

## Chapter 2.2. Quality indicators in macromolecular crystallography: definitions and applications

H. M. EINSPAHR AND M. S. WEISS

### 2.2.1. Introduction

The genesis of this chapter was a perceived need for a single location in the volume in which consensus definitions could be found for the many statistical indicators of quality or figures of merit that have been developed to monitor a macromolecular crystallography (MX) experiment and its final outcome, a model. The evolving experiment has generated a rich collection of *R* values, signal-to-noise indicators, correlation coefficients and other figures of merit. As improvements in data collection, processing and other aspects of the experiment continue, we can expect a continued evolution of new or improved indicators of quality with which to monitor the impact of those improvements.

This chapter, then, attempts to provide a comprehensive list of the indicators of quality currently in use and, for each indicator in the list, a precise definition that conforms to consensus interpretations of the literature and current practice. The authors acknowledge that useful indicators may have been missed in the sweep that produced this list, but hasten to point out that the generation of new indicators is an ongoing process with the newest ones often in obscurity for a period before their utility is recognized and adopted by experimenters. There is also a subset of the indicators in the list whose members stand out as universally accepted and crucial indicators that either every experimental description must include or that every experimenter should be familiar with. A summary of these is given at the end of this article in Table 2.2.11.1. The authors also acknowledge that there may be competing definitions for some indicators. Where they occur, these competing definitions will be pointed out and a discussion will be provided which, at a minimum, will attempt to clarify the differences, but will also, where possible, attempt to arbitrate or discriminate among opposing views. It should be noted that the scope of this chapter is primarily concerned with the crystallographic experiment. While quality indicators useful for model refinement are given in Sections 2.2.8 to 2.2.10, the validation of the refined model is covered more extensively in Part 21 of this volume.

Before proceeding with the quality indicators themselves, it is necessary to discuss a small collection of parameters whose values, typically left to the experimenter to fix, impact on virtually all the indicators of quality we present here. The reason for this wide impact is that they are used to determine which reflections are included in the final data set in question. In their simplest form, the effect is a limit, applied as a cutoff in intensity or resolution, beyond which reflections are excluded from consideration.

When a cutoff is applied based on intensities, the limiting value may be simply a fixed number, for example zero, or it may vary from reflection to reflection, for example some multiple of the standard uncertainty of the reflection intensity. Reflections with intensities below limiting values are excluded. Reflections excluded in this manner are often referred to as unobserved.

When the objective is to estimate electron-density distributions, the exclusion of intensity terms below very low limiting

values is unlikely to have any significant negative impact and may offer a positive savings in computation time. This may not be the case, however, when the objective is refinement. Inclusion of reflections of low intensity, even zero, may have important positive results on the quality of the final model. True, some of these intensities may be poorly known or, if less than zero, physically unrealistic, but given the excellent quality of modern data-collection instrumentation and techniques, few intensities, if any, are without some reasonable estimate of their standard uncertainty, so that weighting procedures can be applied to modulate the impact of individual terms in the final result and judgments can be made about exclusion of intensity values that are truly improbable. With the application of proper weighting procedures, there seems to be little justification for exclusion of reflections from consideration based on intensity (except for the truly improbable mentioned above) when computing indicators of quality. As instrumentation and techniques continue to improve, this is a topic that merits continued attention and debate within the MX community in search of consensus best practices for handling weak reflections and for including them in estimators of quality.

While an acceptable estimate of the nominal resolution of a diffraction data set is widely considered to be a high-value indicator of data quality, assignment of a value to this limit is typically left to the experimenter and is thus prone to subjectivity. Some guidelines for estimation have emerged. One would set the nominal resolution based on the percentage of weak reflections above the limit so that reflections with intensities above that limit would be included. An example might be the resolution at which 70% of unique reflections have intensities above zero or above some multiple of their standard uncertainties. Another way to estimate the nominal resolution which has gained wide acceptance is based on the overall signal-to-noise value where, in its most popular expression, the limit value is the resolution at which the mean signal-to-noise ratio in the outer resolution shell falls to 2. Each of these estimation methods is susceptible to convolution with limits based on intensity leading to, for example, a limit set as the resolution at which the mean signal-to-noise ratio 'of observed reflections' falls to 2. The guidelines for estimating resolution limits – the methods to be applied, the constraining values such as 70% or 2, and the proper integration with limits applied based on intensity – also merit further attention and discussion by the MX community, the goal being definition of consensus best practices. We take this opportunity to suggest that the widely accepted estimate, the resolution at which the mean signal-to-noise in the outer resolution shell falls to 2, calculated without imposition of a cutoff in intensity, has much to recommend it. True, some susceptibility to subjectivity remains in the definition of resolution-shell ranges and limits, but the impact is minimal and hardly worth the effort. On the other hand, modifying the indicator to remove the shells so that the mean signal-to-noise ratio applied to all data above an appropriately adjusted limit value would extinguish that source of subjectivity.

When a cutoff is applied based on nominal resolution, then considerations similar to those for intensity-based cutoffs apply.

## 2.2. QUALITY INDICATORS

If the objective is to estimate electron-density distributions, exclusion of large numbers of weak intensity terms beyond a limiting resolution is unlikely to be of any significant impact except, perhaps, a favourable one in computation time. If a sparse population of more intense reflections is also excluded, the effect may also be positive by reducing aberrations that might interfere with interpretation. If the objective is refinement, the benefit is less clear, except possibly in computation time. The cost for that may be the exclusion of a few more intense terms that provide positive guidance to refinement. Another positive effect of a sharply defined limiting resolution might be improved estimation of the spherical interference function where it is needed. It appears that, in practice, the consensus is that, in applying a cutoff, the exclusion of a few intense terms is at worst of negligible impact overall.

