Structure, Vol. 9, R21–R26, February, 2001, ©2001 Elsevier Science Ltd. All rights reserved. PII S0969-2126(01)00565-2

Entering a New Phase: Using Solvent Halide Ions in Protein Structure Determination

Ways & Means

Zbigniew Dauter*‡ and Miroslawa Dauter†
*Synchrotron Radiation Research Section
Macromolecular Crystallography Laboratory
National Cancer Institute and NSLS
†SAIC Intramural Research Support Program
and NSLS
Brookhaven National Laboratory
Upton, New York 11973

Introduction

As proteins have numerous charged or polar groups on their surface they attract various ions present in the protein solution. Aqueous solutions from which protein crystals are grown usually contain additives, including simple inorganic ions in the form of precipitants, buffers, detergents, cryoprotectants and so on. Such additives influence the process of crystal growth, but in crystals reside mainly in the disordered bulk solvent region and are seldom identified in crystal structures. Indeed, ions such as NH_4^+, OH^-, Na^+ and Mg^{2+} are isoelectronic with water molecules and by their appearance in electron-density maps and B factors are indistinguishable from waters, even when occupying perfectly ordered sites within the first solvent shell around the protein surface. Light metal ions can sometimes be identified by the characteristic coordination of their sites, but ammonium or hydroxyl ions cannot be distinguished as they are capable of participating in hydrogen-bonding networks in the same way as water molecules.

Chloride ions are present in many common protein crystallization conditions, either as precipitants or in various buffers. In the pioneering structure determination of hen egg-white lysozyme [1] one of the ordered solvent sites was described as a chloride ion on the basis of its low temperature factor, and indeed the mother liquor contained 1 M NaCl. Various forms of lysozyme crystal have since been shown to contain sites occupied by halide ions, either on the basis of their anomalous scattering signal (e.g., in the case of iodide [2], bromide [3] or chloride ions [4]) or from the isomorphous difference (e.g., bromide ions [5]). The identification of ordered halide sites using the anomalous signal suggested that these ions could be used for phasing protein structures [3, 6].

Chlorine has an X-ray K absorption edge at 4.39 Å, but at a CuKα wavelength of 1.54 Å still has the anomalous scattering contribution f″ of 0.70 electrons [7]. The anomalous signal of chlorine is a little higher than that of sulfur (0.56 electrons), which has successfully been used to phase macromolecules [8]. However, higher B factors and probable fractional occupancies will lower the phasing power of solvent chloride ions and their practical use would require exceptionally accurate diffraction data. Bromine has one electron more than selenium and its absorption edge at 0.92 Å is similar to that of selenium at 0.98 Å. In the same way that selenomethionine (SeMet) is routinely used to determine protein structures using multiwavelength anomalous diffraction (MAD) methods, bromine has been used as a standard anomalous marker in MAD approaches to solving the crystal structures of nucleic acids [9]. Iodine does not have an absorption edge within the easily accessible wavelength range available at synchrotron beamlines and therefore cannot be used for MAD, but its anomalous signal is quite significant at 1.54 Å where f″ = 6.8 electrons.

Preliminary tests [6] showed that bromide and iodide anions diffuse into the ordered solvent shell of protein crystals when soaked in an appropriate solution, and can successfully be used for phasing. A number of novel protein structures have been recently solved employing this approach.

Procedure

The procedure used for "halide cryosoaking" is very simple. Immediately before freezing, the crystal is dipped for a short time into a cryoprotectant solution that contains a significant concentration of an appropriate halide salt. In practice it is impossible to give a general recipe, as obviously various crystals come from different crystallization conditions and require different cryoprotectants. Moreover, to date, there are more examples for the successful use of bromides than iodides.

Preliminary tests on crystals of xylanase (K.R. Rajashankar, MD and ZD, unpublished results), in which a range of soak times and NaBr concentrations were used, suggest that prolonging the soak time beyond about 20 s does not lead to the more comprehensive incorporation of halide ions. The diffusion of ions, or rather the exchange of solution between the crystal and surrounding liquid, is evidently very fast. This is corroborated by the observation that molecules of glycerol are often observed in high-resolution crystal structures in ordered sites around the protein surface, held by several hydrogen bonds. In some cases, such an effect can be seen after a ten second immersion of a crystal into a cryosolution containing glycerol [10]. The solution exchange time might depend on the crystal size, although the crystals tested were within the range 0.2–0.4 mm and larger specimens are rarely used for data collection.

