Optimizing data collection for structure determination

Ana González

Stanford Synchrotron Radiation Laboratory, 2575 Sand Hill Road, MS 99, Menlo Park, CA 94025, USA

Correspondence e-mail: ana@slac.stanford.edu

The ultimate purpose of diffraction data collection is to produce a data set which will result in the required structural information about the molecule of interest. This usually entails collecting a complete and accurate set of reflection intensities to as high a resolution as possible. In practice, the characteristics of the crystal and properties of the X-ray source can be limiting factors to the data-set quality that can be achieved and a careful strategy has to be used to extract the maximum amount of information from the data within the experimental constraints. In the particular case of data intended for phasing using anomalous dispersion, the synchrotron beamline properties are relevant to determine how many wavelengths (one or more) should be used for the experiment and what the wavelength values should be. This will in turn affect the detailed strategy for data collection, including decisions about the data-collection sequence and how much data to collect at each wavelength. Collection of multiwavelength anomalous dispersion (MAD) data at three different wavelengths can provide very accurate experimental phases. Two-wavelength MAD experiments may offer the best compromise between phase quality and minimizing the effects of radiation damage to the sample. However, MAD experiments are demanding in terms of beamline wavelength range, easy tunability, stability and reproducibility. When the beamline cannot fulfill these demands, single-wavelength experiments may be a better option.

1. Introduction

The most commonly used methods to obtain experimental phases are the following.

(i) Isomorphous replacement, developed by Green *et al.* (1954). The phasing information is extracted from the differences in diffracted intensities between the native protein structure and one or several heavy-atom derivatives. Good isomorphism between the native and derivative crystal forms is essential for the success of IR experiments.

(ii) Multi- or single-wavelength anomalous dispersion (MAD or SAD). This method exploits the differences in diffracted intensities arising from changes in the scattering properties of heavy atoms as a function of incident X-ray wavelength or energy around their absorption edges (James, 1948). The variation in scattering properties results both in differences between reflection intensities at different wavelengths (known as dispersive differences) as well as differ-

ences between Friedel-related intensities at the same wavelength (anomalous, Friedel or Bijvoet differences).[1]

(iii) Direct methods, based on direct analysis of the structure-factor amplitudes (reviewed by Usón & Sheldrick, 1999). Although direct methods are routinely used to extract initial phase information (*i.e.* the heavy-atom sites) necessary to proceed with anomalous dispersion or isomorphous replacement phasing, their application as a stand-alone method to solve protein structures is limited mainly by insufficient high-resolution diffraction from many crystals.

Anomalous dispersion is currently the fastest growing phasing method (Hendrickson, 1999; Ealick, 2000). Although the theoretical foundations of anomalous dispersion phasing were established even before those for isomorphous replacement (Bijvoet, 1949), the practical application of the method to macromolecular structure solution is a more recent development (Guss *et al.*, 1988; Hendrickson, 1991). The reason for this time delay is the difficulty of the experiment from the technical point of view. While isomorphous replacement does not require a particularly sophisticated experimental setup, the development of adequate synchrotron beamline instrumentation was critical in exploiting the anomalous dispersion method to its full potential. The advances in beamline design make it possible in principle to solve new structures from a single crystal in a single experiment with no non-isomorphism problems.

Recent efforts to use anomalous differences measured at a single wavelength have proved that SAD phasing using solvent flattening (Wang, 1985) or direct methods (Fan *et al.*, 1984) to resolve the SAD phase ambiguity can be successful at solving macromolecular structures (Dauter *et al.*, 2002). SAD phasing is especially useful when the source or sample characteristics make multiwavelength experiments impractical and when lack of isomorphism rules out isomorphous replacement phasing.

Although dedicated synchrotron-radiation beamlines make it possible to maximize the anomalous signal in the data, the anomalous contribution to the total scattering factor $f(\lambda) = f_0 + f^A(\lambda)$, where $f^A(\lambda) = f'(\lambda) + if''(\lambda)$ is intrinsically small compared with the elastic scattering contribution f_0. Collection of redundant data at different wavelengths can provide the desired accuracy, but it can result in an experiment several times longer than is required to collect a single complete data set. A longer data collection will result in a higher radiation dose absorbed by the sample and an increased probability of significant radiation damage. It has been demonstrated that ionizing radiation causes an expansion of the unit-cell dimensions (Ravelli *et al.*, 2002; Murray & Garman, 2002) and can induce specific structural changes (Burmeister, 2000; Ravelli & McSweeney, 2000) even before the intensity of the diffracted intensities is significantly reduced. These effects result in a loss of isomorphism during the experiment. If the magnitude of the resultant errors in the calculated anomalous and dispersive differences is larger than the differences themselves it may become impossible to solve the structure, despite the use of cryogenic techniques (Garman, 1999). To reduce the total dose absorbed by the crystal during the experiment, we must make a careful choice of strategy aiming at acquiring a maximum of phasing information with a minimum amount of data. The optimal strategy should be based on the anomalous scattering properties of the sample and the characteristics of the beamline where the experiment is to be carried out.

