Acta Crystallographica Section D
Biological Crystallography
ISSN 1399-0047

Weak data do not make a free lunch, only a cheap meal

Zhipu Luo,[a] Kanagalaghatta Rajashankar[b] and Zbigniew Dauter[a]*

[a]Synchrotron Radiation Research Section, MCL, National Cancer Institute, Argonne National Laboratory, Argonne, IL 90439, USA, and [b]NE-CAT and Department of Chemistry and Chemical Biology, Cornell University, Argonne National Laboratory, Argonne, IL 60439, USA

Correspondence e-mail: dauter@anl.gov

Received 16 May 2013
Accepted 27 September 2013

Four data sets were processed at resolutions significantly exceeding the criteria traditionally used for estimating the diffraction data resolution limit. The analysis of these data and the corresponding model-quality indicators suggests that the criteria of resolution limits widely adopted in the past may be somewhat conservative. Various parameters, such as $R_{\rm merge}$ and $I/\sigma(I)$, optical resolution and the correlation coefficients $CC_{1/2}$ and CC*, can be used for judging the internal data quality, whereas the reliability factors R and $R_{\rm free}$ as well as the maximum-likelihood target values and real-space map correlation coefficients can be used to estimate the agreement between the data and the refined model. However, none of these criteria provide a reliable estimate of the data resolution cutoff limit. The analysis suggests that extension of the maximum resolution by about 0.2 Å beyond the currently adopted limit where the $I/\sigma(I)$ value drops to 2.0 does not degrade the quality of the refined structural models, but may sometimes be advantageous. Such an extension may be particularly beneficial for significantly anisotropic diffraction. Extension of the maximum resolution at the stage of data collection and structure refinement is cheap in terms of the required effort and is definitely more advisable than accepting a too conservative resolution cutoff, which is unfortunately quite frequent among the crystal structures deposited in the Protein Data Bank.

1. Introduction

In single-crystal diffraction data, the average intensities of reflections as well as data-measurement accuracy and precision diminish with increasing resolution. Eventually the signal-to-noise ratio tends to zero and the recorded 'data' contain no information, but there is no consensus in the community in regard to a proper cutoff value. Therefore, the resolution limit of diffraction data measured from crystals of macromolecules cannot be objectively defined. Hence, various criteria have been and still are used for this purpose. It is amazing that in spite of enormous progress in data-acquisition methodology and structure solution and refinement over the past several decades, the resolution limit of the measured data is still judged in a manner similar to that used when data were recorded on photographic film at room temperature. Only recently have attempts been made to overcome this conservatism (Karplus & Diederichs, 2012; Diederichs & Karplus, 2013; Evans & Murshudov, 2013).

The traditional criteria for judging the data resolution limit are the $R_{\rm merge}$ and mean $I/\sigma(I)$ values. Both of these have drawbacks. The value of the commonly used form of $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$ increases with increasing multiplicity of measurements, whereas the accuracy

of the resulting estimation of the averaged intensities obviously improves. Multiplicity-independent versions of the intensity merging factors have been proposed in the form of $R_{\rm meas} = \sum_{hkl}\{N(hkl)/[N(hkl)-1]\}^{1/2}\sum_i |I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$ and $R_{\rm p.i.m.} = \sum_{hkl}\{1/[N(hkl)-1]\}^{1/2}\times\sum_i |I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$ (Diederichs & Karplus, 1997; Weiss & Hilgenfeld, 1997; Weiss, 2001). However, they are not universally used and are not required by the Protein Data Bank (PDB; Berman *et al.*, 2000) or by journals that publish X-ray structures. The mean $I/\sigma(I)$ ratio is in principle a better criterion, but in practice it relies on correctly estimating values of the intensity uncertainties, $\sigma(I)$, which is not always easy to achieve. It is interesting to note that some data-processing programs represent this signal-to-noise parameter as $\langle I\rangle/\langle\sigma(I)\rangle$ and others as $\langle I/\sigma(I)\rangle$, but when intensities are weak both forms of averaging lead to similar numerical values (Supplementary Table S1[1]). In the present text $I/\sigma(I)$ is used to mean either of the two versions of this parameter. Photon-counting detectors, such as Pilatus, reproduce the number of diffracted X-ray quanta, but CCD detectors only provide proportional numerical values. As a result, simple counting statistics may not apply and the estimation of the uncertainties requires certain corrections to be applied. If the data multiplicity is sufficient, the proper level of uncertainties can be estimated from the spread of individual measurements around the average, but when the redundancy is low the estimated uncertainties may not be reliable.

Several earlier reports suggested that the $R_{\rm merge}$ value in the highest resolution range should not exceed 20–40% (Dauter, 1999) or 30–60% (Wlodawer *et al.*, 2008) depending on the crystal symmetry. Most of the authors suggest that the limiting value of $I/\sigma(I)$ is about 2.0 (Dauter, 1999; Evans, 2006; Wlodawer *et al.*, 2008).

There were, however, instances that suggested that these criteria should be relaxed, at least for certain applications. *Ab initio* solution of the c_6 cytochrome structure (Frazão *et al.*, 1995) by direct methods was not possible with 1.2 Å resolution data. Extension of the resolution to 1.1 Å, beyond the 2.0 limit of $I/\sigma(I)$, led to successful solution of this structure with *SHELXS* (Sheldrick, 2008). It was concluded that 'the reflections in the 1.2–1.1 Å shell provided useful information in the refinement and especially in the solution of the structure'. Diffraction data from a crystal of the Shiga-like toxin I B pentamer (Ling *et al.*, 1998) were characterized by an $I/\sigma(I)$ of 2.0 at 3.6 Å resolution, but after analyzing the results it was decided that data beyond this limit contributed favorably to the quality of the electron-density maps and the structure was refined against the 2.8 Å resolution data. However, it is not clear how much improvement resulted from the resolution extension, since 20-fold noncrystallographic symmetry was also used in this case. In fact, about 18% of all PDB depositions report an $I/\sigma(I)$ ratio in the highest resolution shell that is lower than 2.0, but the distribution of $I/\sigma(I)$ among all PDB submissions is very broad (for detailed statistics, see Supplementary Table S2). 48% of depositions have a highest resolution $I/\sigma(I)$ ratio of greater than 3.0, suggesting that many crystallographers underestimate the resolution limit of diffraction data, not using the full potential of their crystals.