The concentration of the halide salt seems to have a more pronounced effect on halide incorporation. Successful phasing has been obtained from soaks in halide concentrations of between 0.25 and 1.0 M. Tests with xylanase led to satisfactory phasing provided the concentration of NaBr exceeded 0.4 M; tests with concentrations of between 0.1 and 0.2 M proved unsuccessful. A higher concentration of halide ions leads to more sites with higher occupancies and to increased phasing power. Again, it should be stressed that the success of

‡To whom correspondence should be addressed (e-mail: dauter@bnl.gov).

the phasing procedure depends on many factors, such as the quality of the diffraction data and resolution, packing density, crystal symmetry, and the presence of pseudo-symmetric arrangements of molecules. It is impossible to give a general prescription as to how strong an anomalous signal must be or how many anomalous scatterers per molecule are necessary to assure the structure solution. A surprising rearrangement of molecules that led to the crystal phase transition from an orthorhombic to hexagonal form was observed after soaking crystals of pepstatin-insensitive carboxyl proteinase (PCP) in NaBr solution [11]; KI or NaCl did not have such an effect.

The multitude of possible crystallization conditions and various protein properties requires each case to be treated individually. The original composition of the mother liquor modified for cryofreezing should at first be preserved and halide salt added without changing the concentration of other components. If the mother liquor contains a large amount of salt it can, in part, be substituted by bromide or iodide, particularly if the original salt is NaCl. This was done in the first tests with lysozyme, which is normally crystallized from a 1 M NaCl solution. Crystals of lysozyme were stable in a 1 M solution of NaBr without NaCl, and could be easily crystallized from such a mother liquor. By contrast, 1 M KI had clearly adverse effects on the mosaicity and diffraction power of lysozyme crystals and it was necessary to mix KI and NaCl to obtain good quality diffraction data [6].

Another possibility, not yet deeply explored, is the use of halide salt as a cryoprotectant without the addition of other protecting ingredients, like glycerol, glycol or sugars. It is known that solutions of salts in high concentration freeze without the formation of ice (e.g., solutions of ammonium formate or lithium sulfate [12]). Indeed, it was possible to collect good quality data from a crystal of lysozyme frozen in a 4 M solution of NaBr without the addition of glycerol.

These results were obtained with crystals soaked in solutions in the pH range 3.2 to 7.5. It might be speculated that low pH should favor the binding of negatively charged halide anions, as the protein should acquire more positive charge. This assertion has not yet been rigorously confirmed experimentally.

As many different factors might affect the binding of solvent halides to various proteins, it is difficult to give a general protocol that is applicable for all cases. Instead, some experimentation with soak conditions is recommended.

Phasing

The data collected from halide-soaked crystals can be used for phasing by MAD, multiple or single isomorphous replacement with anomalous scattering (MIRAS or SIRAS) or for single anomalous diffraction (SAD), as appropriate. Bromine has a very convenient K X-ray absorption edge at 0.92 Å and, as mentioned before, has been used for MAD phasing of nucleic acids following the substitution of thymine with almost isostructural bromouracil [9]. In the vicinity of the absorption edge, both real (or dispersive f′) and imaginary (f″) corrections to the atomic scattering factor vary abruptly with changing wavelength, and these differences are utilized in MAD phasing. The imaginary component is proportional to the fluorescence of the sample, the dispersive component follows the first derivative of the fluorescence curve. The imaginary scattering contribution f″ of bromine is about four electrons beyond the edge and, depending on the experimental conditions, might reach up to six electrons at the white line (i.e., immediately above the absorption edge), whereas the dispersive correction f′ may be as much as minus eight electrons at the inflection point of the fluorescence spectrum. There is, however, a complication with setting up a MAD experiment with bromine-soaked crystals. The anomalous signal of bromide ions occupying ordered sites within the first solvent shell around the protein might differ slightly from that of the completely disordered ions within the bulk solvent in the crystal or in the drop of cryosolution frozen around it. The fluorescence spectrum therefore shows the signal of disordered bromide ions, which are in large excess. It can be recommended to shift the observed values of the inflection point and peak of the spectrum by about 3 eV toward the low energies (or longer wavelength). Obviously back-soaking cannot be used, as it would also remove most of the ordered halide sites. It has been shown that the use of a single wavelength (SAD) was enough to solve structures of bromine-soaked crystals [3, 11, 13]. In such an approach there is no need to know the precise positions of the peak or edge points of the spectrum, the near-remote wavelength can be used instead.