2. Synchrotron beamline characteristics

The characteristics of the synchrotron beamline will determine to a large extent whether a particular data-collection strategy will be feasible or likely to succeed. The most relevant properties from the point of view of experimental phasing are the effective wavelength range of the beamline, the bandpass and the stability and reproducibility of the selected wavelength.

2.1. Wavelength range

For phasing biological macromolecules, the wavelength spectrum of the incident X-ray beam should ideally be tunable to cover the absorption edges of most common heavy elements naturally present or easily introduced into proteins. This allows optimization of the anomalous differences, which are related to the value of the imaginary component of the anomalous scattering factor f'', which attains a local maximum just above the absorption edge. Additionally, data collection at an energy a couple of keV away from the absorption edge[1] would optimize the dispersive differences between reflections measured at different wavelengths. The dispersive differences are proportional to the difference in the real component $\Delta f'$; f' is minimum (most negative) at the inflection point of the edge and rises to its maximum value away from the edge. Maximizing f'' and $\Delta f'$ achieves a maximum theoretical separation between the centers of the phasing circles and more accurate phases (Hendrickson, 1985).

In practice, there is a limit for wavelengths longer than about 2 Å imposed by the difficulty of carrying out routine macromolecular crystallography experiments in this spectral region. The problems encountered arise both from the lower X-ray beam flux resulting from increased absorption of the X-rays from beamline elements, mainly the beryllium windows used to insulate vacuum sections, and errors in the diffracted intensities arising from absorption by the sample itself. Although some anomalous dispersion experiments are feasible at long wavelengths (Lehmann *et al.*, 1993; Kahn *et al.*,

---

[1] The energy E and wavelength λ of light are related by the expression $E = hc/\lambda$, where h is the Planck constant and c is the speed of light. Macromolecular crystallographers traditionally characterize X-rays in terms of their wavelength in Å. On the other hand, scattering factors are commonly tabulated as a function of energy in eV or keV. Moreover, because of the inverse relationship between E and λ, fixed increments or intervals of energy will have very different values in Å over the full spectrum, making it inconvenient to use wavelengths on these occasions. Therefore, in some instances throughout this paper energy units will be used when convenient, although most of the discussion will be in terms of wavelength. The formula $\lambda = 12398/E$ can be used to translate the energy in eV to the corresponding wavelength in Å.

2000), they require a specialized instrumentation setup: mounting the sample in an enclosure filled with helium to decrease air absorption, use of image-plate detectors, which are more sensitive to softer X-rays than CCDs, and a cylindrical detector surface in order to be able to measure high-angle reflections (Kahn *et al.*, 2000). These limitations make it difficult to use the full anomalous signal from lighter heavy atoms without absorption edges in the hard X-ray region (such as S or P) for structure solution. For these experiments, highly redundant data collection at a single wavelength below 2 Å would be a reasonable strategy (Dauter & Adamiak, 2001; Dauter *et al.*, 1999).

Most dedicated MAD beamlines can access short wavelengths down to 0.8–0.7 Å. The limit is imposed by the critical energy of the source or by the use of a mirror to absorb shorter-wavelength X-rays with the aim of protecting other beamline elements from heat overload. Some beamlines, especially at high-energy sources in third-generation synchrotrons, can reach shorter wavelengths. Sometimes the beamline X-ray wavelength range falls between the *L* and *K* absorption edges of interest for a particular experiment so that neither edge can be accessed. This can happen often for elements with atomic numbers between 41 and 56, including heavy atoms useful for derivatization, such as Xe, I *etc*. In this case, tunability over a large wavelength range is still advantageous, since a single-wavelength experiment can be carried out at a wavelength providing an optimal compromise between a high f'' value and small absorption errors (Panjikar & Tucker, 2002).

Absorption edges of all elements with atomic number between 25 and 39 and between 59 and 92 are accessible at a typical dedicated beamline tunable between 2.0 and 0.7 Å, allowing multiwavelength experiments with optimized anomalous and dispersive differences.[2] The range above includes a high proportion of common heavy atoms present in proteins (see Table 1).

2.2. Wavelength resolution, stability and reproducibility

The natural absorption edge width $(\Delta\lambda/\lambda)_N$ is between 1.5 and 3×10^{-4} for the edges covered at typical crystallography beamlines (Graber *et al.*, 1998). The observable width of the edge and any 'white-line' features near it will be given by the

Table 1
Absorption edges available at a beamline tunable between 2 and 0.7 Å for some common heavy atoms; theoretical f'' above the edge.

In the cases where a white line is often present, this can significantly increase the f'' value.

