at a minimum, Ago2, the enzyme Dicer, and a double-stranded RNA binding protein such as TRBP (see the figure). Other proteins, such as the chaperone complex Hsc70-Hsp90 (*3*) or C3PO (*4*), are also recruited by Ago2 and accelerate RISC-loading or increase turnover rates, respectively. Most notably, Ago2 recruits glycine-tryptophan (GW) proteins, the key players of miRNA-mediated translational repression and mRNA degradation, to the mRNA transcript.

How are Ago proteins able to recruit such an impressive number of binding partners? How does the cell regulate these events? These key questions can now be addressed more directly based on the crystal structures of human Ago2 and *Kp*Ago as well as previous work on Ago homologs. Past crystal structures of archaeal and bacterial Argonaute proteins (*5*–*7*), together with crystal structures of eukaryotic Ago2 domains (*8*, *9*), revealed two lobes in the protein's architecture and a multidomain conformation (see the figure). Similar to these earlier structures, human Ago2 has four domains in which the N-terminal and PAZ domains form one lobe and the MID and PIWI domains form the second lobe. The PAZ domain binds the 3′ end of the guide RNA, while the MID domain provides a binding pocket for the 5′-terminal phosphate group of the guide. Furthermore, both biochemical and structural studies have revealed a crucial role for the PIWI domain in slicing. The crystal structure of *Kp*Ago shows that the slicing activity is regulated by conformational changes of Ago upon guide loading. This rearrangement leads to formation of a catalytic tetrad in the enzyme's active site and allows endonucleolytic activity. Interestingly, a similar and presumably induced rearrangement of the active site can be seen in the structure of human Ago2, suggesting this as a regulatory mechanism for RISC specificity.

Although their overall architectures are similar to those of their archaeal and bacterial counterparts, the human Ago2 and *Kp*Ago crystal structures are a key step toward a complete structural and mechanistic understanding of RNAi pathways. These crystal structures open the way for designing synthetic RNAs or potent small molecules as new drugs for efficient RNAi-based therapies. It will now be possible to map mutations and interaction surfaces onto protein structures with established biochemical functions. The first hints of the insights that will be forthcoming are from experiments by Schirle and MacRae to elucidate the nature of human Ago2 binding to tryptophan, in an effort to mimic binding to tryptophan-rich GW proteins. The interaction between Ago2 and GW proteins, which is essential for efficient and robust mRNA silencing (*10*, *11*), is an attractive target for potential manipulation of gene silencing efficiency. For example, it might be possible to design therapeutic peptides or small molecules to compete with GW protein binding and thereby inhibit downstream processes. Future research can now build upon a solid structural foundation to define the molecular mechanisms that enable cells to use RNA for the control of protein expression levels.

References

1. N. T. Schirle, I. J. MacRae, *Science* **336**, 1037 (2012).
2. K. Nakanishi, D. E. Weinberg, D. P. Bartel, D. J. Patel, *Nature* (2012); 10.1038.nature11211.
3. S. Iwasaki *et al.*, *Mol. Cell* **39**, 292 (2010).
4. X. Ye *et al.*, *Nat. Struct. Mol. Biol.* **18**, 650 (2011).
5. J. J. Song *et al.*, *Science* **305**, 1434 (2004).
6. Y. Wang *et al.*, *Nature* **456**, 921 (2008).
7. Y. R. Yuan *et al.*, *Mol. Cell* **19**, 405 (2005).
8. A. Boland *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 10466 (2011).
9. F. Frank *et al.*, *Nature* **465**, 818 (2010).
10. J. E. Braun *et al.*, *Mol. Cell* **44**, 120 (2011).
11. M. R. Fabian *et al.*, *Nat. Struct. Mol. Biol.* **18**, 1211 (2011).

10.1126/science.1223549

BIOCHEMISTRY

Resolving Some Old Problems in Protein Crystallography

Two methods improve the quality and ease of structural modeling by showing how to include diffraction data that are often thrown away.

