Phasing in iodine for structure determination

Zbigniew Dauter

A new method for solving X-ray crystal structures relies on the introduction of iodine into proteins in the form of *p*-iodophenylalanine.

Iodine, most familiar as a brownish liquid in the back of the medicine chest, may soon become a favorite tool of X-ray crystallographers. In this issue, Xie *et al.*[1] show that protein crystal structures can be solved using the phasing signal produced by an iodine-containing amino acid incorporated into the protein, providing an alternative to the most popular method of solving crystal structures.

In the not-so-distant past, determination of the three-dimensional structure of a protein required months or years of tedious work by highly trained specialists. This situation has changed dramatically in the past few years because of rapid progress in the methodologies of macromolecular crystallography and nuclear magnetic resonance, as documented in the exponential increase of protein structures in the Protein Data Bank (http://www.wwpdb.org/).

These technological developments have been catalyzed in part by structural genomics projects, which aim to rapidly elucidate all existing unique protein folds and to connect protein folds with their biological functions. Unlike traditional X-ray crystallography, these new initiatives, organized in recent years by several dedicated consortia in the United States, Europe and Japan, are attempting to solve large numbers of novel protein structures of unknown function in a high-throughput mode with a high degree of automation at every step of the process. To maximize the chances of solving novel structures, the target proteins should have less than 30% sequence identity to any protein of known structure. With the initial pilot studies largely completed, the initiatives are moving to the production stage, which should result in the elucidation of hundreds of novel structures every year.

When a protein crystal is exposed to an X-ray beam, X-rays scatter off the electrons of each atom in the crystal, producing a complex diffraction pattern of spots (**Fig. 1**). Reconstructing the atomic structure of the protein from the diffraction pattern can be challenging. Although the diffraction data captures the positions and intensities of the diffracted X-rays, information about their phases is lost. In the case of scattered visible radiation, images of objects can be reconstructed by the use of lenses, as in the human eye, but no lenses exist for X-rays, and the phases of the scattered rays must be estimated computationally.

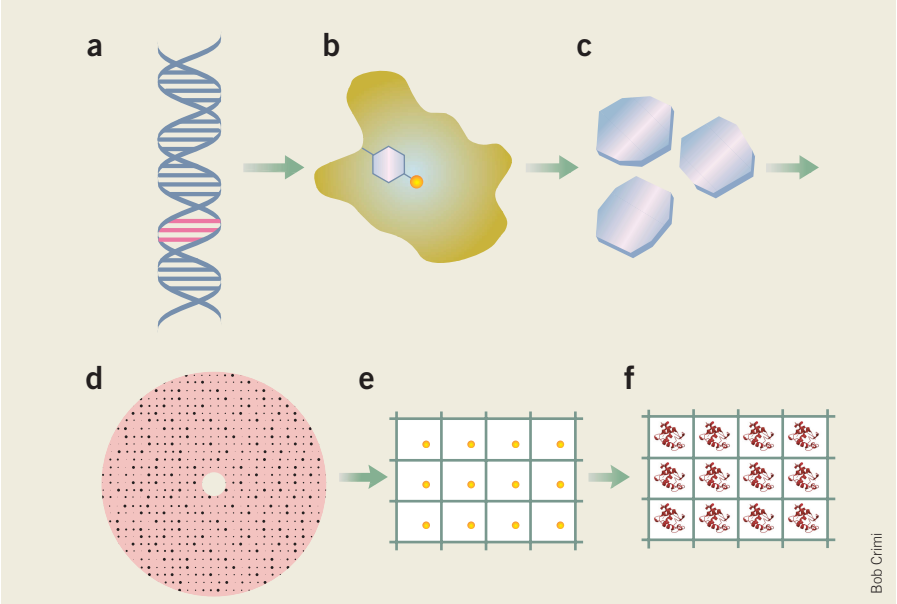

Figure 1 Scheme of the method in Xie *et al.* (**a**–**c**) Modified DNA containing an amber codon (**a**) was used to produce a mutant protein containing an iodophenylalanine residue (**b**), which was then crystallized (**c**). (**d**–**f**) Diffraction data were collected in-house with a copper source of X-rays (**d**). The anomalous signal contained in the data was used to locate iodine atoms in the crystal lattice (**e**), and the protein structure was solved by SAD phasing (**f**).

