Then calculate CC_{free}:

```
select only col R-free-flags = 0
correl col IOBS F-modelsq SHELLS 14 RESOLUTION 999 1.42
```

Supplementary Text

Derivation of the CC versus $CC_{1/2}$ relationship*

To calculate the intra-data set correlation coefficient $CC_{1/2}$, the measurements belonging to each unique reflection of the experimental data set are randomly assigned to two half-data sets. This assignment is only performed for those unique reflections which have at least two measurements. If the number of available measurements is even, each half-data set receives half of the measurements; if it is odd, a randomly chosen half-data set obtains the extra measurement. Next, within each half-data set, the average intensity is calculated for each unique reflection. We thus obtain two half-data sets with intensities I_1 and I_2.

Using acute brackets to denote averages taken over the unique reflections in a given resolution bin, we can now consider the following quantities defined for an, in principle, infinitely large population of measurements:

$J - \langle J \rangle := \tau$: "true" measurements with mean zero and variance σ_τ^2

ε_1 : independent errors with mean zero and variance σ_ε^2

ε_2 : independent errors with mean zero and variance σ_ε^2

We assume that the variance of ε_1 equals that of ε_2, and that τ, ε_1, and ε_2 are mutually independent. Now, consider $I_1 - \langle I_1 \rangle := y_1 = \tau + \varepsilon_1$ and $I_2 - \langle I_2 \rangle := y_2 = \tau + \varepsilon_2$. Then both y_1 and y_2 have mean zero and variance $\sigma_\tau^2 + \sigma_\varepsilon^2$. The correlation between y_1 and y_2 is given by:

$$CC_{1/2} := Corr(I_1, I_2) = Corr(y_1, y_2) = \frac{Cov(y_1, y_2)}{\sqrt{(\sigma_\tau^2 + \sigma_\varepsilon^2)(\sigma_\tau^2 + \sigma_\varepsilon^2)}}$$

$$= \frac{E[y_1 y_2] - E[y_1]E[y_2]}{\sigma_\tau^2 + \sigma_\varepsilon^2}$$

$$= \frac{E[y_1 y_2]}{\sigma_\tau^2 + \sigma_\varepsilon^2}$$

$$= \frac{E[(\tau + \varepsilon_1)(\tau + \varepsilon_2)]}{\sigma_\tau^2 + \sigma_\varepsilon^2}$$

$$= \frac{E[\tau^2 + \varepsilon_1\tau + \varepsilon_2\tau + \varepsilon_1\varepsilon_2]}{\sigma_\tau^2 + \sigma_\varepsilon^2}$$

$$= \frac{E\left[\tau^2\right]}{\sigma_\tau^2 + \sigma_\varepsilon^2}$$

$$= \frac{\sigma_\tau^2}{\sigma_\tau^2 + \sigma_\varepsilon^2}$$

Next, let

$$y = \frac{y_1 + y_2}{2}$$

$$\varepsilon = \frac{\varepsilon_1 + \varepsilon_2}{2}$$

so that *y* has mean zero and variance $\sigma_\tau^2 + \frac{\sigma_\varepsilon^2}{2}$. Then

$$CC_{true} := Corr(\frac{\mathrm{I}_1 + \mathrm{I}_2}{2}, \mathrm{J}) = Corr(y, \tau) = \frac{Cov(y, \tau)}{\sqrt{\sigma_\tau^2 + \sigma_\varepsilon^2 / 2}\sqrt{\sigma_\tau^2}}$$

$$= \frac{E[y\tau] - E[y]E[\tau]}{\sqrt{\sigma_\tau^2 + \sigma_\varepsilon^2 / 2}\sqrt{\sigma_\tau^2}}$$

$$= \frac{E[(\frac{y_1 + y_2}{2})\tau]}{\sqrt{\sigma_\tau^2 + \sigma_\varepsilon^2 / 2}\sqrt{\sigma_\tau^2}}$$

$$= \frac{E[(\frac{\tau + \varepsilon_1 + \tau + \varepsilon_2}{2})\tau]}{\sqrt{\sigma_\tau^2 + \sigma_\varepsilon^2 / 2}\sqrt{\sigma_\tau^2}}$$

$$= \frac{E[(\tau + \varepsilon)\tau]}{\sqrt{\sigma_\tau^2 + \sigma_\varepsilon^2 / 2}\sqrt{\sigma_\tau^2}}$$

$$= \frac{\sigma_\tau^2}{\sqrt{\sigma_\tau^2 + \sigma_\varepsilon^2 / 2}\sqrt{\sigma_\tau^2}}$$