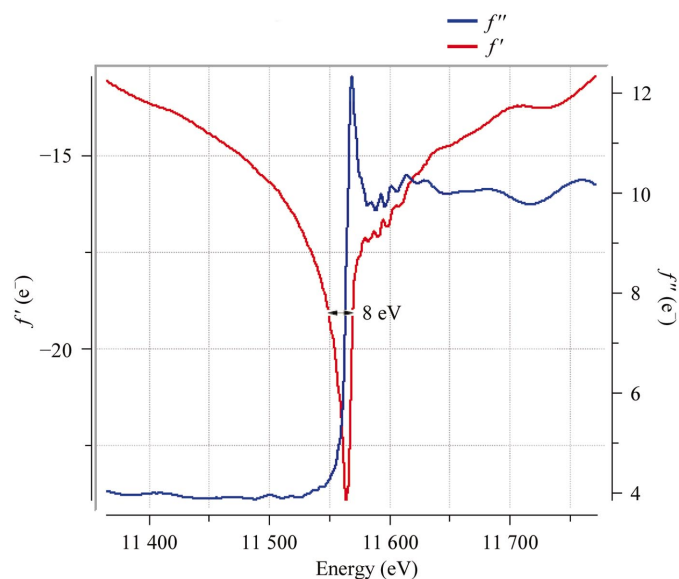| Element | | *K* | | | L_{III} | | | L_{II} | | | L_I | | |
No.	Symbol	*E* (eV)	λ (Å)	f'' (e⁻)	*E* (eV)	λ (Å)	f'' (e⁻)	*E* (eV)	λ (Å)	f'' (e⁻)	*E* (eV)	λ (Å)	f'' (e⁻)
26	Fe	7112	1.7433	4.0									
27	Co	7709	1.6083	3.9									
29	Cu	8979	1.3808	3.9									
30	Zn	9659	1.2836	3.9									
34	Se	12658	0.9795	3.8									
35	Br	13474	0.9201	3.8									
36	Kr	14326	0.8654	3.8									
62	Sm							7312	1.6956	12.4	7737	1.6024	12.8
78	Pt				11564	1.0721	10.2	13273	0.9341	11.3	13880	0.8932	12.0
79	Au				11919	1.0402	10.2	13734	0.9027	11.2	14353	0.8638	12.0
80	Hg				12284	1.0093	10.2	14209	0.8725	11.1	14839	0.8355	11.9
82	Pb				13035	0.9511	10.1	15200	0.8157	10.9	15861	0.7817	11.8

Figure 1
Absorption edge from a Pt-containing crystal measured at the SSRL wiggler beamline 9-2. Although the natural width of the edge is less than 1.5 eV, the energy dispersion of the beam [determined primarily by the bandpass of the Si(111) monochromator] is responsible for the observed edge width.

convolution of the natural edge width and the wavelength spread of the monochromatic X-ray beam, determined primarily by the bandpass of the monochromator. In order to avoid experimental broadening of the edge, an instrumental $(\Delta\lambda/\lambda)_I$ much smaller than the natural width of the edge of interest would be required. An Si(311) monochromator, which can theoretically achieve $(\Delta\lambda/\lambda)_I$ of the order of 10^{-5} (Table 2), would thus be ideal to maximize the size of the anomalous signal. In practice, source-size effects and the X-ray beam divergence set a limit on the minimum bandpass and many MAD beamlines, particularly at second-generation sources, use wider bandpass monochromators such as Si(111) and Si(220). While these monochromators broaden the width

---

[2] This statement refers to *K* and L_{II} and L_{III} absorption edges edges. L_I edges are less useful for MAD experiments, because the typical value of f' at the inflection point is only about -9 e⁻, compared with approximately -19 and -12 e⁻ for the L_{III} and L_{II} absorption edges, respectively (in the presence of a white line, f' becomes even more negative at these edges). The smaller $|f'|$ at the edge results in smaller dispersive differences.

Table 2
Intrinsic wavelength bandpasses for different silicon reflections, assuming a perfectly cut crystal.

The actual bandpass of the beamline will be determined by the size of the source and the divergence of the beam.

Type	Reflection	$\Delta\lambda/\lambda$
Si	111	1.33×10^{-4}
Si	220	5.6×10^{-5}
Si	311	2.7×10^{-5}

of the edge (Fig. 1) and decrease the observed magnitude of both f'' and f' in absorption edges with a white line, they also provide a more intense beam, which is essential to permit the study of small crystals and weakly diffracting samples at less intense sources and makes it possible to collect more accurate data.

The acceptable values for the beam wavelength drift and reproducibility are dependent on the beam bandpass. The narrower the wavelength bandpass, the more stable the wavelength needs to be. A certain amount of wavelength instability caused by small beam movements or temperature gradients on the surface of the monochromator is not uncommon. If the resulting wavelength drift is equal or larger than the bandpass, there will be a reduction in anomalous and dispersive signal at the edge wavelengths. With a bandpass of 2–3×10^{-4}, typical of many beamlines with Si(111) monochromators, a total wavelength drift $\Delta\lambda/\lambda$ of approximately 10^{-4} during an experiment at the Se K edge would be acceptable. Appropriate wavelength stability is achieved with adequate cooling of optical elements matching the heat load of the incident X-ray beam. On very intense beamlines, liquid-nitrogen cooling is required. On weaker beamlines on second-generation synchrotron sources water cooling is often sufficient, as proved by the success of MAD experiments at these beamlines (see review by Hendrickson & Ogata, 1997).