Iodine is not appropriate for MAD, because its K edge is at 0.37 Å and L_I edge at 2.39 Å—beyond the capability of most synchrotron beamlines. It retains, however, a significant anomalous effect at a wavelength corresponding to the copper source of 1.54 Å with an f″ of about seven electrons. Iodine can be used as a heavy atom in the SIRAS (or MIRAS) approach (U. Derewenda et al., unpublished results) and can also be employed for SAD phasing [14].

Any of the existing phasing programs can be used to estimate phases based on the signal of soaked halides. Several have been used successfully and the choice depends mostly on the preferences and experience of the researcher. There is, however, one difference between SeMet and halide phasing that must be taken into consideration. In the classic SeMet case, the expected number of anomalous scatterers is known and the sites are fully occupied, whereas in the halide-soak approach there is an unknown number of halide sites with varying occupancies, often much lower than unity. When direct or Patterson search methods are applied to SeMet data (based on the anomalous or dispersive differences), usually the expected number of strong sites are identified followed by much lower noise peaks. In the case of halides, there is no sharp boundary between the strong, highly occupied sites and noise. The number of sites selected for phasing is arbitrary and can be accepted successively if they appear in the appropriate residual maps.

Example: Human Acyl Protein Thioesterase

The human acyl protein thioesterase (hAPT) provides a good example to illustrate the procedures that can lead

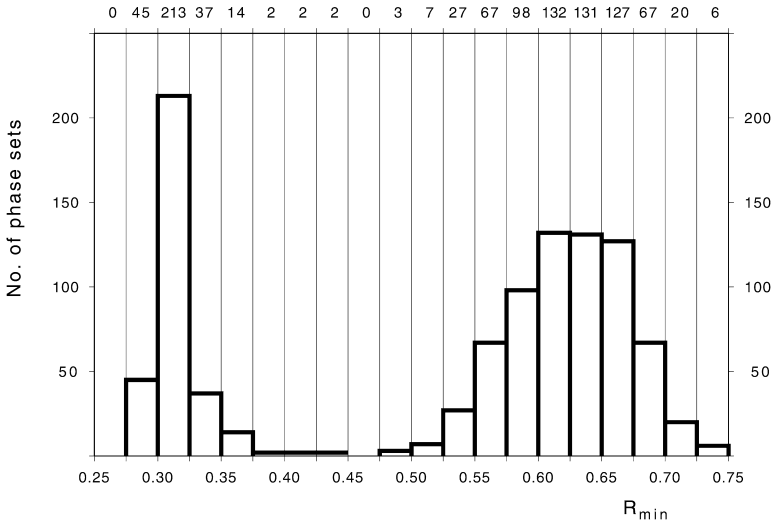

Figure 1. The Solution of the Bromide Substructure in hAPT by SnB

The histogram shows the frequency of phase sets as a function of the R_{min} figure of merit. The number of phase sets in each bin is shown at the top of the histogram.

to a smooth crystal structure determination using SAD of cryosoaked bromides [13]. The 225-residue enzyme crystallized in space group $P2_1$ with two molecules in the asymmetric unit (total molecular weight 56 kDa). The crystal was soaked for 20 s in a solution containing 42% saturated ammonium sulfate (1.8 M), sodium acetate buffer pH 5.0, 20% glycerol and 1 M NaBr.