In the early days of protein crystallography, phasing and refinement algorithms were relatively simple and were not based on advanced statistical and probabilistic approaches. The inclusion of very weak reflections with no reliable uncertainties could bring more harm than benefit to the computational procedures, so it was sensible to avoid such noisy, weak, very high-resolution data. However, almost all contemporary crystallographic programs (with the exception of *SHELXL*) used for the phasing and refinement of macromolecular structures are based on statistically elaborate principles, *e.g.* maximum likelihood. *SHELXL* does not utilize the maximum-likelihood principle, but employs elaborate weighting of reflections. These algorithms properly utilize reflection uncertainties and are able to reliably weigh the influence of even very weak and relatively inaccurate reflection intensities. This suggests that the resolution limit of diffraction data may be liberally extended further than the current conservative criteria suggest.

Recent publications have addressed this issue (Karplus & Diederichs, 2012; Diederichs & Karplus, 2013, subsequently referred to as KD; Evans & Murshudov, 2013). KD suggested that instead of R factors or $I/\sigma(I)$, it might be more correct to use the statistically well established correlation coefficient $CC_{1/2}$ between the reflection intensities merged within two randomly split halves of the whole data set or, even better, the estimation of the true correlation between the measured and ideal intensities $CC^* = [2CC_{1/2}/(1 + CC_{1/2})]_{1/2}$. These correlation coefficients are comparable to the correlation coefficients $CC_{\rm work}$ and $CC_{\rm free}$ between the experimental intensities and those calculated from the atomic model. In contrast, there is no statistical relationship between the data $R_{\rm merge}$ and the refinement R and $R_{\rm free}$ values.

KD reprocessed the diffraction images obtained for two crystal structures, PDB entries 3e4f and 3n0s, from the Center for Structural Genomics of Infectious Diseases (CSGID) archive (http://csgid.oprg/csgid/) at a much higher resolution limit than in the original investigations (Klimecka *et al.*, 2011). Using as a measure of success the values of R and $R_{\rm free}$ obtained at different resolutions but compared at equal resolution, they concluded that the extension of resolution is beneficial up to a $CC_{1/2}$ as small as 0.1, when the $R_{\rm merge}$ reaches above 300% and the $I/\sigma(I)$ ratio is close to 0.3. Evans & Murshudov (2013) used various simulated data and concluded that changing the resolution cutoff over a range of 0.3 Å makes only a small difference in terms of the interpretability of electron density.

It has been shown that the extension of resolution beyond the limit of the experimentally measured reflections by inclusion of the calculated structure factors is beneficial to phasing and model-building procedures (Yao *et al.*, 2005; Caliandro *et al.*, 2005). This procedure was implemented in *SHELXE* and was named the 'free-lunch algorithm' (Sheldrick, 2008).

[1] Supporting information has been deposited in the IUCr electronic archive (Reference: TZ5036).

It is interesting to determine whether the potential benefit observed by including the weak data extending beyond the traditional resolution limit is analogous to the free-lunch effect. We performed calculations similar to those executed by KD. In addition to analyzing the effect of observed weak reflections, we also performed refinements with randomly shuffled reflection intensities in the highest resolution bins. The latter should result in inferior refinement statistics if the weak data do have a positive effect on refinement. On the other hand, if the weak data do not play a significant role in refinements the resulting statistics in either case should be similar. In addition to the R and $R_{\rm free}$ factors, we also investigated the behavior of other criteria, such as the optical resolution, the maximum-likelihood refinement target function and the difference electron-density maps.

2. Materials and methods

The four data sets used in the calculations were obtained through independent integration of diffraction images. Two of them, 3e4f and 3n0s (Klimecka *et al.*, 2011), were the same as those used by KD in their work, with the images obtained from the CSGID archive. In addition, we used diffraction images corresponding to the structure of the *Thermus aquaticus* single-stranded DNA-binding protein with PDB code 2fxq (SSB; Jędrzejczak *et al.*, 2006) and a data set measured from a crystal of thaumatin.

The details of the diffraction data collection for 3e4f, 3n0s and 2fxq are available in the corresponding publications. The data from thaumatin were collected on SER-CAT beamline 22BM at the Advanced Photon Source (Argonne National Laboratory) using a MAR225 CCD detector and a highly attenuated X-ray beam of 1.000 Å wavelength.

All images were integrated and scaled using the *HKL*-2000 platform (Otwinowski & Minor, 1997). The scaling was additionally executed in the 'no merge original index' mode for the alternative merging and calculation of data statistics with *phenix.merging_statistics* and *phenix.cc_star* (Adams *et al.*, 2010). The resulting detailed diffraction data statistics for