The effect of a wavelength instability larger than the bandpass will be most pronounced on the dispersive differences in MAD data, because any wavelength excursions about the inflection point of the absorption edge will make the minimum value of f' increase rapidly and therefore limit the size of the maximum $\Delta f'$ achieved in the experiment. Data collected at a white-line feature will also show reduced anomalous differences. On the other hand, anomalous differences collected away from the edge will not be affected by small drifts in wavelength, making single-wavelength data collection at the high-energy (short-wavelength) side of the absorption edge a suitable strategy in this case. Collection at an energy 50 eV away from the edge is probably the safest strategy if the wavelength stability is suspect.

3. Optimal choice of wavelengths

In order to optimize both anomalous and dispersive wavelengths in MAD experiments, data collection at three wavelengths is necessary in the general case. Although the location of the absorption edge and the characteristics of the beamline,

as described in §2, should ultimately decide exactly where to collect data, in many cases these wavelengths are as follows.

(i) A wavelength above the absorption edge (short-wavelength or high-energy side of the edge), often referred to as 'peak wavelength' even when the edge of interest does not have a white-line feature. For L edges without a white line (Au, Hg, Pb etc.) the maximum f'' is reached above the L_I edge. This is also the optimal wavelength for single-wavelength experiments (SAD or SIRAS).

(ii) The wavelength at the inflection point of the edge. In the case of L edges, the L_{III} edge is the one where the minimum f' is reached.

(iii) A wavelength away from the edge ('remote wavelength'). The choice of the remote wavelength is discussed in more depth below.

In two-wavelength MAD experiments, the best strategy is to optimize the dispersive, rather than the anomalous differences (González et al., 1999; Peterson et al., 1996). This means that data collection at the inflection and remote wavelengths is preferred in this case.

3.1. Choosing the remote wavelength

In order to achieve maximum separation between the phasing circles and therefore optimize the phasing power of the MAD experiment, the anomalous scattering factor at the remote wavelength should be such that it maximizes the quantity $f'' \times \Delta f'$, where $\Delta f'$ is the difference in f' between the remote and the minimum f' wavelength in the data set. Generally, the farther away from the absorption edge, the more effective the remote wavelength will be. For example, for an Se edge, collecting the remote wavelength at $\lambda = 0.9$ Å instead of 0.96 Å results in a 25% increase in $\Delta f'$, while f'' only decreases by 13%. At some point (in terms of energy, at between 4 and 5 keV above K edges), the loss in f'' starts offsetting the gain in $\Delta f'$ and there is then no point in going to shorter wavelengths. In most cases, there are also practical concerns to consider when choosing the remote wavelength related to the diffracted intensity (lower at shorter wavelengths) and the experimental setup (the larger the wavelength change within the experiment, the more hardware parameters will have to be adjusted, from undulator gaps to detector position). Therefore, it may be impractical to collect very widely separated wavelengths in the same experiment. For K edges, choosing the remote energy between 500 and 1000 eV away from the edge is often a good compromise between measuring large dispersive differences and keeping the experiment simple.

For L edges, the choice of remote wavelength can be more complex. An energy about 200–300 eV above the L_I edge is a good choice, but if this edge is not accessible at the beamline or the X-ray beam intensity is too low to make it practical to collect data, a long remote wavelength on the low-energy side of the L_{III} edge would be the second best choice. The f'' value is large below the edge because of the contribution from the M edge in the soft X-ray region. The Cu $K\alpha$ emission wavelength could be suggested as a practical limit for the longest suitable

remote wavelength, although the limit could also be decided for each particular experiment, taking into account the dimensions of the crystal and intensity of the X-ray beam as a function of the wavelength (Teplyakov *et al.*, 1998). Beyond this limit, the problems derived from data collection at long wavelengths described in §2.1 outweigh the higher anomalous and dispersive signals. The wavelengths with a local f' maximum below the L_I and L_{II} edges would also be reasonable, if less optimal choices for remote wavelengths when collecting on samples with L edges. As an example, Fig. 2 shows appropriate choices for the remote wavelength for a mercury experiment.

Finally, if the absorption edge has a very high white line, the inflection point of the descending edge of the white line is also a local f' maximum and this wavelength could also be a suitable remote (Shapiro *et al.*, 1995). This could be decided on a case-by-case basis after analysis of the fluorescence scan and comparison of the f' value between this wavelength and those suggested above.