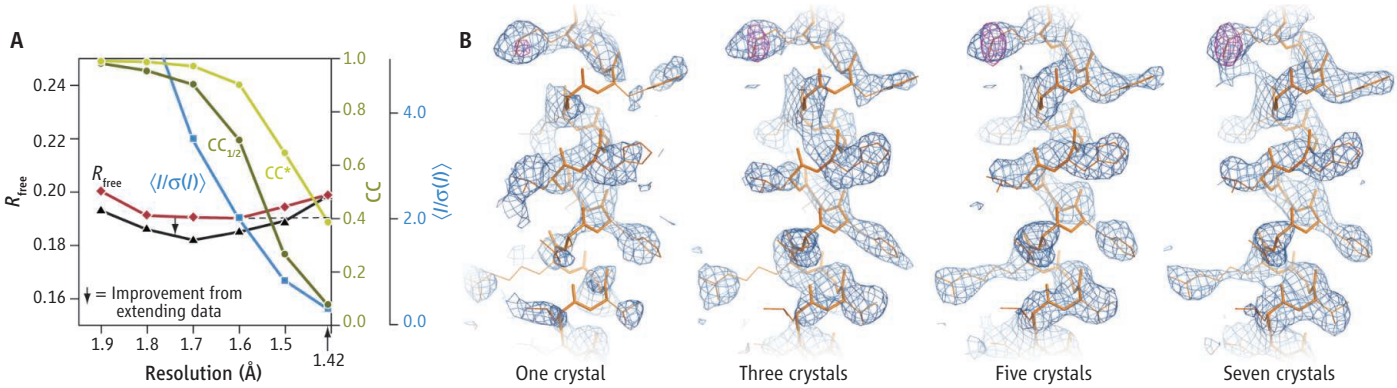
Phil Evans

Scientific conclusions should be supported by the observed data. However, in x-ray crystallography, the raw diffraction data are rather remotely connected to the final coordinates of the molecule because the experimental data undergo a Fourier transform during the analysis. Thus, any individual feature of the structural model—where a particular atom is located—depends on all of the measured diffraction intensities. Also, the phase information essential for this reconstruction is lost in the experiment (the "phase problem"). Although the coordinate model is repeatedly tested against the data in the course of structure solution, it is common practice to choose what data to use early in the process. Two papers in this issue suggest in different ways that crystallographers have often been excluding useful data from structure determination. On page1030, Karplus and Diederichs (*1*) show that the "resolution" of data sets is frequently underestimated, so that the final model is not as good as it could be. On page 1033, Liu *et al.* (*2*) show that averaging data from multiple crystals can give helpful information for solving the phase problem by using intrinsic sulfur atoms in the protein, circumventing the need to introduce heavier atoms.

The high-resolution diffraction data are found at high scattering angles and give the precise atomic positions in a structure, but in these outer regions of the pattern, the intensity I of diffraction fades away into the background noise because of various disorder effects. A high-resolution cutoff is typically applied to the data, and ideally should be at the point where adding more observations would not add significant information. Unfortunately, this point is hard to estimate.

An obvious criterion is the average signal-to-noise $\langle I/\sigma(I)\rangle$ as a function of resolution. Estimates of the noise term $\sigma(I)$ are generally rather unreliable, but even so, this guide for applying a cutoff is often used. However, another commonly used criterion based on a measure of internal consistency, R_{merge}, is particularly ill-suited to this purpose, despite its widespread use. Diederichs and Karplus point out that R_{merge} cannot be compared with the crystallographic R-factor used to compare observed and calculated data (R_{cryst} or R_{free}), because as the relative error increases with higher resolution, R_{merge} tends to infinity, whereas R_{cryst} tends to a constant value. R_{merge}, or its improved multiplicity-weighted cousin R_{meas} [which the same authors and others introduced some time ago (*3*, *4*)], can be useful as a general guide and for comparison between different data sets, but are not suit-

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, UK. E-mail: pre@mrc-lmb.cam.ac.uk



Waste not, want not. Two types of crystallographic data normally thrown away can be put to use. (**A**) Karplus and Diederichs showed that as the resolution of the data is extended, the overall fit between model and observations improved, as measured by R_{free} from paired refinements extending the resolution in steps of 0.1 Å . The red line is R_{free} before extension, and the black line is after extension. The improvement extends to the maximum resolution of 1.42 Å, even though the measures of signal/noise (blue line; dashed line indicates $\langle I/\sigma(I)\rangle = 2.0$) and of internal consistency ($CC_{1/2}$ and CC*, green lines) are getting worse (scale bars on the right). (**B**) Liu *et al*. demonstrate improvement in electron density using phases from S anomalous scattering from increasing numbers of crystals of the protein TorS. Purple contours show Bijvoet-difference (anomalous) density for a methionine S. Thirteen crystals were used, and at least seven were needed to locate the S atoms, so the maps shown for one, three, and five crystals used information only available by using seven or more crystals.

able metrics for deciding the "real resolution."