$$= \sqrt{\frac{\sigma_\tau^2}{\sigma_\tau^2 + \sigma_\varepsilon^2 / 2}}$$

From the above follows that

$$\frac{2}{[CC_{true}]^2} = \frac{2(\sigma_\tau^2 + \sigma_\varepsilon^2 / 2)}{\sigma_\tau^2} = 1 + \frac{\sigma_\tau^2 + \sigma_\varepsilon^2}{\sigma_\tau^2} = 1 + \frac{1}{CC_{1/2}}$$

which yields

$$CC_{true} = \sqrt{\frac{2CC_{1/2}}{1+CC_{1/2}}}$$

Provided the sample size is large, this relationship will be approximately fulfilled for a sample drawn from the population.

Thus we can consider $CC^* := \sqrt{\frac{2CC_{I/2}}{1+CC_{I/2}}}$, when calculated for a finite sample, as an estimate of CC_{true} .

Systematic errors may invalidate one or more of the assumptions of independence of τ, ε_1,and ε_2. We note that, depending on their type, systematic errors will often increase, and in some cases decrease, CC* relative to CC_{true}. For example, an increase of CC* over CC_{true} would result if ε_1 and ε_2 have the same sign, for significantly more than half of the reflections.

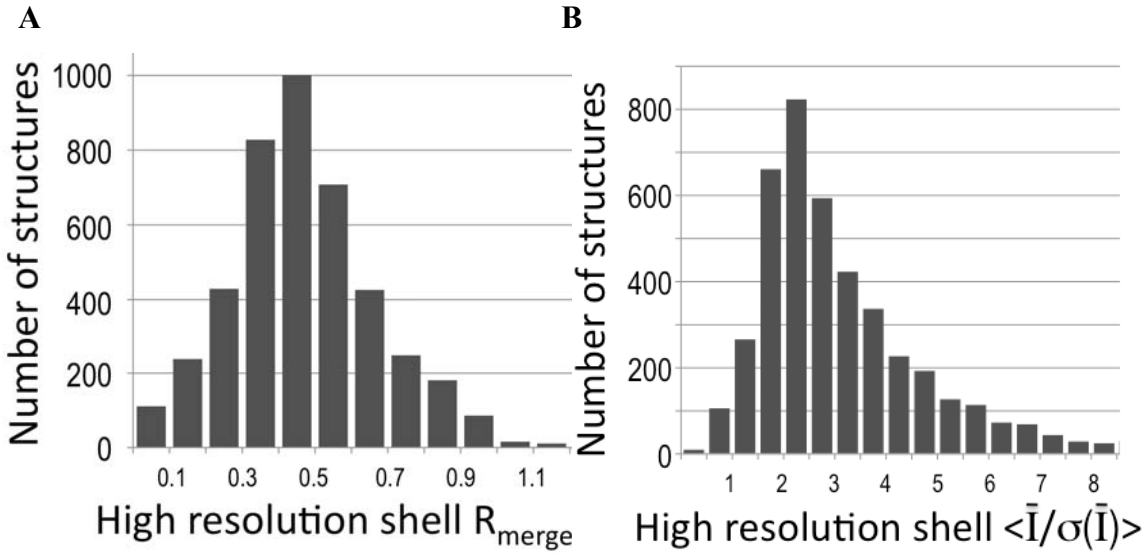

Figure S1. Highest-resolution shell statistics for recently determined protein structures. Histograms are based on all structures in the Protein Data Bank (PDB) (*28*) with a 2010 deposition date. (A) Histogram of the highest resolution shell R_{merge} values in 4,304 structures with a value reported in the PDB entry. Of these, 93% have $R_{merge} < 0.80$ in the highest resolution shell. (B) Histogram of the highest resolution shell $\langle \bar{I}/\sigma(\bar{I}) \rangle$ values in 4,193 structures with a value reported in the PDB entry. Of these, 91% have $\langle \bar{I}/\sigma(\bar{I}) \rangle > 1.5$ in the highest resolution shell.