Diffraction data were collected on an ADSC Quantum4 CCD detector at the National Synchrotron Light Source (NSLS) beamline X9B, using a wavelength of 0.9167 Å (13524 eV), 50 eV beyond the bromine K edge in the high-energy near-remote region. Data processed with HKL2000 [15] showed a significant amount of anomalous signal, with an R_{merge} of 4.8% overall (12.3% for the highest resolution shell of 1.8 Å) and R_{anom} of 4.0% (9.0% at 1.8 Å). The average $\Delta F^{anom}/F$ value was $\sim$4.0% in the low-resolution region (5.4% overall), which corresponds roughly to the effect of about six fully occupied bromine atoms. For accurately measured data sets, the formula $\Delta F^{anom}/F = \sqrt{2}\times(f''\times\sqrt{N_A})/(6.7\times\sqrt{N_P})$ [9] usually reflects the number of anomalous scatterers quite well at low resolution [6].

The Bijvoet differences were submitted to the direct methods program SnB [16] with a request to find eight bromide sites. The program ran very successfully, giving 315 good solutions per 1000 trial phase sets (Figure 1). These eight bromide sites were submitted to SHARP [17], which revealed ten more bromide sites in the residual map. Following a second run, SHARP identified six more sites and the third round of SHARP, based on 22 bromide sites, produced the phase set with a figure of merit (FOM) of 0.40. SHARP phasing was performed using only one set of single wavelength data. The phases were further improved by density modification using DM [18], and the final FOM was 0.85.

The arrangement of bromide sites clearly displayed the noncrystallographic symmetry (Figure 2). This could have been used for further phase improvement by noncrystallographic symmetry averaging, but the DM map was very clear and the wARP procedure [19] was utilized for automatic model building. As a result, 347 of an expected 450 residues were built automatically and it was straightforward to build the complete protein model.

The solution of the hAPT structure took a couple of days from data collection to model refinement, with a minimal amount of human effort. It was not necessary to produce the SeMet variant of the protein or to collect more than one diffraction data set. Furthermore, it was not necessary to spend an excessive amount of time building the initial model at the graphics display.

Halide Sites

When soaked into crystals, halide ions move to sites within the ordered solvent shell around the protein surface, and share them with water molecules. The occupancy of their sites varies. A small number of sites might be almost fully occupied, but there are more sites with lower occupancy down to noise level. This is illustrated in Figure 2, which shows the hAPT dimer with peaks of the anomalous Fourier map plotted at the 10, 5 and 3σ level.

The experience gained from test structures [6] and from the new structures shows that the number of sites is roughly proportional to the size of the protein. It seems that halide anions fill most of the available places at the surface of the protein molecule with varying occupancy. These ions are most probably in a state of equilibrium with the halide ions present in the bulk solvent and their occupancies might depend on several factors. Preliminary tests (K. R. Rajashankar, M. D. and Z. D., unpublished results) show that the diffusion process is rapid and soak times longer than ten seconds do not visibly increase the number or occupancy of identifiable sites. In the crystal of hAPT, following a 20 s soak in 1 M NaBr, one of the bromide ions penetrated to the anion hole in the active site of the enzyme, which is located at the tight monomer–monomer interface. The concentration of halides in the soaking solution seems to influence their occupancy more significantly. In addition, there are other factors that might be relevant, such as the pH of the solution or the presence of other cations, anions or neutral additives. The evaluation of their importance requires further tests.

All halide ions share their sites with water molecules. Their coordination, appearance in electron-density maps