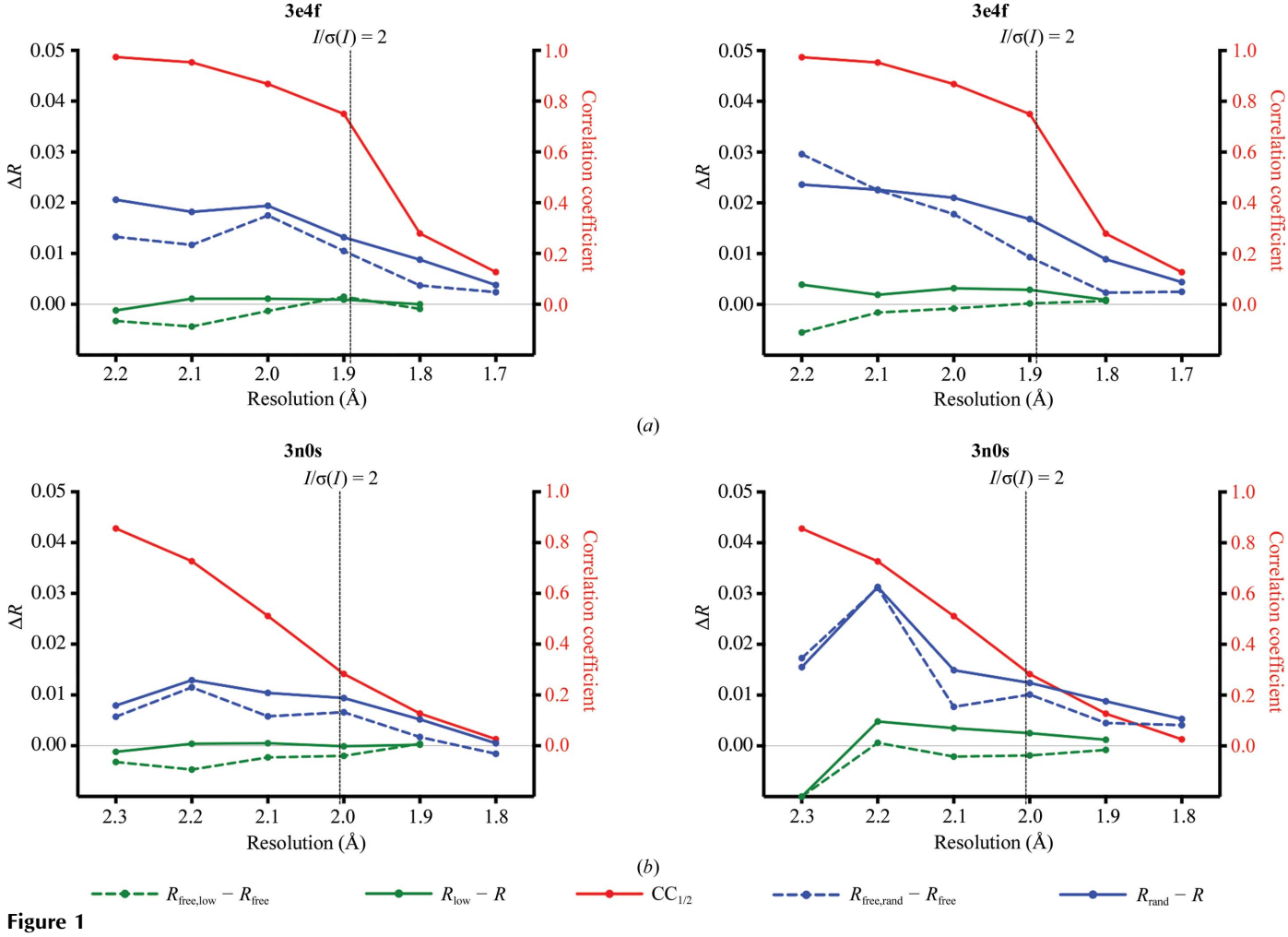

Figure 1
Differences between pairs of R factors for the four structures refined at different resolutions with *phenix.refine* (left) and *REFMAC* (right): 3e4f (*a*), 3n0s (*b*), 2fxq (*c*) and thaumatin (*d*). The green solid lines show the differences between the R factors obtained from two refinements, firstly at resolution d (labelled R) and secondly at a resolution extended by 0.1 Å but calculated at the same resolution (labelled $R_{\rm low}$). The analogous differences in $R_{\rm free}$ factors are indicated as green dashed lines. The blue continuous and dashed lines illustrate the differences between the R and $R_{\rm free}$ factors, respectively, calculated from refinement at resolution d with data randomized within the 0.1 Å-wide highest resolution shell (*i.e.* between d and d + 0.1) and refinement at the same resolution with the original data. The values of $CC_{1/2}$ are also shown as red solid lines.

all crystals are presented in Supplementary Table S1. For each data set, the optical resolution was calculated using *SFCHECK* (Vaguine *et al.*, 1999) and the obtained values in resolution ranges are presented in the same table.

Additional versions of each data set were prepared by randomizing the data in the highest resolution shells. For example, six modified data sets for 3e4f were obtained in which the intensities of reflections beyond 2.3, 2.2, 2.1, 2.0, 1.9 or 1.8 Å were randomly reshuffled. The modification of reflection intensities was performed as follows. All reflections beyond the selected resolution limit were sorted based on resolution and the intensities within each consecutive block of 50 reflections were randomly interchanged. This procedure ensured that the overall intensity statistics and their Wilson plot values were preserved. The reflections reserved for the calculation of $R_{\rm free}$ were not modified, *i.e.* they were excluded from randomization, to keep the same set of $R_{\rm free}$ reflections for all refinements.

All structures were refined analogously to the procedure adopted by KD with *phenix.refine* (Adams *et al.*, 2010) and *REFMAC* (Murshudov *et al.*, 2011). The details of the refinement process are included in the Supporting Information. Each structure was refined at several resolution limits, each differing by 0.1 Å, and the results are presented in Fig. 1. The differences between the R factors obtained from two refinements, the first at resolution d and the second at a resolution extended by 0.1 Å but calculated at the same resolution d, are shown as green solid lines. The analogous differences in $R_{\rm free}$ factors are indicated as green dashed lines.

The same figure includes differences between R factors calculated from refinement at resolution d with data randomized within the 0.1 Å-wide highest resolution shell (*i.e.* between d and $d + 0.1$) and refinement at the same resolution with the original data. The continuous blue line refers to differences in R factors and the dashed blue line shows differences in $R_{\rm free}$ factors.