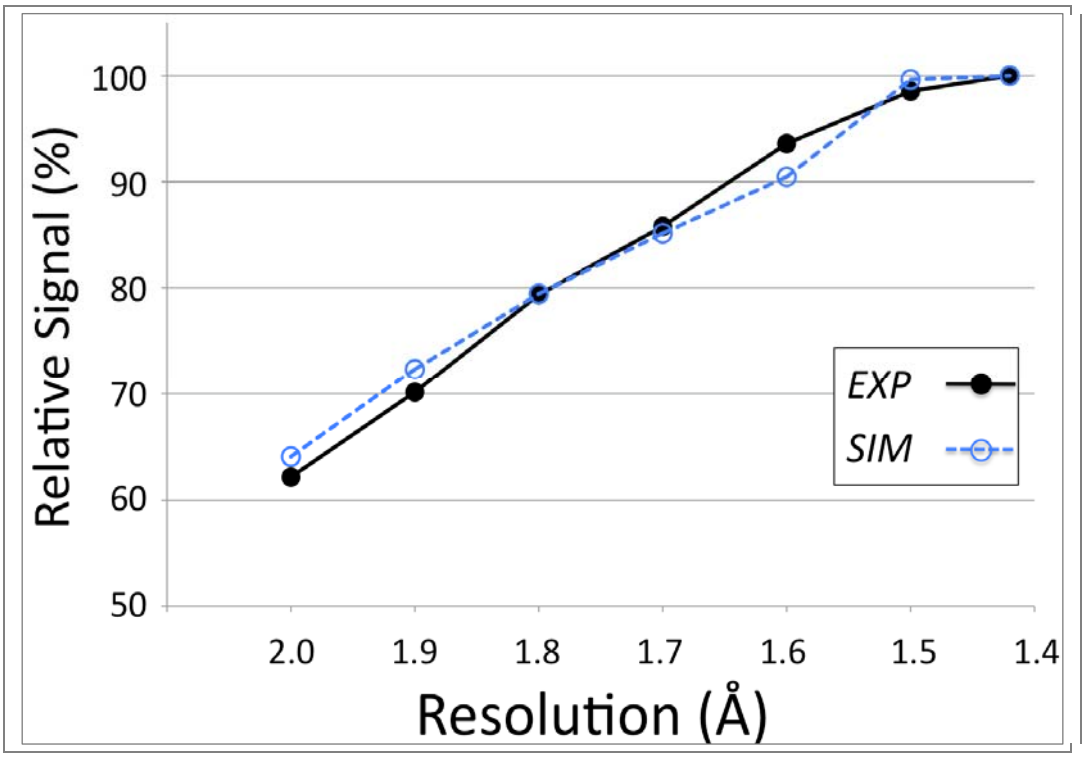

Figure S2. Data set signal as a function of resolution as seen in isomorphous difference electron density maps. Plotted is the relative signal present as a function of resolution in difference maps based on the ***EXP*** (black circles) and ***SIM*** (blue open circles) data sets. Signal is measured as electron density peak heights in standard deviations, with 100% for each data set being defined as the tallest peak height obtained among the maps calculated at the resolutions indicated. In both cases the highest peak height occurred for the 1.42 Å resolution map. The results show that signal is present strongly out to 1.5 Å resolution, with a small further increase between 1.5 and 1.42 Å. The isomorphous difference Fourier maps were calculated between the ***EXP*** or ***SIM*** data set and a 1.5 Å resolution refined model for unliganded CDO [PDB 2B5H (*9*)]. Since the phases for these maps come from the unliganded structure, the maps are unbiased with regard to the electron density signal for the ligand and differ only in the high resolution cutoff of the experimental data. All difference maps were calculated on the same grid, and show the largest peak associated with a 0.5 Å shift in the active site iron.