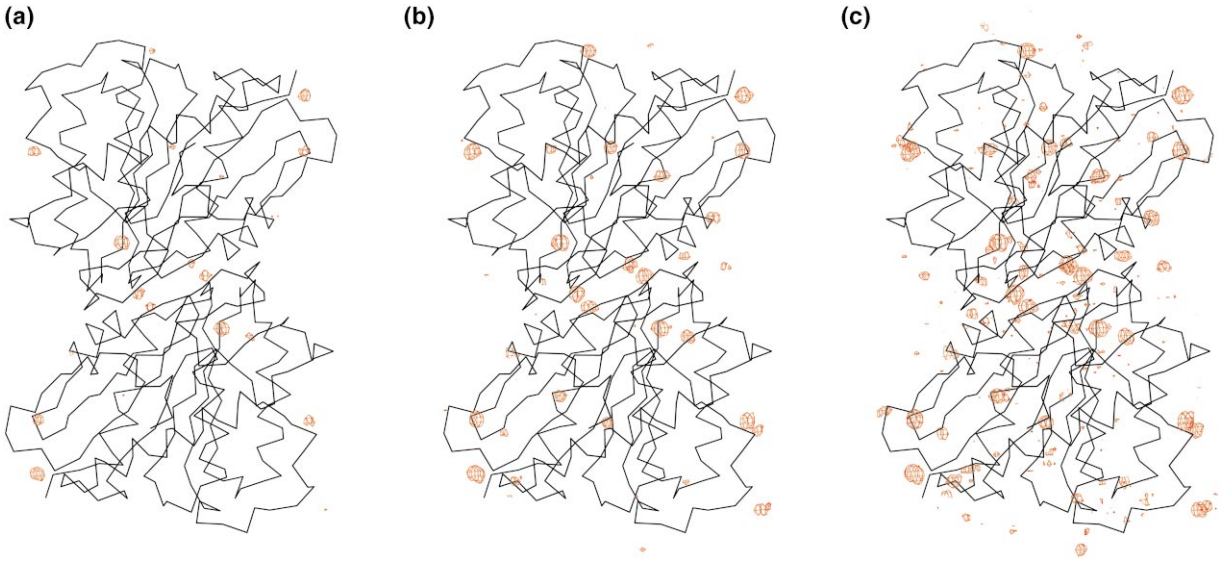

Figure 2. Cα Trace of the hAPT Dimer Viewed Along its Noncrystallographic 2 fold Axis
The anomalous difference map ($\Delta F^{\pm}$, $\phi-\pi/2$) is superimposed at (a) 10σ, (b) 5σ and (c) 3σ level. (The figure was prepared using BOBSCRIPT [21].)

and behavior during structure refinement is almost identical to that of fully occupied water molecules and only rarely is it possible to differentiate bromide or chloride ions from waters, especially if the sites are partially occupied. These ions can, however, be easily identified by their anomalous scattering signal. The presence of only one chloride ion was postulated in the original structure of lysozyme [1] on the basis of its low temperature factor, whereas even the rather weak anomalous signal of chlorine at a wavelength of 1.54 Å led to the identification of eight chloride ion sites [4].

The coordination geometry of halide ions is not specific. Several bromide sites around the molecule of hAPT are shown in Figure 3. The halide sites observed in other structures are very similar, in that every site has a different environment. Halide ions usually accept hydrogen bonds from various donor groups from the protein and neighboring water molecules. In addition, they make van der Waals contacts with non-polar protein atoms. This is in marked contrast to classic heavy-atom sites. Metal ions usually require more or less specific coordination by electron-donating ligands [20]. In some cases the ligands can be identified, for example, the typical octahedral arrangement of carbonyl or carboxyl oxygen atoms around calcium (or similar) ions. In derivatives of metals such as platinum, gold or osmium, it is not normally possible to locate all of the metal ligands owing to disorder, exchange with solvent and the low resolution of diffraction data. Mercury might bind covalently to cysteine sulfhydryl groups, but can also coordinate to carboxylate groups [10]. Xenon or krypton under pressure penetrate into hydrophobic cavities and do not tolerate the neighborhood of solvent water molecules. Complex anions, such as selenates, which are potentially good anomalous scattering markers, bind through their oxygen atoms and avoid the non-polar surface. The halide anions are monoatomic and polarizable, and consequently able to engage in both polar and hydrophobic interactions.

In spite of the non-specific geometry of halide coordination, the chemistry of their sites is quite specific. This can be seen in Figure 2, where the highest bromide peaks around the hAPT dimer are located in analogous sites around both monomers. Among the 22 sites used for phasing, only four weak sites (ranked as 12, 16, 19 and 22) did not have local symmetry counterparts. The other 18 sites are arranged as nine pairs of analogous sites around two monomers, following the noncrystallographic symmetry with an root mean square difference of 0.21 Å.