If two refinements with and without randomized data in the highest resolution shell lead to identical R factors, it means that the reflections in this last shell are too weak to influence the process of model refinement. On the other hand, if the R factors calculated for the model refined with the highest resolution data randomized are higher than those for the model refined with the original data, it suggests that the randomization of the weak data degrades the quality of the refined structure.

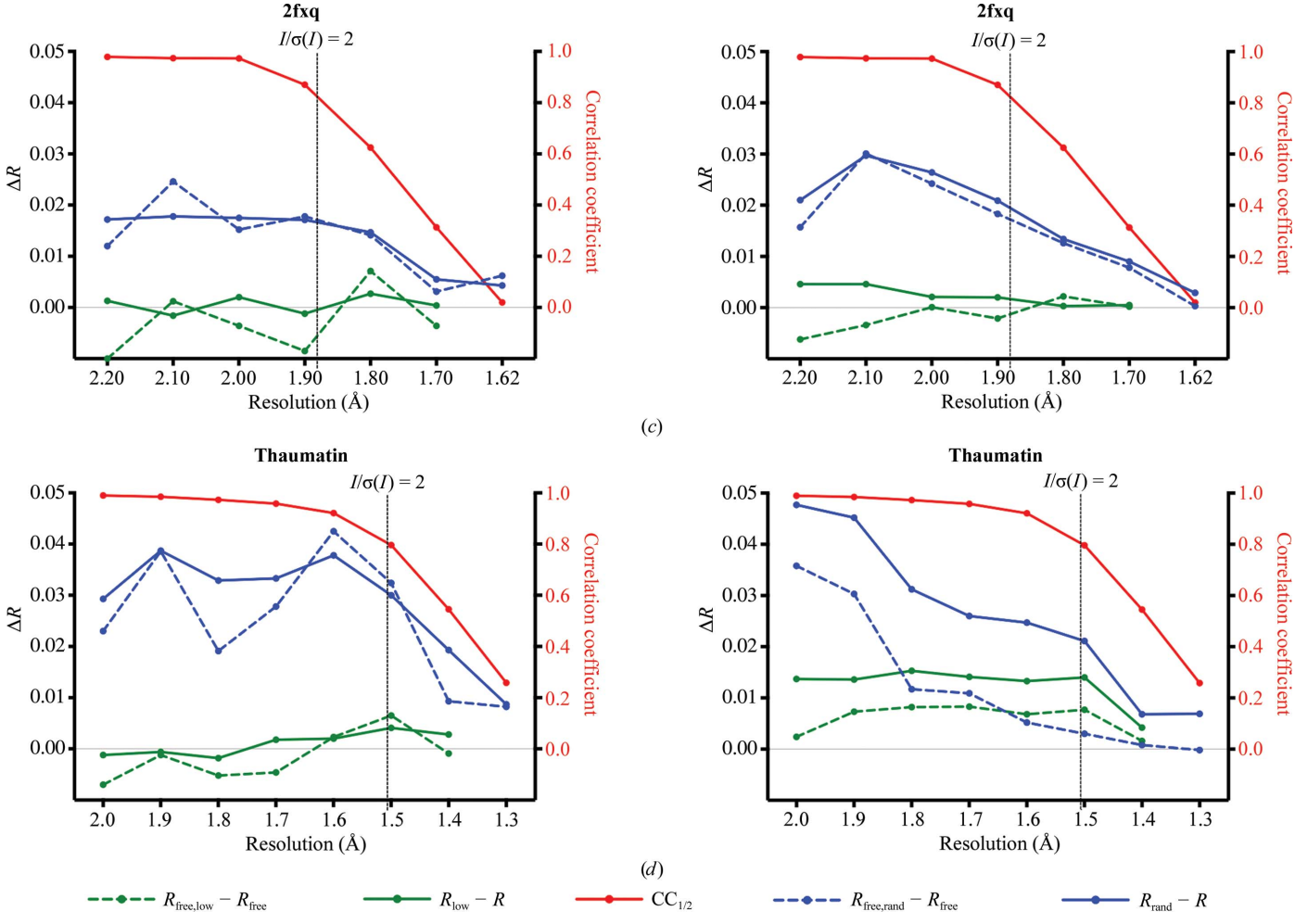

Figure 1 (continued)

The electron-density maps obtained from various combinations of data and resolutions were plotted with *PyMOL* (DeLano, 2002) and compared in terms of their 'real-space' correlation coefficients using *phenix.get_cc_mtz_pdb* (Adams *et al.*, 2010). The maximum-likelihood refinement target-function values were taken from the appropriate refinement log files of *phenix.refine* and *REFMAC*.

3. Results and discussion

The data resolution limits for the three investigated structures, as quoted in the relevant publications, are 2.0 Å for 3e4f, 2.15 Å for 3n0s and 1.85 Å for 2fxq; for thaumatin this limit is estimated as 1.6 Å. The reprocessed data for these structures extended to 1.7, 1.8, 1.62 and 1.3 Å resolution, respectively.

3.1. R and R_{free} factors

The statistics of various R factors are presented in Fig. 1 for the four investigated structures refined with *phenix.refine* (Adams *et al.*, 2010) and *REFMAC* (Murshudov *et al.*, 2011). Each panel shows in green the differences between the values of the R and R_{free} factors calculated at the same resolution (marked on the graph) after refinement at the same resolution and extended by 0.1 Å resolution. These graphs are analogous to those presented by KD. For all four structures, the differences in both the R and R_{free} factors are close to zero but are scattered on the positive and negative sides. Inclusion of data in the additional 0.1 Å resolution range does not seem to have any conclusive effect on the resulting R factors.

Fig. 1 also shows differences between the R and R_{free} factors (blue lines) obtained from two refinements with and without the reflection intensities in the highest 0.1 Å resolution shell randomized. In all of the illustrated cases the differences are positive, showing that randomization degrades the quality of the final structures. However, these differences are most pronounced for lower resolution refinements and diminish with increasing resolution limit of the diffraction data. For all four structures randomization has only a minimal effect for refinements against data extending to the two highest resolution values.