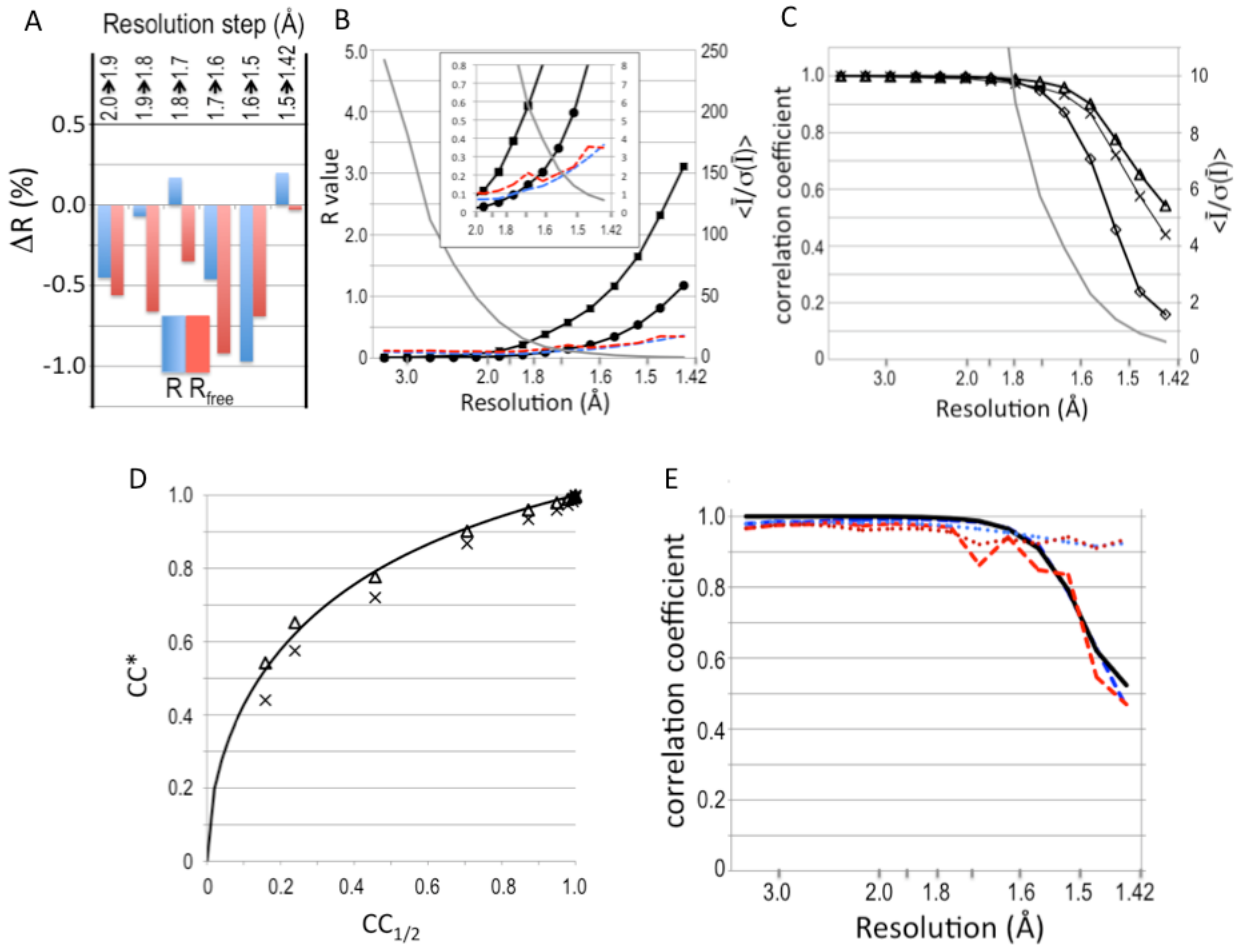

Figure S3. The SIM data set behavior qualitatively matches that of the EXP data set. (**A**) The same as Figure 1, but for the ***SIM*** data set. For each incremental step of resolution from X->Y (top legend), the pair of bars gives the changes in overall R_{work} (blue) and R_{free} (red) for the model refined at resolution Y with respect to those for the model refined at resolution X, with both R values calculated at resolution X. (**B**) Same as Figure 2, but for the ***SIM*** data set. R_{meas} (squares) and R_{pim} (circles) are compared with R_{work} (blue) and R_{free} (red) from 1.42 Å resolution refinements. $<\bar{I}/\sigma(\bar{I})>$ (grey) is also plotted. Inset is a close-up of the plot beyond 2 Å resolution. (**C**) Same as Figure 3, but for the ***SIM*** data set. Plotted as a function of resolution are $CC_{1/2}$ (open diamonds), CC for ***SIM*** compared with the underlying true data set from which ***SIM*** was generated (X's), CC for ***SIM*** compared with a related simulated data set but with about 15-fold higher intensity (open triangles), and $<\bar{I}/\sigma(\bar{I})>$ (grey). The latter CC is equivalent to the comparison of ***EXP*** with 3ELN shown in Figure 3. (**D**) Same as Figure 4A, but for the ***SIM*** data set. Plotted is the analytical relationship (eqn. 3) between $CC_{1/2}$ and CC* (black curve). Also roughly following the CC* curve are the CC values for ***SIM*** data comparisons as defined in panel C (X's and open triangles). (**E**) Same as Figure 4B, but for the ***SIM*** data set. Plotted as a function of resolution are CC* (black), and CC_{work} (blue dashed) and CC_{free} (red dashed) from the 1.42 Å refinement, as well as CC_{work} (blue dotted) and CC_{free} (red dotted) for the refined model against the underlying true data set from which ***SIM*** was generated.