Of the sites that are best for phasing, most contain halide ions that are hydrogen-bonded to amide nitrogen atoms, either from the protein mainchain or asparagine and glutamine sidechains. In addition, good sites often make ionic pairs with arginine or lysine residues. Sometimes, hydrogen bonds to the hydroxyl groups of threonine or serine residues can also be observed. All halide ions are in contact with water molecules, which can be ordered or in the bulk solvent region.

Conclusions

Halide ions present in the cryo-solution in molar or near submolar concentrations are able to rapidly diffuse into protein crystals and adopt ordered, albeit partially occupied, sites within the first solvent shell. The anomalous scattering signal within the diffraction data collected from such crystals can be utilized for structure solution. Bromides are appropriate for MAD phasing, iodides can be used for SIRAS and all halides, including chlorides, can be used for SAD phasing.

The number of sites accepted for phasing can be selected arbitrarily from initial direct methods or Patterson search procedure and from successive anoma-

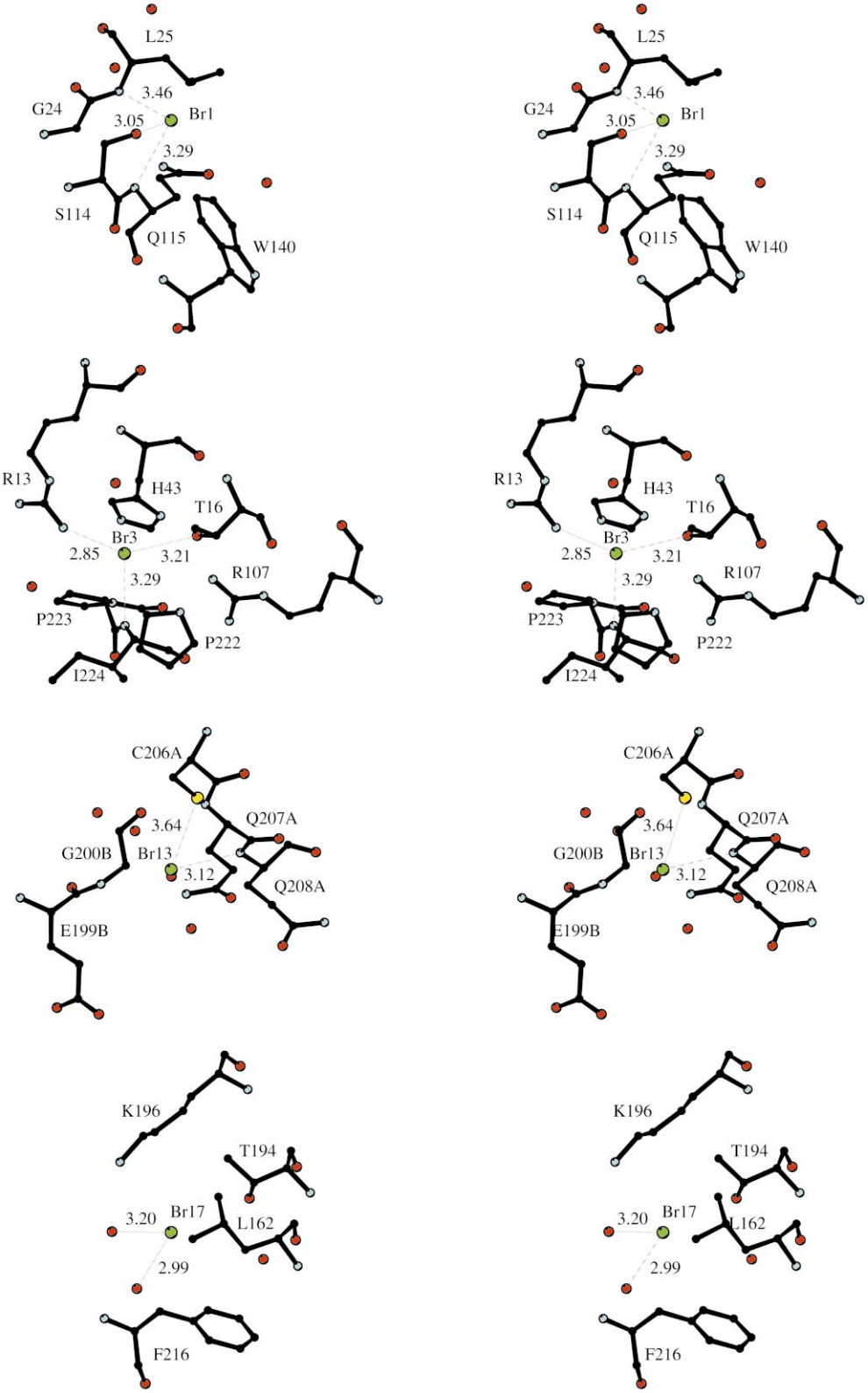

Figure 3. Bromide Sites in the hAPT Structure

A selection of bromide ions identified in the hAPT structure, illustrating the diversity of their environments. Surrounding residues are shown if at least one atom lies closer than 5 Å from a bromide site. Hydrogen bonds to bromide ions and their lengths (in Å) are marked. The figures are in stereoview. (The figure was prepared using BOBSCRIPT [21].)

lous difference syntheses, as long as the new sites improve the phasing process.

This approach is quick and requires little effort. It can be proposed as an alternative way to prepare derivatives when a protein does not bind heavy-metal atoms or is not amenable to the preparation of a SeMet variant. Because it is easy and rapid, this method is likely to be of particular relevance for high-throughput structural genomics projects.

References