4. Data redundancy and completeness

Experimental phasing will be more successful the more reflections are accurately phased. Often, a unique completeness of at least 90% is needed and even greater completeness is required for poorly diffracting crystals or low anomalous signal (González, 2003). The data sets must contain a high proportion of Friedel-related reflections. The Friedel completeness usually must be close to 100% for SAD phasing, where anomalous differences provide all the phase information. For MAD, the Friedel pair completeness can be lower. For most cases studied, values between 45 and 80% have been found to be sufficient for two- or three-wavelength MAD phasing (González, 2003). Thus, although the experimental phases will always be significantly better with collection of a complete Friedel set, this should not be strictly necessary when
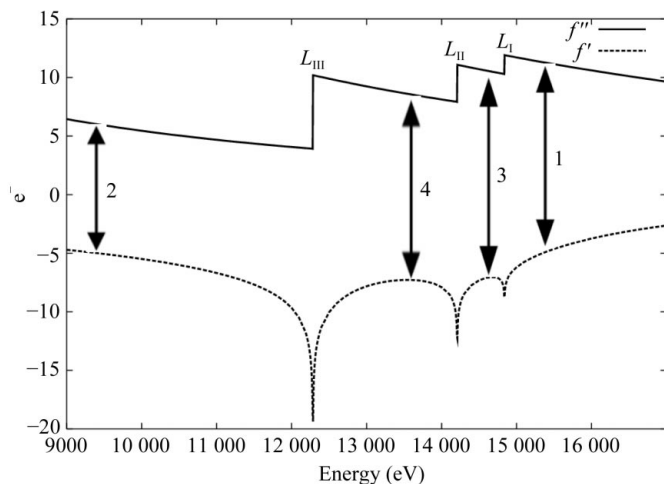
minimizing the radiation damage or finishing an experiment in a short time is the highest priority.

Additional data redundancy improves the experimental phases by decreasing the error in the merged data and thus in the calculated amplitude differences. Having said that, it is better to collect few good well resolved intense diffraction spots for each reflection than many poor ones that overlap with neighboring reflections or are barely above the background intensity of the image. Ensuring that the exposure time is adequate and the sample-to-detector distance is appropriate is important in order to obtain a high $I/\sigma(I)$ for the measured intensities. For multiwavelength experiments, it is very important to collect each reflection under similar conditions at each wavelength (Hendrickson, 1985, 1991). This is easily achieved by the standard practice of using the same crystal and oscillation range for data collection. This procedure causes systematic errors in the measured intensities to partially cancel out when calculating the dispersive differences. Achieving a systematic error reduction to the same extent for anomalous measurements appears to be more difficult. Setting the crystal so that Bijvoet-related reflections are collected in the same diffraction image is not always possible. Collecting an 'inverse-beam' oscillation pass (with the crystal rotated 180° with respect to the original pass) can lengthen the experiment considerably, particularly when the crystal symmetry and orientation makes it possible to optimize Friedel and unique completeness with the same rotation angle (Dauter, 1997); in this particular case, an inverse-beam pass will most likely be superfluous for MAD phasing. For unfavorable crystal symmetry and orientations, extremely small anomalous signals and for SAD phasing inverse-beam collection is useful, although sophisticated scaling methods (Evans, 1997; Friedman *et al.*, 1995) make it possible to obtain a similar result when collecting all the needed data in a continuous rotation wedge. As a precaution against radiation damage, inverse-beam data collection should only be undertaken for a MAD experiment once the pass providing good unique completeness has been collected at all the wavelengths (Rice *et al.*, 2000).

5. Data-collection sequence

For MAD data collection it is possible to collect a full data set at each wavelength at once or collect the data progressively at all wavelengths in wedges of a few degrees. The advantage of the first method is that having a complete or almost complete data set may facilitate structure solution with just one wavelength if severe radiation damage takes place or if the data collection has to be interrupted for any reason. The best wavelength to start with would be the 'peak' wavelength in order to provide optimized SAD phasing. On the other hand, if the structure cannot be solved by SAD with the data available, this strategy often prevents use of the dispersive differences, either because not enough reflections have been collected or because the radiation damage makes it impossible to scale together the intensities measured at different wavelengths.

Figure 2
Possible choices for the remote energy on a mercury MAD experiment, ordered from most optimal to least optimal.

Data collection in wedges results in better preservation of the dispersive differences because the effects of radiation damage would be similarly spread over the data at all wavelengths and it would be easier to scale the data sets and treat the radiation damage like an additional source of systematic error. However, wedge data collection at three wavelengths will demand the collection of more frames than are actually needed to solve the structure with two or one wavelengths (González, 2003). For low-symmetry space groups, there would be a high risk of ending up with three incomplete data sets from which no interpretable maps could be calculated. A good compromise would be to collect complete data sets in wedges at two wavelengths and then, if still possible, continue with collection at a third wavelength. As stated in §3, the best wavelengths to collect simultaneously would be the inflection and the remote wavelengths.