The inclusion of weak high-resolution reflections does not influence the behavior of the R factors in the low-resolution ranges. As seen in the example shown in Supplementary Fig. S1, the values of the R and R_{free} factors are very similar in lower resolution ranges irrespective of whether the refinements were performed with high or limited resolution data.

However, the R_{free} criterion was originally introduced for evaluating the overall refinement protocols and preventing them from overfitting, and this parameter is less useful for judging the results of individual refinement runs owing to its unknown precision. We repeated analogous refinements of the 3e4f structure ten times at various resolution limits against data with different (random) selections of R_{free} reflections and the results are presented in Fig. 2. Whereas the spread of R factors is characterized by an r.m.s.d. of about 0.001, the R_{free} factors differ by up to 0.02, with an r.m.s.d. of about 0.006. This reflects the lower population of the R_{free} reflections, which constitute 5% of the whole set.

The estimated uncertainties of R and R_{free} are comparable to the differences in the R and R_{free} factors presented in Fig. 1. This considerably lowers the usefulness of the behavior of these factors for evaluating the results of individual refinement runs and evaluating the proper data resolution limit, in spite of these differences (as well as those presented by KD) being obtained from refinements with the same set of R_{free} reflections.

Fig. 1 also shows the values of $CC_{1/2}$ from data processing. Analysis of these values, which are also numerically quoted in Supplementary Table S1, does not provide conclusive results. The behavior of the R factors suggests that the useful reflections extend to about 1.8, 1.9, 1.7 and 1.4 Å resolution for 3e4f, 3n0s, 2fxq and thaumatin, respectively, but that the respective values of $CC_{1/2}$ at these resolution limits are about 0.3, 0.1, 0.3 and 0.5 for these structures. Judging from these discrepancies, is not possible to formulate numerical values for $CC_{1/2}$ that

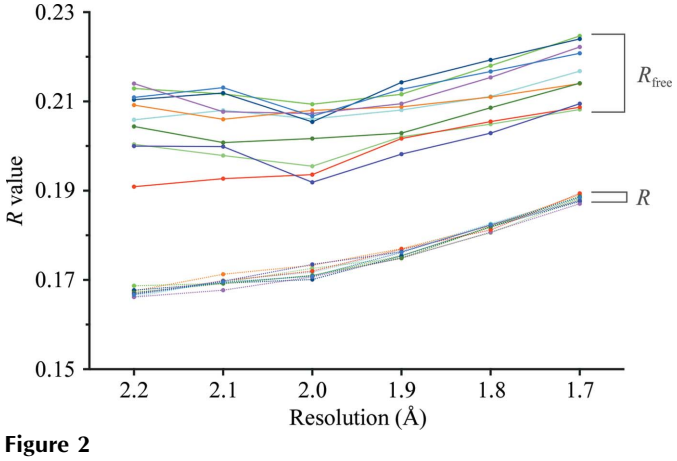

Figure 2
R and R_{free} factors for ten refinements of PDB entry 3e4f at each resolution limit with different random selections of R_{free} reflections.

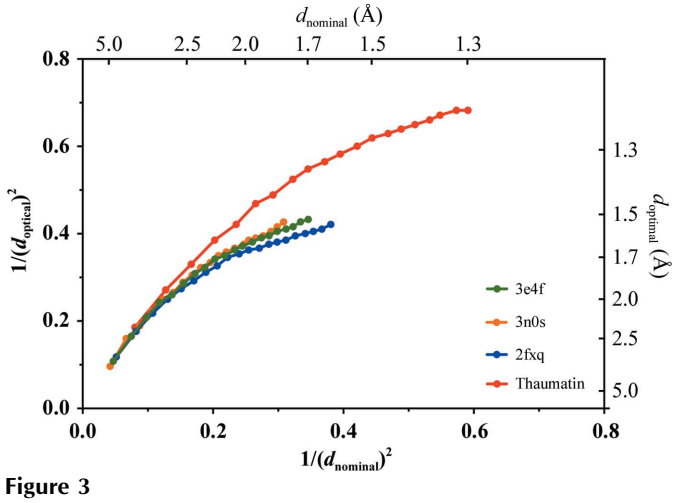

Figure 3
Relationship between the nominal resolution ($d_{nominal}$) and the optical resolution (d_{opt}) for the four analyzed data sets. The optical resolution was calculated by *SFCHECK* (Vaguine *et al.*, 1999).

Table 1
Resolution limits (Å) of the four data sets estimated according to different criteria.

		3e4f	3n0s	2fxq	Thaumatin
$d_{published}$	From original publication	2.0	2.15	1.85	1.6
$d_{reprocessed}$	Reprocessed data	1.7	1.8	1.62	1.3
d_{est}	Estimated from R and R_{free}	1.8	1.9	1.7	1.4
d_{opt}	Optical resolution (*SFCHECK*)	1.56	1.59	1.58	1.24
d_{opt}/d_{est}		0.87	0.84	0.93	0.89

would reliably suggest the appropriate resolution limit of the useful diffraction data.

3.2. Optical resolution

The parameter that is potentially useful for the evaluation of diffraction data is the optical resolution, which is available from *SFCHECK* (Vaguine *et al.*, 1999). The optical resolution estimated for different data sets is always higher than the nominal resolution, regardless of the resolution limit at which the data are truncated (Fig. 3, Table 1). However, extension of the highest nominal resolution is accompanied by a smaller extension of the optical resolution. This suggests that the inclusion of very weak highest resolution data is not very beneficial for the interpretation of fine features in electron-density maps. The optical resolution, *i.e.* the smallest distance between two peaks in the Fourier map that can still be resolved (Rupp, 2010), is estimated in *SFCHECK* as $d_{opt} = (\sigma^2_{Patt} + \sigma^2_{sph})^{1/2}$, where σ_{Patt} is the standard deviation of the Gaussian function describing the Patterson origin peak and $\sigma_{sph} \simeq 0.356 \times d_{nominal}$ is the standard deviation of the analogous peak of the Fourier transform of a sphere of radius $1/d_{nominal}$ (Vaguine *et al.*, 1999).