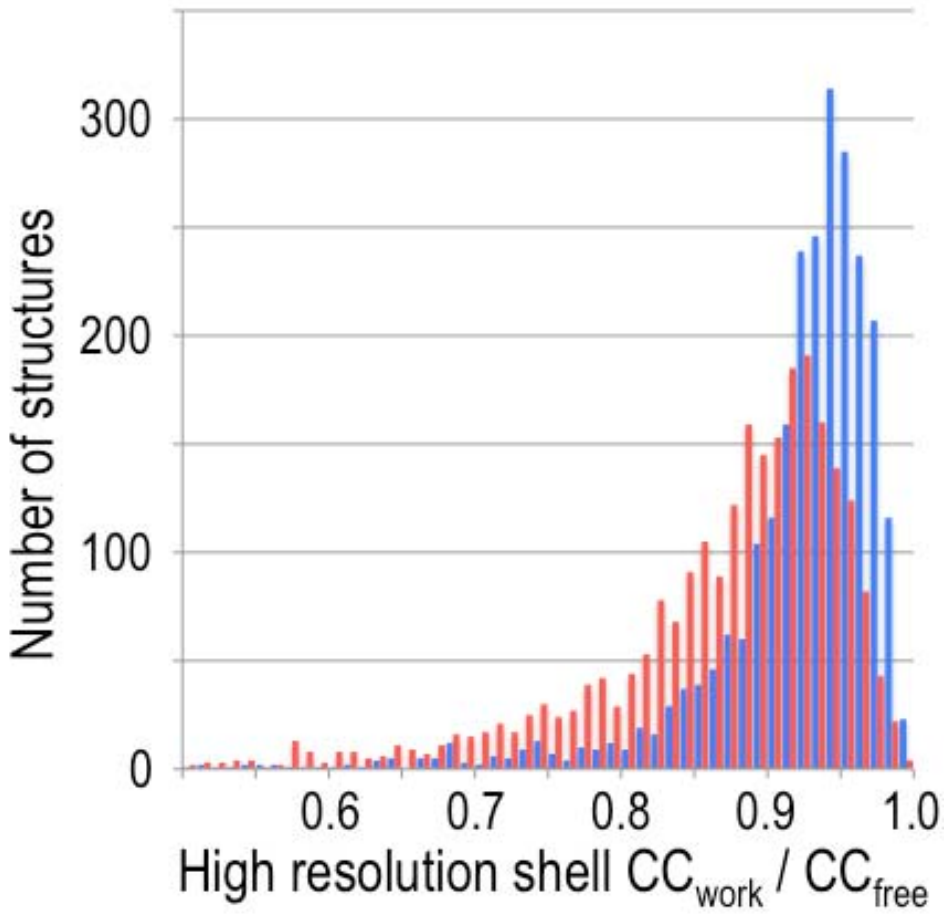

Figure S4. CC_{work} and CC_{free} in the high resolution shell of recent structures as an indicator of CC*. 2,524 X-ray structures deposited in 2010 and re-refined in the PDB_REDO project (*29*) were used to calculate CC_{work} (blue bars) and CC_{free} (red bars). Since PDB_REDO uses state-of-the-art algorithms and avoids overfitting, CC_{work} can be expected to be a reliable lower-bound estimate for CC*. Of the deposited structures, 90% have a highest resolution bin with $CC_{work} \geq 0.85$, proving that currently used high-resolution cutoffs are too conservative and discard many reflections that would enhance model accuracy.

Table S1: Data statistics for the *EXP* and 3ELN data sets[a]

	EXP		3ELN[a]
Resolution (Å)	40 - 1.80 (1.86-1.80)	40-1.42 (1.46-1.42)	20-1.42 (1.44-1.42)
Unique reflections	19874 (1929)	39483 (2494)	39569 (-)
R_{meas}	0.109 (0.612)	0.148 (4.378)	0.095 (0.86)
R_{pim}	0.027 (0.148)	0.038 (2.182)	0.043 (0.365)
$CC_{1/2}$ outer shell; # pairs	0.975 ; n=1929	0.088; n=2101	-
$<\bar{I}/\sigma(\bar{I})>$	44.9 (7.8)	23.1 (0.28)	40.3 (3.0)
Completeness (%)	100.0 (100.0)	99.7 (96.3)	100.0 (100.0)
Multiplicity	17.1 (17.0)	13.8 (3.3)	37.5 (22.9)

[a] for reference, statistics are also shown here for the strong 3ELN reference data set taken from Simmons *et al*. (*7*) and which diffracts to 1.42 Å by conventional standards. For these data, $CC_{1/2}$ is not available and R_{mrgd-F} is reported in place of R_{pim}.