6. Data resolution

Data collection to the maximum diffraction resolution limit improves map interpretability, facilitating automated model building (Morris *et al.*, 2002) and making it possible to observe fine structural details resulting from unbiased experimental phases (Burling *et al.*, 1996; Schmidt *et al.*, 2002). When the purpose of the experiment is to solve the structure *de novo*, however, high-resolution experimental phases are not necessary to obtain high-resolution maps. A more time-efficient strategy is to collect medium- or low-resolution data for phasing and use phase extension to extend the phases to the resolution limit of the crystal. This is straightforward when the high-resolution data are collected from the same crystal (collection at the last wavelength in a three-wavelength MAD data would be an ideal point to try to extend the data resolution). However, if the crystal deteriorates quickly, data from a different crystal or the native can also be used. Lack of isomorphism is usually not a concern once good experimental phases have been obtained. In the most difficult cases, molecular replacement is often successful in locating the molecule in a different cell or space group. This approach has been successful with experimental phases to a resolution as low as 5 Å (Bass *et al.*, 2002).
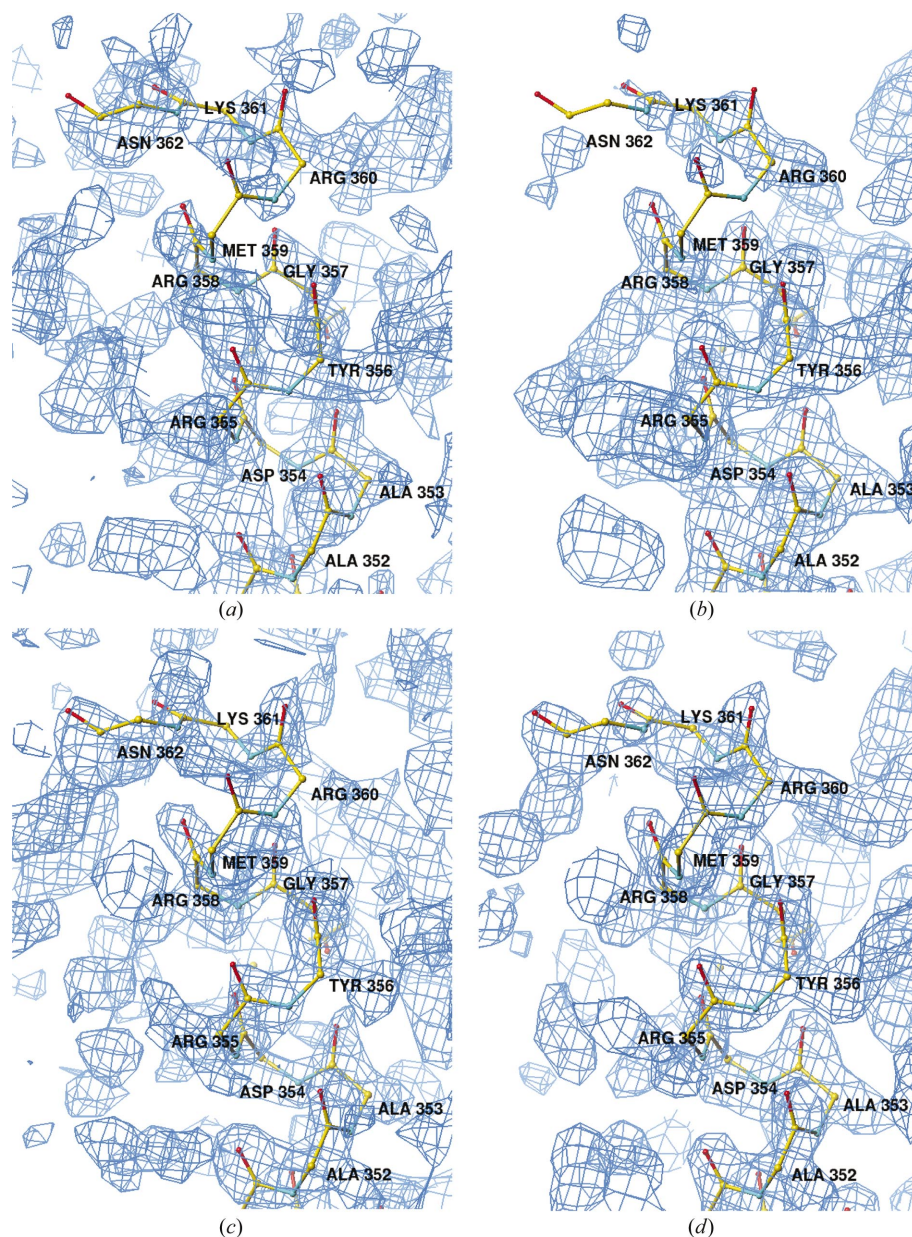

Figure 3
Comparison of maps calculated from SAD and MAD phases using the same total amount of data (52.5°, space group *I*422). (*a*) SAD experimental map. (*b*) SAD map after density modification. (*c*) Two-wavelength MAD experimental map. (*d*) Two-wavelength MAD after density modification. The helical backbone of the protein (PDB code 1kq3) is displayed with the map. Note the areas of disconnected density near the solvent boundary between residues 359 and 362 in the SAD map after density modification (*b*). The density is somewhat better defined in this area in the experimental SAD map (*a*), which implies that in this case the SAD phases are not accurate enough to define the molecular envelope well. The MAD maps tend to show better continuity.