Because many imposed limits that exclude data are dependent on estimates of standard uncertainty, procedures for estimation of standard uncertainty need to be considered. As suggested earlier, a diffraction data set without a standard uncertainty for each intensity measurement is certainly the exception in current practice. The methods used by data-processing programs to estimate these standard uncertainties may be difficult, even impossible, to discern, but at core they must all be based on counting statistics. With the continuing trend toward data sets of high multiplicity (also referred to as redundancy), however, estimates of standard uncertainties are available from distributions of replicate measurements about means. Both of these estimates have value and express impacts of somewhat different sources of error. It would seem therefore that, where applicable, the best way to accommodate both would be to calculate weighted average values of individual intensities, where the weights are derived from the standard uncertainties from data processing, and then to estimate the standard uncertainties for the weighted averages by application of standard propagation-of-error procedures. This, finally, is the third area we highlight here that deserves focused discussion within the MX community with the objective of defining consensus best practices.

Little remains to be included in these introductory preliminaries. It should be understood that most of these indicators of quality may be cast in terms of either structure factors or intensities. We draw little attention to this as we define individual indicators, except that, where one of the two possibilities seems to dominate in usage, we tend to define that form alone. It is also true that many of these indicators have counterparts in which individual reflection terms are weighted. In current practice, these forms find little use in application to biological macromolecules and we ignore them here. Finally, many of these indicators may be expressed as fractions, that is, as numbers between one and zero. They are also often expressed in the percentage form, that is, as numbers between one hundred and zero. While we confine our use to the former here, we make no statement of preference.

The remainder of this chapter consists primarily of individual sections that reflect the various steps from the crystallographic data-collection experiment to the refinement of the final model. Relevant quality indicators and definitions are given for each of the steps and the most commonly used are collected in Table 2.2.11.1.

### 2.2.2. Quality indicators for diffraction data

Once useful crystalline samples have been obtained, the collection of X-ray diffraction data is the next (and the last) experi-

mental step in a structure determination. Although the greatest care may be taken to collect data of as high quality as possible, there remain circumstances and influences that limit the quality of the data. Over time, many indicators have been defined to describe various aspects of diffraction data quality. The most important ones are discussed here.

**Nominal resolution,  $d_{\min}$ .** The resolution of a diffraction data set describes the extent of measurable data and is calculated by Bragg's law [equation (2.2.2.1)] based on the maximum Bragg angle  $2\theta$  included in the data set for a given data-collection wavelength  $\lambda$ . As discussed above, the nominal resolution is a limit set by the experimenters and is well known to be prone to subjective judgment. A number of suggestions have been made to reduce the subjectivity associated with this limit. One defines the limit as the resolution within which the intensities of a fraction of the unique reflections, for example 70%, are above a threshold, for example zero or three times their standard uncertainties. Another suggested limiting criterion, discussed further below, recommends that the nominal resolution be set as the midpoint of the resolution range of the shell at which the mean signal-to-noise ratio falls below 2.

$$d_{\min} = \lambda/2 \sin \theta. \quad (2.2.2.1)$$

**True resolution,  $d_{\text{true}}$ .** The true resolution of a diffraction data set is defined as the minimum distance between two objects in a crystal that permits their images in the resultant electron-density map to be resolved. Often,  $d_{\text{true}}$  is approximated as  $d_{\min}$ .

To illustrate this crucial distance, represent two equivalent atoms by equal overlapping Gaussians. One might then consider that the distance between them that just permits distinguishing them as individual atoms might be the distance at which the electron-density value at the midpoint between the atoms drops to a value just below that at the positions of the atoms. For a normal distribution, this distance is  $2\sigma$ , twice the standard deviation of the distribution.

Another perspective is provided by the realization that, when a Fourier synthesis is terminated at a resolution cutoff  $d_{\min}$ , successive spheres of negative and positive density of decreasing amplitude surround the maxima of positive density at atomic positions in that synthesis. It has been shown that the distance from the centre of such a maximum to the first zero is  $0.715d_{\min}$  (James, 1948), which is a useful estimate of the limiting distance between distinguishable features in the electron-density map or  $d_{\text{true}}$ . Similar estimates are used in other areas, notably for defining resolution in astronomy. A more recent re-evaluation suggests that a limit of  $0.917d_{\min}$  is a better value, especially when the effects of form factors and atomic displacement parameters are considered (Stenkamp & Jensen, 1984). Add to that the effects of errors in experimental amplitudes and derived phases, and the approximation of  $d_{\text{true}}$  as  $d_{\min}$  seems quite reasonable.

**Optical resolution,  $d_{\text{opt}}$ .** The optical resolution  $d_{\text{opt}}$  is calculated from the standard deviation of a Gaussian fitted to the origin peak of the Patterson function ( $\sigma_{\text{Patt}}$ ) of the diffraction data set and the standard deviation of another Gaussian fitted to the origin peak of the spherical interference function ( $\sigma_{\text{sph}}$ ). This definition is based on Vaguine *et al.* (1999) and is implemented in the program *SFCHECK*. The optical resolution is intended to account for uncertainties in the data, atomic displacement factors, effects of crystal quality and series-termination effects by means of a propagation-of-error-like approach (Blundell & Johnson, 1976; Vaguine *et al.*, 1999). It has been suggested that  $d_{\text{opt}}$  is a better approximation of  $d_{\text{true}}$  than