Table S2: Refinement statistics for the *EXP* data set

Overall R_{work} / R_{free} values for isotropic refinements [a]

High-resolution limit for R value calculation (Å)	High-resolution limit for refinement (Å)						
	2.0	**1.9**	**1.8**	**1.7**	**1.6**	**1.5**	**1.42**
2.0	0.1621/ 0.1988	0.1583/ 0.1954	-	-	-	-	0.1646/ 0.1909
1.9	-	0.1581/ 0.1968	0.1619/ 0.1916	-	-	-	0.1643/ 0.1878
1.8	-	-	0.1653/ 0.1967	0.1668/ 0.1936	-	-	0.1678/ 0.1918
1.7	-	-	-	0.1724/ 0.2014	0.1729/ 0.2001	-	0.1714/ 0.1975
1.6	-	-	-	-	0.1828/ 0.2092	0.1781/ 0.2083	0.1787/ 0.2045
1.5	-	-	-	-	-	0.1877/ 0.2168	0.1877/ 0.2127
1.42	-	-	-	-	-	-	0.1996/ 0.2230

Overall R_{work} and R_{free} for anisotropic refinements[b]

2.0	0.1430/ 0.2004	0.1396/ 0.1931	-	-	-	-	0.1410/ 0.1761
1.9	-	0.1380/ 0.1915	0.1362/ 0.1863	-	-	-	0.1388/ 0.1715
1.8	-	-	0.1375/ 0.1906	0.1411/ 0.1819	-	-	0.1401/ 0.1747
1.7	-	-	-	0.1466/ 0.1902	0.1419/ 0.1852	-	0.1419/ 0.1801
1.6	-	-	-	-	0.1509/ 0.1943	0.1476/ 0.1889	0.1491/ 0.1884
1.5	-	-	-	-	-	0.1573/ 0.1986	0.1582/ 0.1981
1.42	-	-	-	-	-	-	0.1701/ 0.2085[c]

[a] The rms deviations from ideality are 0.015 to 0.017 Å for bond lengths and 1.5 to 1.6° for bond angles, with a systematic trend that the models refined at higher resolution have better ideality.
[b] The rms deviations from ideality are 0.014 to 0.016 Å for bond lengths and 1.39 to 1.5° for bond angles, with a systematic trend that the models refined at higher resolution have better ideality.
[c] The CC_{work}/CC_{free} in the highest resolution shell (1.46-1.42 Å) are 0.382/0.212 .

Table S3: Data statistics for the SIM data set

Resolution (Å)	100-1.6 (1.66-1.60)	100-1.42 (1.46-1.42)
Unique reflections	27821 (2706)	38352 (2145)
R_{meas}	0.027 (0.718)	0.034 (3.128)
R_{pim}	0.0007 (0.195)	0.009 (1.176)
$CC_{1/2}$ outer shell; # pairs	0.893; n=2701	0.159; n=1942
$<\bar{I}/\sigma(\bar{I})>$	79.9 (4.4)	57.7 (0.6)
Completeness (%)	100.0 (100.0)	97.4 (79.8)
Multiplicity	16.9 (14.2)	14.6 (6.1)

Table S4: Refinement statistics for the SIM data set

Overall R_{work} and R_{free} for isotropic refinement[a]

High-resolution limit for R value calculation (Å)	High-resolution limit for refinement (Å)						
	2.0	**1.9**	**1.8**	**1.7**	**1.6**	**1.5**	**1.42**
2.0	0.1282/ 0.1537	0.1237/ 0.1481	-	-	-	-	0.1178/ 0.1339
1.9	-	0.1248/ 0.1488	0.1241/ 0.1422	-	-	-	0.1179/ 0.1334
1.8	-	-	0.1253/ 0.1445	0.1225/ 0.1388	-	-	0.1190/ 0.1352
1.7	-	-	-	0.1244/ 0.1428	0.1225/ 0.1400	-	0.1207/ 0.1373
1.6	-	-	-	-	0.1250/ 0.1433	0.1225/ 0.1435	0.1234/ 0.1409
1.5	-	-	-	-	-	0.1282/ 0.1475	0.1291/ 0.1451
1.42	-	-	-	-	-	-	0.1370/ 0.1527