7. Minimizing radiation damage: SAD or two-wavelength MAD?

In cases when the crystal is very sensitive to radiation damage or there is a limited time for the experiment, both one- or two-wavelength experiments will be a better option than the three-wavelength counterpart because they require fewer data for phasing. As Table 3 shows, for many examples SAD and two-wavelength MAD require a similar amount of data for phasing, making it difficult to predict which will be the best

Table 3
Amount of data needed for phasing with optimized SAD data (collected at the peak wavelength) and two-wavelength MAD data (collected at the inflection and remote wavelengths) for three different proteins.

Phasing	Myoglobin		TM0423		TM1083	
	SAD	MAD	SAD	MAD	SAD	MAD
Space group	$P6$		$I422$		$P4_332$	
Anomalous scatterer	Fe		Se		Se	
No. anomalous scatterers in AU	1		7		1	
No. residues in AU	153		364		124	
Oscillation (° per wavelength)	170	55	52.5	25	20	10
Unique (%)	100	96	100	91	99	91
Acentric (%)	100	64	97	64	97	70
Multiplicity	9.4	3.3	4.1	2.0	4.7	2.7
Map correlation						
Experimental	0.32	0.37	0.37	0.43	0.27	0.35
DM	0.61	0.67	0.60	0.66	0.62	0.65

strategy to shorten the experiment on a case-by-case basis. In terms of total dose absorbed during the experiment, which appears to be the ultimate factor determining radiation damage (Garman & Nave, 2002), two-wavelength data collection at the inflection and remote wavelengths offers the advantage of avoiding data collection at the maximum f'' wavelength, which is where the maximum absorption per incident photon takes place. This suggests that this strategy would be better than SAD, where the entire collection is performed at a wavelength with a high f'' value.

Regarding map quality, MAD experiments also appear to be the better strategy. Although SAD maps experience a relatively greater improvement after density modification (Table 3), SAD phases are more likely to lack the accuracy required to determine an accurate molecular envelope. This translates into disconnected density areas in the maps after solvent flattening (see Fig. 3).

On the other hand, a single-wavelength experiment may be further shortened and therefore the preferred strategy to reduce radiation damage if the SAD method is complemented by existing phase information from additional sources (for example, with isomorphous differences with a native data set, direct methods (Langs *et al.*, 1999; Foadi *et al.*, 2000) or a partial model of the structure.

8. Summary

The optimal strategy for anomalous dispersion experiments depends on the properties of the sample, beamline characteristics and ultimately on the purpose of the experiment. When the data-collection objective is to obtain a very accurate picture of structural details, a fairly redundant (with an average multiplicity of 4 or higher) data collection to high resolution at three wavelengths is the best option.

When radiation damage is a serious concern (small, weakly diffracting, very radiation sensitive or low-symmetry crystals), a reasonably strategy would be to collect a complete data to low resolution at the remote and inflection wavelengths in wedges of a few degrees at each wavelength. The total oscillation range would be chosen to maximize the unique completeness of the data and when the symmetry allows it, the Friedel-pair completeness. The Friedel completeness can then be optimized for at least one of the wavelengths if the crystal is still diffracting well once a complete set of phases is secured. The phases can be further improved by collection of the peak wavelength or additional data redundancy at the first two wavelengths. Finally, collection of additional data to higher resolution might also be advantageous. A fully capable MAD beamline, delivering a very stable beam over a sufficiently wide wavelength spectrum, is required for this strategy to give optimal results.

Another alternative procedure would be data collection at the peak wavelength for SAD phasing. In this case, Friedel completeness is more important than for the two-wavelength MAD strategy. Selecting a strategy which maximizes the number of Friedel-related reflections in the least time is important and inverse-beam mode collection can be very useful. If the stability of the beamline cannot be guaranteed or if the energy range available is not wide enough to access a good remote energy, this strategy would be the preferred option. The two-wavelength MAD strategy will provide better maps, comparable in quality to three-wavelength MAD phases and it is possible that by avoiding data collection at the maximum f'', the dose absorbed by the crystal will also be minimized, although more experiments are necessary to prove this.