To solve the crystal structure of a novel protein, it is necessary to have markers in the form of heavy or 'anomalous' atoms[2]. The scattering signal of such atoms stands out from that of the bulk of hundreds or thousands of lighter C, O, N and H atoms because of their large number of electrons or their special, anomalous scattering properties. Since the 'substructure' of heavy atoms is small, it is much easier to locate them and use their calculated scattering contribution for the initial estimation of phases, which can then be improved in an iterative fashion to gradually reveal the complete structure of the macromolecule in the crystal.

The traditional means of obtaining heavy-atom-containing proteins is to soak the protein crystal in solutions containing salts of Hg, Pt, Au, etc. The heavy atom will diffuse into the crystal though solvent channels and bind to the protein at one or more sites. The initial phases are then deduced from differences between the diffraction measured from the native and the derivatized crystals. As an alternative to heavy-atom soaks, heavy atoms naturally present in the crystal can give a suitable signal in some cases, as in some metalloproteins.

Yet another technique is to incorporate heavy atoms into the protein before crystallization. Currently, the most popular approach to the initial phase evaluation uses the heavy atom selenium, in the form of selenomethionine. This amino acid can be incorporated into proteins by expressing them in the presence of selenomethionine in the medium. Selenium produces an anomalous diffraction signal, allowing the structure to be solved by a method called multi-wavelength anomalous diffraction (MAD)[3].

Zbigniew Dauter is at the Synchrotron Radiation Research Section of the National Cancer Institute, Brookhaven National Laboratory, Bldg 725A, Upton, New York 11973, USA. e-mail: dauter@bnl.gov

Anomalous scattering is a resonant effect and occurs when the energy of X-rays equals or exceeds the excitation energy of electrons in some atoms, so that X-rays are absorbed and then emitted by these anomalous scatterers in the form of fluorescence. In MAD, the phases are deduced from differences in the signals of anomalous scatterers at several X-ray wavelengths in the vicinity of the X-ray absorption edge of these atoms. MAD experiments can be done only at synchrotron beam lines, where the X-ray wavelength can be tuned to precise values at the selected absorption edge.

An alternative method that is gaining in popularity is single-wavelength anomalous diffraction (SAD). In SAD, it is enough to measure diffraction data with only one wavelength, not necessarily tuned to the absorption edge of the anomalous scatterers. Synchrotron radiation may be useful but is not a prerequisite for SAD phasing, and diffraction data can be collected at copper-anode X-ray sources found in university laboratories. SAD phasing requires fewer data than MAD and is methodologically simpler.

Iodine has been used as a heavy atom in the past, either by soaking crystals in KI_3[4] or N-iodosuccinimide[5], which iodinates tyrosines, or by cryo-soaking crystals in iodide salts[6]. With these techniques, however, the level of iodine inclusion is often low. In addition, the iodine can perturb the protein structure or cause the crystals' diffraction properties to deteriorate.

The ingenious method of Xie *et al.*—a compelling application of the unnatural-amino-acid technology developed by this group[7,8]—circumvents these problems and allows convenient production of iodine-containing proteins for SAD phasing (**Fig. 1**). The unique pair $tRNA^{Tyr}_{CUA}$-tyrosyl tRNA synthetase was imported from *Methanococcus jannaschii* to *Escherichia coli* and made specific for iodophenylalanine by iterative mutations and selection[7]. Xie *et al.* used this approach to generate a mutant bacteriophage T4 lysozyme in which a single phenylalanine residue was replaced by iodophenylalanine. The fidelity of iodophenylalanine incorporation was almost 100%.

The mutant lysozyme was crystallized, diffraction data were collected in-house and the structure was solved using SAD. One iodine atom among 164 residues resulted in a strong anomalous signal, about 3% of the total diffraction intensities, comparing favorably with the level routinely achieved with selenomethionine-containing crystals at synchrotron beam lines. The presence of iodophenylalanine in the hydrophobic core of lysozyme did not perturb the molecular structure.