Overall R_{work} and R_{free} for anisotropic refinement[b]

2.0	0.1105/ 0.1790	0.1060/ 0.1610	-	-	-	-	0.0906/ 0.1151
1.9	-	0.1065/ 0.1622	0.0988/ 0.1321	-	-	-	0.0895/ 0.1143
1.8	-	-	0.0991/ 0.1339	0.1008/ 0.1304	-	-	0.0886/ 0.1145
1.7	-	-	-	0.1015/ 0.1334	0.0969/ 0.1242	-	0.0889/ 0.1166
1.6	-	-	-	-	0.0980/ 0.1262	0.0883/ 0.1193	0.0905/ 0.1195
1.5	-	-	-	-	-	0.0934/ 0.1239	0.0954/ 0.1236
1.42	-	-	-	-	-	-	0.1037/ 0.1321[c]

[a] The rms deviations from ideality are 0.015 to 0.018 Å for bond lengths and 1.55 to 1.62° for bond angles, with a systematic trend that the models refined at higher resolution have better ideality.
[b] The rms deviations from ideality are 0.013 to 0.016 Å for bond lengths and 1.37 to 1.53° for bond angles, with no systematic trends with resolution.
[c] The CC_{work}/CC_{free} in the highest resolution shell (1.46-1.42 Å) are 0.462/0.470.

Table S5. Data quality and paired refinement statistics for extending the resolution limits of two medium resolution structures from the literature.[a]

PDB code	3E4F				3N0S			
Crystal form	$P2_12_12_1$ (a=36.31Å, b=108.05Å, c=132.81Å)				$P2_1$ (a=72.04Å, b=109.44Å, c=74.05Å, β=111.86°)			
Outer shell data reduction statistics								
Data source	*Published*	Reprocessed			*Published*	Reprocessed		
Outer shell (Å)	*2.03-2.0*	2.03-2.0	1.83-1.8	1.73-1.7	*2.19-2.15*	2.19-2.15	2.04-2.0	1.89-1.85
Multiplicity	*8.1*	7.9	5.0	3.5	*2.8*	4.2	4.3	4.2
Completeness (%)	*100*	100.0	99.9	93.4	*95.9*	99.1	99.2	98.7
R_{merge}	*0.496*	0.534	1.490	2.656	*0.487*	0.689	1.133	2.338
$\langle I/\sigma(I)\rangle$	*4.5*	4.0	0.9	0.4	*1.9*	2.5	1.5	0.7
$CC_{1/2}$; # pairs	-	0.909 (1787)	0.393 (2376)	0.166 (2469)		0.708 (3259)	0.491 (3823)	0.194 (4822)
Refinement statistics								
Resolution range (Å)	*50-2.0*	50-2.0	50-1.8	50-1.7	*50-2.15*	50-2.15	50-2.0	50-1.85
R_{work}	*0.172*	0.1675	0.1781	0.1932	*0.174*	0.1679	0.1784	0.1935
R_{free}	*0.226*	0.2040	0.2122	0.2227	*0.228*	0.2194	0.2252	0.2365
ΔR_{work} pair[b]	-	-	-0.0004	+0.0011	-	-	+0.0008	+0.0008
ΔR_{free} pair[b]	-	-	-0.0010	-0.0015	-	-	-0.0038	-0.0069
rmsd bonds	*0.017*	0.012	0.012	0.012	*0.021*	0.015	0.014	0.014
rmsd angles	*1.7*	1.4	1.4	1.4	*1.7*	1.6	1.6	1.6

[a] Published values are taken from Klimecka *et al.* (*27*). For selection criteria and analysis protocols, see Materials and Methods.

[b] Each ΔR reports the change compared with the refinement done at the published resolution limit when calculated at that same resolution limit.

References and Notes

1. A. J. C. Wilson, Largest likely values for the reliability index. *Acta Crystallogr.* **3**, 397 (1950). doi:10.1107/S0365110X50001129

2. A. T. B. Brünger, Free *R* value: A novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472 (1992). doi:10.1038/355472a0 Medline

3. U. W. Arndt, R. A. Crowther, J. F. W. Mallett, A computer-linked cathode-ray tube microdensitometer for x-ray crystallography. *J. Phys. E Sci. Instrum.* **1**, 510 (1968). doi:10.1088/0022-3735/1/5/303 Medline

4. K. Diederichs, P. A. Karplus, Improved *R*-factors for diffraction data analysis in macromolecular crystallography. *Nat. Struct. Biol.* **4**, 269 (1997). doi:10.1038/nsb0497-269 Medline