References

Bass, B. R., Strop, P., Barclay, M. & Rees, D. C. (2002). *Science*, **298**, 1582–1587.
Bijvoet, J. M. (1949). *Proc. Acad. Sci. Amst.* **52**, 313.
Burling, F. T., Weis, W. I., Flaherty, K. M. & Brünger, A. T. (1996). *Science*, **271**, 72–77.
Burmeister, W. P. (2000). *Acta Cryst.* D**56**, 328–341.
Dauter, Z. (1997). *Methods Enzymol.* **276**, 326–344.
Dauter, Z. & Adamiak, D. A. (2001). *Acta Cryst.* D**57**, 990–995.
Dauter, Z., Dauter, M., La Fortelle, E. de, Bricogne, G. & Sheldrick, G. M. (1999). *J. Mol. Biol.* **289**, 83–92.
Dauter, Z., Dauter, M. & Dodson, E. (2002). *Acta Cryst.* D**58**, 494–506.
Ealick, S. E. (2000). *Curr. Opin. Chem. Biol.* **4**, 495–499.
Evans, P. R. (1997). *Proceedings of the CCP4 Study Weekend. Recent Advances in Phasing*, edited by K. S. Wilson, G. Davies, A. W. Ashton & S. Bailey, pp. 97–102. Warrington: Daresbury Laboratory.
Fan, H. F., Han, F. S., Qian, J. Z. & Yao, J. X. (1984). *Acta Cryst.* A**40**, 489–495.
Foadi, J., Woolfson, M. M., Dodson, E. J., Wilson, K. S., Jia-xing, Y. & Chao-de, Z. (2000). *Acta Cryst.* D**56**, 1137–1147.
Friedman, A. M., Fischman, T. O. & Steitz, T. A. (1995). *Science*, **268**, 1721–1727.
Garman, E. (1999). *Acta Cryst.* D**55**, 1641–1653.
Garman, E. & Nave, C. (2002). *J. Synchrotron Rad.* **9**, 327–328.
González, A. (2003). *Acta Cryst.* D**59**, 315–322.
González, A., Pédelacq, J.-D., Sola, M., Gomis-Rüth, F. X., Coll, M., Samama, J.-P. & Benini, S. (1999). *Acta Cryst.* D**55**, 1449–1458.
Graber, T., Mini, S. M. & Viccaro, P. J. (1998). *Proc. SPIE*, **3448**, 21–22.

Green, D., Ingram, V. & Perutz, M. F. (1954). *Proc. R. Soc. London Ser. A*, **255**, 287–307.

Guss, J. M., Merritt, E. A., Phizackerley, R. P., Hedman, B., Murate, M., Hiodgson, K. O. & Freeman, H. C. (1988). *Science*, **241**, 806–811.

Hendrickson, W. A. (1985). *Trans. Am. Crystallogr. Assoc.* **21**, 11–21.

Hendrickson, W. A. (1991). *Science*, **254**, 51–58.

Hendrickson, W. A. (1999). *J. Synchrotron Rad.* **6**, 845–851.

Hendrickson, W. A. & Ogata, C. M. (1997). *Methods Enzymol.* **276**, 326–344.

James, R. W. (1948). In *The Optical Principles of the Diffraction of X-rays.* London: G. Bell & Sons.

Kahn, R., Carpentier, P., Berthet-Colominas, C., Capitan, M., Chesne, M.-L., Fanchon, E., Lequien, S., Thiaudière, D., Vicat, J., Zielinski, P. & Stuhrmann, H. (2000). *J. Synchrotron Rad.* **7**, 131–138.

Langs, D. A., Blessing, R. H. & Guo, D. Y. (1999). *Acta Cryst.* A**55**, 755–760.

Lehmann, M. S., Müller, H. H. & Stuhrmann, H. B. (1993). *Acta Cryst.* D**49**, 308–310.

Morris, R. J., Perrakis, A. & Lamzin, V. S. (2002). *Acta Cryst.* D**58**, 968–975.

Murray, J. & Garman, E. (2002). *J. Synchrotron Rad.* **9**, 347–354.

Panjikar, S. & Tucker, P. A. (2002). *J. Appl. Cryst.* **35**, 261–266.

Peterson, M. R., Harrop, S. J., McSweeney, S. M., Leonard, G. A., Thompson, A. W., Hunter, W. N. & Helliwell, J. R. (1996). *J. Synchrotron Rad.* **3**, 24–34.

Ravelli, R. B. G. & McSweeney, S. (2000). *Structure*, **8**, 315–328.

Ravelli, R. B. G., Theveneau, P., McSweeney, S. & Caffrey, M. (2002). *J. Synchrotron Rad.* **9**, 355–360.

Rice, L. M., Earnest, T. N. & Brünger, A. T. (2000). *Acta Cryst.* D**56**, 1413–1420.

Schmidt, A., González, A., Morris, R. J., Costabel, M., Alzari, P. M. & Lamzin, V. S. (2002). *Acta Cryst.* D**58**, 1433–1441.

Shapiro, L., Fannon, A. M., Kwong, P. D., Thompson, A. M., Lehman, M. S., Grubel, G., Legrand, J.-F., Als-Nielsen, J., Colman, D. R. & Hendrickson, W. A. (1995). *Nature (London)*, **374**, 327–337.

Teplyakov, A., Oliva, G. & Polikarpov, I. (1998). *Acta Cryst.* D**54**, 610–614.

Usón, I. & Sheldrick, G. M (1999). *Curr. Opin. Struct. Biol.* **9**, 643–648.

Wang, B.-C. (1985). *Methods Enzymol.* **115**, 90–112.