A better and more reliable measure of internal consistency is the correlation coefficient between two data sets generated by randomly splitting the data into halves, $CC_{1/2}$. Diederichs and Karplus make a strong case that $CC_{1/2}$ provides a good criterion for deciding where to apply a resolution cutoff. Most importantly, they show that the improvement in the fit of the model to the data by adding another shell of data correlates well with $CC_{1/2}$ remaining significantly above zero, based on a series of paired refinements that extend the resolution in steps (see the figure, panel A).

The use of half–data set correlation measures is not particularly new: The idea of splitting observations in half is a very old one in statistics; a similar measure (Fourier shell correlation) has been used in single-particle electron microscopy since the early 1980s [see, for example, (*5*, *6*)]; and its use in x-ray crystallography has been suggested both for assessing anomalous differences [within or between data sets (*7*)] and for the present purpose (*8*, *9*). However, it is encouraging to see here a direct link between data quality and model quality, and CC*, the estimate of the "true" CC derived from $CC_{1/2}$, can be compared directly to a correlation coefficient with data calculated from the model.

The lesson for the crystallographic community is that we should not prematurely exclude too much high-resolution data. Anecdotal evidence suggests that including higher-resolution data can make automated model-building methods work better. There is also a lesson for referees not to complain on the basis of R_{merge}, about authors "overstating the resolution." For the wider community of structure users, we need to wean ourselves from interpreting the nominal resolution of a crystallographic data set as a single number representing the quality and reliability of a structure. Resolution tells us how many observations were used. Although higher resolution will in general be better, the answers we want from a structure are about local conformations, and their correctness cannot be indicated by global scores—neither measures of internal consistency of the data nor of the overall fit of the model to the data.

Hendrickson's group [Liu *et al*. (*2*)] describe an improved approach to the phase problem that uses data from multiple crystals to solve structures, where the phase information comes from the anomalous diffraction from intrinsic sulfur (S) atoms. In the absence of a related model to use for "molecular replacement," the crystallographic phase problem for macromolecules is generally solved by using a few marker atoms heavier than carbon, nitrogen, or oxygen, and measuring their effect on the diffraction pattern, often from their so-called anomalous diffraction. The most common method used at present relies on replacing the native amino acid methionine with engineered selenomethionine, a method introduced earlier by Hendrickson and co-workers (*10*).

In most unmodified proteins, the heaviest atoms are S, but rather few structures have been solved by S anomalous scattering because the very small signal is easily swamped by noise. Although the accuracy of measuring this signal can be improved by measuring several crystals and averaging the results, crystallographers have often shied away from combining data from many crystals because of worries that they may not be the same (nonisomorphous). Liu *et al*. now show that with careful data collection (longish wavelength and avoidance of radiation damage) and cluster analysis to eliminate nonconforming crystals, even quite large proteins with poorly diffracting crystals can be solved just by using the S anomalous signal (see the figure, panel B). This result is of practical importance because introducing traditional heavy atoms, including selenomethionine, can be difficult, whereas almost all proteins contain a reasonable number of S atoms (nucleic acids contain phosphorus, which is equally good). It is commonplace to collect many data sets from equivalent crystals, so combining all of the data together to help the structure solution is worth trying.

Referees and authors alike will need to assess carefully the quality of a structure determination when these two methods are used because the inclusion of weaker, noisier data can give better structures only when used in a statistically robust way. For this reason, it is desirable that referees should be given access to the unmerged diffraction intensities (and preferably the diffraction images) and that these data should be deposited in public databases.

References

1. P. A. Karplus, K. Diederichs, *Science* **336**, 1030 (2012).
2. Q. Liu *et al*., *Science* **336**, 1033 (2012).
3. K. Diederichs, P. A. Karplus, *Nat. Struct. Biol.* **4**, 269 (1997).
4. M. S. Weiss, *J. Appl. Cryst.* **34**, 130 (2001).
5. P. B. Rosenthal, R. Henderson, *J. Mol. Biol.* **333**, 721 (2003).
6. R. Henderson *et al*., *Structure* **20**, 205 (2012).
7. T. R. Schneider, G. M. Sheldrick, *Acta Crystallogr.* **D58**, 1772 (2002).
8. P. R. Evans, *Acta Crystallogr.* **D62**, 72 (2006).
9. P. R. Evans, *Acta Crystallogr.* **D67**, 282 (2011).
10. W. A. Hendrickson, J. R. Horton, D. M. LeMaster, *EMBO J.* **9**, 1665 (1990).

10.1126/science.1222162

Published by AAAS