Looking to the future, it is possible that the method of Xie *et al.* could be exploited in conjunction with the recently introduced method, radiation-damage induced phasing[9]. Irradiation of macromolecular crystals by intense X-ray beams causes certain specific chemical changes in crystalline samples, including radiolysis of relatively weak bonds, such as in disulfide bridges and in halide-substituted aromatic rings. Gradually vanishing heavy atoms (in analogy to the famous Cheshire cat) create isomorphous changes that can be used to evaluate protein phases. It may be expected that iodophenylalanine will behave analogously to bromouracil[9] and iodotyrosine[10], so that the disappearance of iodines would provide a useful additional phasing signal.

1. Xie, J., Wang, L., Brock, A., Spraggon, G. & Schultz, P.G. *Nat. Biotechnol.* **22**, 1297–1301 (2004).
2. Blundell, T.L. & Johnson, L.N. *Protein Crystallography* (Academic Press, New York, 1976).
3. Hendrickson, W.A. *J. Synchrotron Rad.* **6**, 845–851 (1997).
4. Sigler, P.B. *Biochemistry* **9**, 3609–3617 (1970).
5. Brzozowski, A.M. *et al. Nature* **351**, 491–494 (1991).
6. Dauter, Z., Dauter, M. & Rajashankar, K.R. *Acta Crystallogr.* **D56**, 232–237 (2000).
7. Wang, L., Brock, A., Herberich, B. & Schultz, P.G. *Science* **292**, 498–500 (2001).
8. Chin, J.W. *et al. Science* **301**, 964–967 (2003).
9. Ravelli, R.B.G. & McSweeney, S.M. *Structure* **8**, 315–328 (2000).
10. Zwart, P.H., Banumathi, S., Dauter, M. & Dauter, Z. *Acta Crystallogr.*, in press (2004).

A universal TANGO?

Vijay S Pande

A new algorithm can predict the propensity of proteins to aggregate.

Many diseases have been linked to protein aggregation, including Creutzfeldt-Jakob, Alzheimer, Huntington and Parkinson diseases[1]. With the growing list of aggregation-related diseases, it is tempting to ask whether protein aggregation is a universal phenomenon. In other words, are there common aggregation-related aspects to these diseases? The strongest way to test a hypothesis is to use its predictive aspects; this approach also has the important and imminently practical implication of creating a technology with numerous diagnostic applications. The work of Serrano and coworkers[2] in this issue is a beautiful test of the hypothesis of the existence of universal aspects in protein aggregation–related diseases. By developing TANGO, a novel means for identifying β-sheet aggregation propensity, the authors have found that this propensity can be predicted. Additionally, they show that this propensity can predict which peptides and their mutants will aggregate in various amyloidogenic diseases.

Several other recent studies have provided tantalizing evidence that common aggregation-related characteristics exist among diseases. For example, Dobson and coworkers[1] have shown that many different proteins that are normally well-behaved can be induced to aggregate. Moreover, Glabe and coworkers[3] have reported that an antibody raised against aggregates involved in Alzheimer disease (Aβ oligomers) can also bind oligomers of peptides believed to be involved in several diseases, including Parkinson (α-synuclein) and Huntington (poly-glutamine) diseases and type II diabetes (islet amyloid polypeptide), but does not bind to the monomers or fibrils of any of these peptides[3]. If an antibody recognizes this broad range of protein aggregates, it may indicate that protein aggregate structures are similar in some way.

How does TANGO work? It is informative to first look at a precursor of TANGO, AGADIR[4], which addresses the helical propensities of peptides through a statistical mechanics perspective. AGADIR uses statistical mechanics to calculate the relative probability of finding a helical versus coiled state for a given peptide sequence, with just a few empirical parameters. The sequence-specific elements are critical, as different side chains will lead to different helical propensities. In a sense, TANGO can be thought of as an AGADIR for β-sheet aggregation. Using a strategy similar to AGADIR, TANGO models proteins using a set of four discrete states (unfolded, helix, turn, aggregated) per residue and calculates the partition function for this system based on a few empirically derived parameters (see **Fig. 1**). With four states, the possible number of configurations becomes computationally intractable with even relatively short proteins (a 20-mer peptide would have trillions of possible states). Therefore, a double stretch

Vijay S. Pande is in the Departments of Chemistry and Structural Biology, Stanford University, Stanford, California 94305, USA. e-mail: pande@stanford.edu