According to James (1948), the minimum distance between two details in the electron-density map that can still be resolved (*i.e.* the optical resolution) is $0.715 \times d_{min}$. Subsequent analysis by Stenkamp & Jensen (1984) yielded a larger distance of $0.917 \times d_{min}$. Table 1 presents values of the estimated and optical resolutions for the four investigated data sets and the ratio of the optical resolution, d_{opt}, to the limiting value of data resolution, d_{est}, estimated, as above, from the behavior of R factors. On average, the estimated resolution limits are in qualitative agreement with the analysis of Stenkamp and Jensen. This suggests that the limit of the useful diffraction data corresponds to the data resolution, reaching a value about 10–15% higher than the optical resolution ($\sim d_{opt}/0.9$).

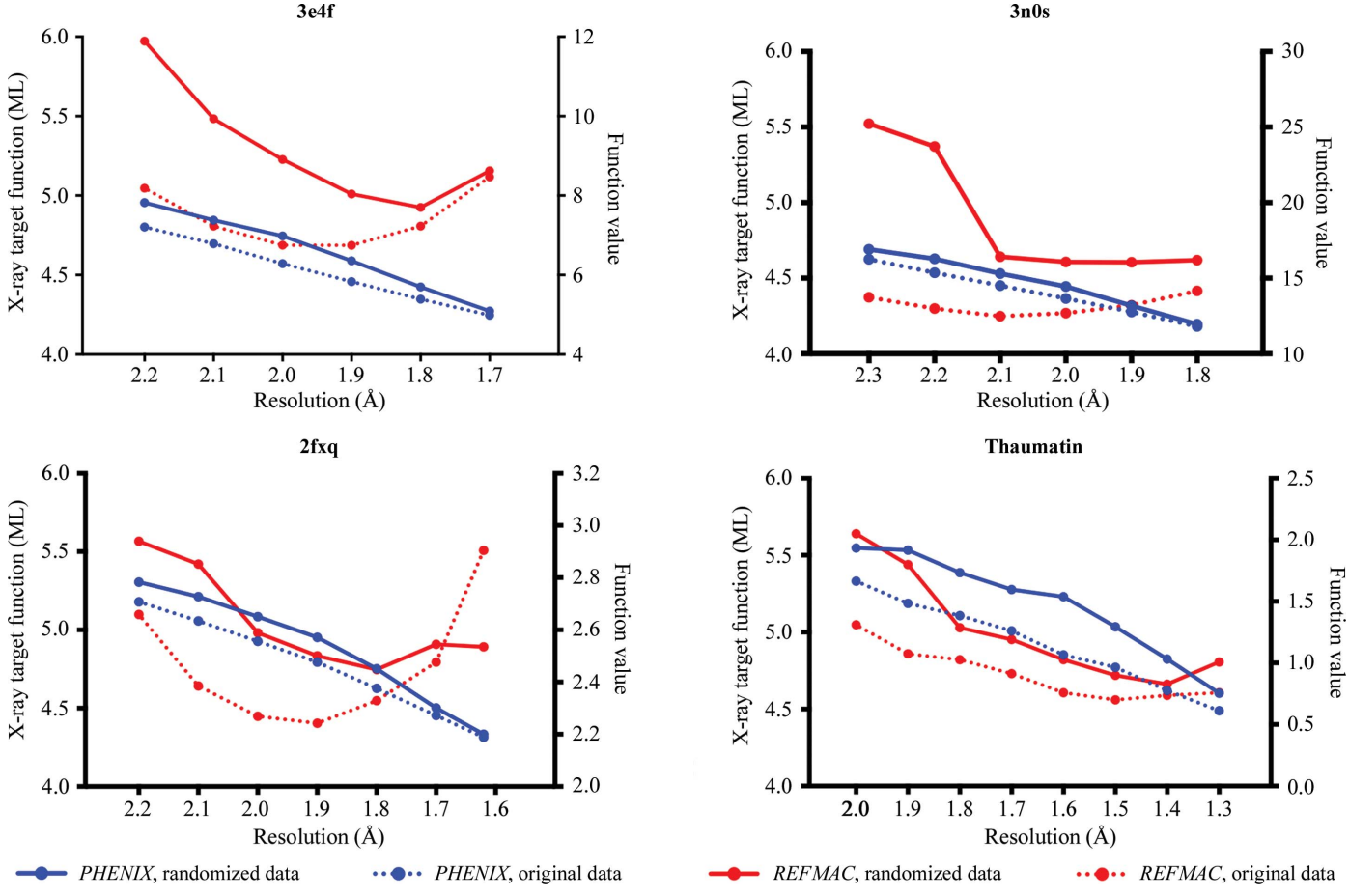

Figure 4
Values of the maximum-likelihood target function after refinement of the four investigated structures at different resolution with *phenix.refine* (blue) and *REFMAC* (red). The dashed lines correspond to the original data and the solid lines correspond to data randomized in the highest 0.1 Å resolution shell.