5. M. S. Weiss, Global indicators of X-ray data quality. *J. Appl. Cryst.* **34**, 130 (2001). doi:10.1107/S0021889800018227

6. P. R. Evans, An introduction to data reduction: Space-group determination, scaling and intensity statistics. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 282 (2011). doi:10.1107/S090744491003982X Medline

7. C. R. Simmons *et al*., A putative Fe2+-bound persulfenate intermediate in cysteine dioxygenase. *Biochemistry* **47**, 11390 (2008). doi:10.1021/bi801546n Medline

8. Materials and methods are available as supplementary material on *Science* Online.

9. C. R. Simmons *et al*., Crystal structure of mammalian cysteine dioxygenase. A novel mononuclear iron center for cysteine thiol oxidation. *J. Biol. Chem.* **281**, 18723 (2006). doi:10.1074/jbc.M601555200 Medline

10. J. Wang, Inclusion of weak high-resolution X-ray data for improvement of a group II intron structure. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 988 (2010). doi:10.1107/S0907444910029938 Medline

11. P. Evans, Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 72 (2006). doi:10.1107/S0907444905036693 Medline

12. N. A. Rahman, *A Course in Theoretical Statistics* (Griffin, London, 1968).

13. T. R. Schneider, G. M. Sheldrick, Substructure solution with SHELXD. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 1772 (2002). doi:10.1107/S0907444902011678 Medline

14. P. B. Rosenthal, R. Henderson, Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333**, 721 (2003). doi:10.1016/j.jmb.2003.07.013 Medline

15. P. Bobko, *Correlation and Regression: Applications for Industrial Organizational Psychology and Management* (Sage Publications, Thousand Oaks, 2001).

16. H. G. Gauch Jr., Prediction, Parsimony and noise. *Am. Sci.* **81**, 468 (1993).

17. V. Luzzati, Resolution d'une structure cristalline lorsque les positions d'une partie des atoms sont connues: Traitement statistique. *Acta Crystallogr.* **6**, 142 (1953). doi:10.1107/S0365110X53000508

18. D. W. J. Cruickshank, Remarks about protein structure precision. *Acta Crystallogr. D Biol. Crystallogr.* **55**, 583 (1999). doi:10.1107/S0907444998012645 Medline

19. R. A. Steiner, A. A. Lebedev, G. N. Murshudov, Fisher's information in maximum-likelihood macromolecular crystallographic refinement. *Acta Crystallogr. D Biol. Crystallogr.* **59**, 2114 (2003). doi:10.1107/S0907444903018675 Medline

20. K. Diederichs, Simulation of X-ray frames from macromolecular crystals using a ray-tracing approach. *Acta Crystallogr. D Biol. Crystallogr.* **65**, 535 (2009). doi:10.1107/S0907444909010282 Medline

21. R. Fisher, *Statistical Methods for Research Workers* (Oliver and Boyd, Edinburgh, 1925), §33–34.

22. Ten independent random partitionings of the data into the two subsets for calculating $CC_{1/2}$ yielded standard deviations of <0.02 in all resolution ranges, and agreed reasonably with the expected standard error as calculated by $\sigma(CC)=(1 - CC^2)/\sqrt{(n - 1)}$ where n is the number of observations contributing to the CC calculation (*21*).

23. W. Kabsch, XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125 (2010). doi:10.1107/S0907444909047337 Medline

24. W. Kabsch, Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 133 (2010). doi:10.1107/S0907444909047374 Medline

25. P. D. Adams *et al*., PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213 (2010). doi:10.1107/S0907444909052925 Medline

26. M. D. Winn *et al*., Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235 (2011). doi:10.1107/S0907444910045749 Medline

27. M. M. Klimecka *et al*., Structural analysis of a putative aminoglycoside *N*-acetyltransferase from *Bacillus anthracis*. *J. Mol. Biol.* **410**, 411 (2011). doi:10.1016/j.jmb.2011.04.076 Medline

28. H. M. Berman *et al*., The Protein Data Bank. *Nucleic Acids Res.* **28**, 235 (2000). doi:10.1093/nar/28.1.235 Medline

29. R. P. Joosten, T. Womack, G. Vriend, G. Bricogne, Re-refinement from deposited X-ray data can deliver improved models for most PDB entries. *Acta Crystallogr. D Biol. Crystallogr.* **65**, 176 (2009). doi:10.1107/S0907444908037591 Medline