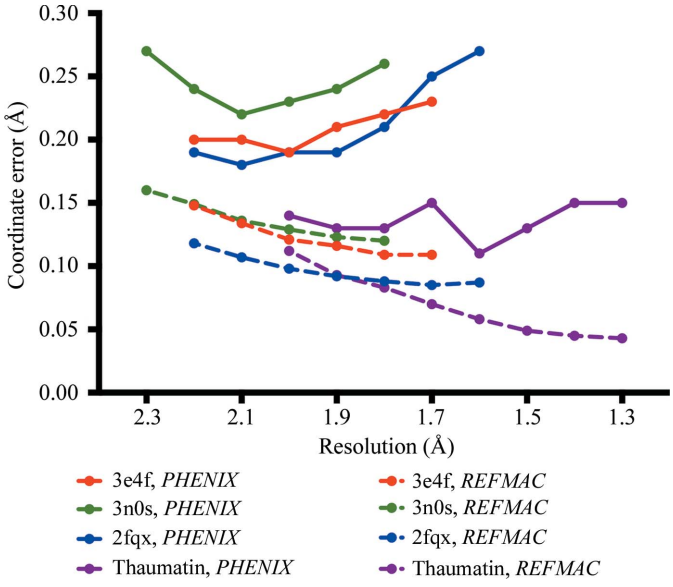

Figure 5
Overall coordinate errors based on maximum likelihood estimated from *phenix.refine* (solid lines) and *REFMAC* (dashed lines) obtained from the refinement of four structures at different data resolutions.

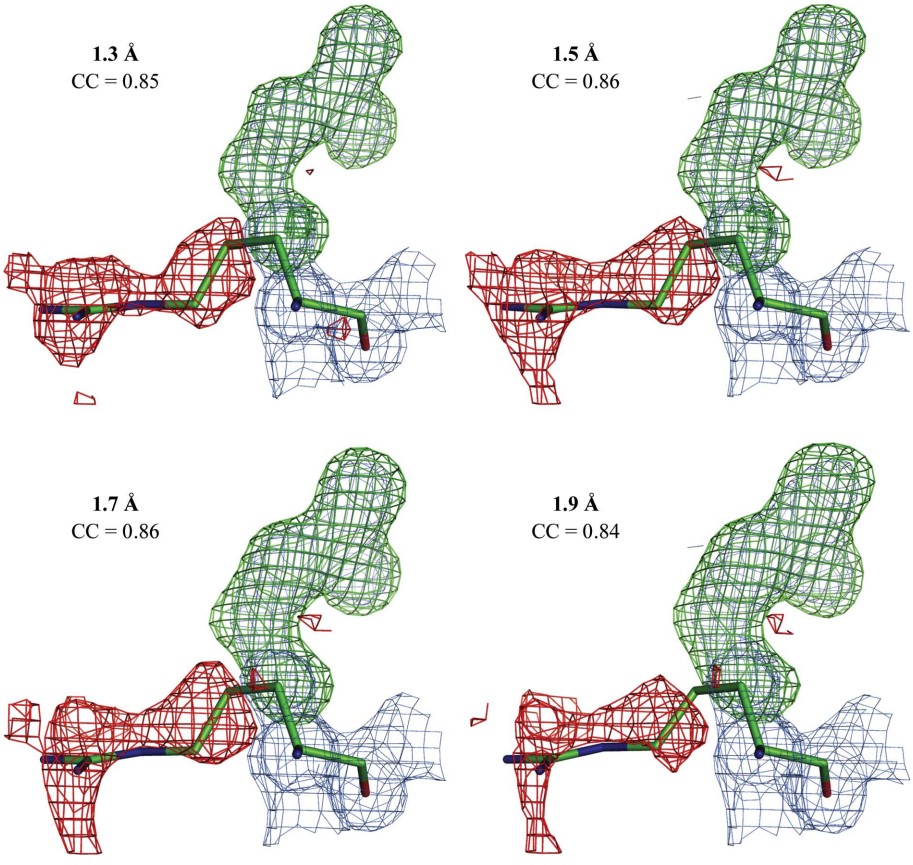

Figure 6
$2F_o - F_c$ (blue, 1.5σ contour level) and $F_o - F_c$ (green, +3.0σ contour level; red, −3.0σ contour level) maps in the vicinity of the Arg175 residue in thaumatin with its side chain deliberately moved from the correct position, after refinement to convergence with data limited to different resolutions. Correlation coefficients between these maps and the refined model around this residue are also given.

3.3. Maximum likelihood

The R and R_{free} factors are not ideal criteria for judging the process of refinement performed with contemporary programs based on minimization of maximum-likelihood (ML) targets. In practice during refinement the R-factor values often slightly increase in the last stages of the refinement process, whereas the ML function is minimized further. The values of the ML target function obtained from *phenix.refine* and *REFMAC* for the four investigated structures are presented in Fig. 4.

The behavior of the ML function is different for the two programs. The values obtained from *phenix.refine* continuously diminish with extension of resolution, but the *REFMAC* values reach a minimum and then rise for the two highest resolution refinement runs. The results from *REFMAC* may suggest that an optimum resolution limit exists and corresponds to values close to those resulting from the $I/\sigma(I) = 2.0$ criterion. However, this is not confirmed by the results obtained from *phenix.refine*. These two programs evidently utilize differently constructed maximum-likelihood target functions and it is not possible to draw generally valid conclusions about the optimum data resolution limit from the results obtained with them.

Both *phenix.refine* and *REFMAC* give an estimation of the average coordinate error based on maximum likelihood. Fig. 5 illustrates the parameters obtained for the four investigated structures refined at various resolutions. The numerical values obtained from the two programs show different behaviors. The *REFMAC* values diminish monotonically with extending resolution, whereas the *phenix.refine* values are about twice higher and initially decrease but eventually increase at higher resolution. The *REFMAC* results suggest that the inclusion of very weak data improves the accuracy of the structural models, whereas the *phenix.refine* results show that the acceptance of very weak reflections is counterproductive. The conclusions from the behavior of ML coordinate errors and ML target-function values are opposite and suggest that these criteria are not useful for estimating the data resolution limit.

3.4. Electron-density maps

Correlation coefficients between the $2F_o - F_c$ electron-density maps calculated at various resolution limits with maps calculated at the same resolution but simulated from the 'best' model refined at the highest resolution are presented in Supplementary Table S3. Refinement against data extended from

resolution d by 0.1 Å marginally improves (by about 0.1%) the correlation coefficients of maps calculated at resolution d. However, the maps obtained by refinement against data randomized in the highest resolution shell are considerably degraded.

The R factors, data and overall map correlation coefficients are global quality criteria of the structures, but do not offer helpful information regarding the modeling of local structural features that potentially differ in detail as a result of increasing data resolution. Inspection of electron-density maps obtained from refinement at various resolution limits did not permit the identification of any locations where more details could be modeled confidently. The extension of resolution and acceptance of very weak diffraction amplitudes marginally improves the global agreement between the observed and calculated amplitudes and the electron-density maps, but does not lead to the possibility of modeling more details in 'problematic' localities of the structures. This is in agreement with the observations of Evans & Murshudov (2013).

The appearance of difference maps calculated after refinement at different resolutions is presented in Fig. 6. One of the arginine side chains of thaumatin was deliberately rotated away from its correct position and the structure was refined with *phenix.refine* to convergence at various resolutions. The $F_o - F_c$ maps obtained with data extending to 1.9, 1.7, 1.5 and 1.3 Å resolution differ only marginally and, together with the $2F_o - F_c$ map, clearly identify the correct location of this side chain.

4. Conclusions

None of the existing criteria for judging the resolution limit of diffraction data are entirely reliable and completely satisfactory, and it is presently not possible to define the 'true' resolution cutoff. Among several possible parameters, such as R_{merge} and its variants, $CC_{1/2}$ and CC*, and optical resolution, the relatively most useful seems to be the signal-to-noise ratio $I/\sigma(I)$. Traditionally, the resolution cutoff used to be applied where the $I/\sigma(I)$ value dropped to 2.0, but it may be advisable to measure data to a resolution limit higher by about 0.2 Å than the thus specified boundary.

The extension of resolution beyond the traditional conservative limit does not degrade the quality of structures refined with programs based on the ML principle. The inclusion of very weak high-resolution reflections is not harmful but may be important in cases of highly anisotropic diffraction, when setting an appropriate resolution limit for the direction of stronger diffraction will result in the inclusion of very weak reflections from other regions of reciprocal space. No 'ellipsoidal' resolution cutoff would be necessary in such cases.

The extension of resolution and the inclusion of very weak reflections does not seem to be considerably beneficial in terms of the R and R_{free} factors of the refined models and does not significantly improve the interpretability of detailed features in the electron-density maps. The inclusion of the weakest reflections requires only a small effort at the stage of data collection but does not seem to be significantly beneficial for the resulting structures. It is therefore not quite a free lunch; it is a 'cheap but low-calorie meal'.

This project was supported by the Intramural Research Program of the National Institutes of Health, National Cancer Institute, Center for Cancer Research and with Federal funds from the National Cancer Institute, National Institutes of Health (contract No. HHSN261200800001). KRR is supported by a grant from the National Institute of General Medical Sciences (8 P41 GM103403-10) of the National Institutes of Health. Use of the Advanced Photon Source was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences under contract No. W-31-109-Eng-38.

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.
Caliandro, R., Carrozzini, B., Cascarano, G. L., De Caro, L., Giacovazzo, C. & Siliqi, D. (2005). *Acta Cryst.* D**61**, 556–565.
Dauter, Z. (1999). *Acta Cryst.* D**55**, 1703–1717.
DeLano, W. L. (2002). *PyMOL*. http://www.pymol.org.
Diederichs, K. & Karplus, P. A. (1997). *Nature Struct. Biol.* **4**, 269–275.
Diederichs, K. & Karplus, P. A. (2013). *Acta Cryst.* D**69**, 1215–1222.
Evans, P. (2006). *Acta Cryst.* D**62**, 72–82.
Evans, P. R. & Murshudov, G. N. (2013). *Acta Cryst.* D**69**, 1204–1214.
Frazão, C., Soares, C. M., Carrondo, M. A., Pohl, E., Dauter, Z., Wilson, K. S., Hervás, M., Navarro, J. A., De la Rosa, M. A. & Sheldrick, G. M. (1995). *Structure*, **3**, 1159–1169.
James, R. W. (1948). *Acta Cryst.* **1**, 132–134.
Jędrzejczak, R., Dauter, M., Dauter, Z., Olszewski, M., Długołęcka, A. & Kur, J. (2006). *Acta Cryst.* D**62**, 1407–1412.
Karplus, P. A. & Diederichs, K. (2012). *Science*, **336**, 1030–1033.
Klimecka, M. M., Chruszcz, M., Font, J., Skarina, T., Shumilin, I., Onopryienko, O., Porebski, P. J., Cymborowski, M., Zimmerman, M. D., Hasseman, J., Glomski, I. J., Lebioda, L., Savchenko, A., Edwards, A. & Minor, W. (2011). *J. Mol. Biol.* **410**, 411–423.
Ling, H., Boodhoo, A., Hazes, B., Cummings, M. D., Armstrong, G. D., Brunton, J. L. & Read, R. J. (1998). *Biochemistry*, **37**, 1777–1788.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Rupp, B. (2010). *Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology*, p. 453. New York: Garland Science.
Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.
Stenkamp, R. E. & Jensen, L. H. (1984). *Acta Cryst.* A**40**, 251–254.
Vaguine, A. A., Richelle, J. & Wodak, S. J. (1999). *Acta Cryst.* D**55**, 191–205.
Weiss, M. S. (2001). *J. Appl. Cryst.* **34**, 130–135.
Weiss, M. S. & Hilgenfeld, R. (1997). *J. Appl. Cryst.* **30**, 203–205.
Wlodawer, A., Minor, W., Dauter, Z. & Jaskolski, M. (2008). *FEBS J.* **275**, 1–21.
Yao, J., Woolfson, M. M., Wilson, K. S. & Dodson, E. J. (2005). *Acta Cryst.* D**61**, 1465–1475.