1. Blake, C.C.F., Mair, G.A., North, A.C.T., Phillips, D.C., and Sarma, V.R. (1967). On the conformation of the hen egg-white lysozyme molecule. Proc. R. Soc. Lond. B. Biol. Sci. *167*, 365–377.
2. Steinrauf, L.K. (1998). Structures of monoclinic lysozyme iodide at 1.6 Å and of triclinic lysozyme nitrate at 1.1 Å. Acta Crystallogr. D *54*, 767–779.
3. Dauter, Z., and Dauter, M. (1999). Anomalous signal of solvent bromides used for phasing of lysozyme. J. Mol. Biol. *289*, 93–101.
4. Dauter, Z., Dauter, M., de La Fortelle, E., Bricogne, G., and Sheldrick, G.M. (1999). Can anomalous signal of sulfur become a tool for solving protein crystal structures? J. Mol. Biol. *289*, 83–92.
5. Lim, K., Nadarajah, A., Forsythe, E.L., and Pusey, M.L. (1998). Locations of bromide ions in tetragonal lysozyme crystals. Acta Crystallogr. D *54*, 899–904.
6. Dauter, Z. Dauter, M., and Rajashankar, K.R. (2000). Novel approach to phasing proteins: derivatization by short cryo-soaking with halides. Acta Crystallogr. D *56*, 232–237.
7. Cromer, D.T. (1983). Calculation of anomalous scattering factors at arbitrary wavelengths. J. Appl. Crystallogr. *16*, 437–438.
8. Hendrickson, W.A., and Teeter, M.M. (1981). Structure of the hydrophobic protein crambin determined directly from the anomalous scattering of sulfur. Nature *290*, 107–113.
9. Hendrickson, W.A., and Ogata, C.M. (1997). Phase determination from multiwavelength anomalous diffraction measurements. Methods Enzymol. *276*, 494–523.
10. Lubkowski, J., et al., and Wlodawer, A. (1999). Atomic resolution structures of the core domain of avian sarcoma virus integrase and its D64N mutant. Biochemistry *38*, 13512–13522.
11. Dauter, Z., Li, M., and Wlodawer, A. (2001). Practical experience with the use of halides for phasing macromolecular structures: a powerful tool for structural genomics. Acta Crystallogr. D *57*, in press.
12. Rubinson, K.A., Ladner, J.E., Tordova, M., and Gilliland, G.L. (2000). Cryosalts: suppression of ice formation in macromolecular crystallography. Acta Crystallogr. D *56*, 996–1001.
13. Devedjiev, Y., Dauter, Z., Kuznetsov, S.R., Jones, T.L.Z., and Derewenda, Z.S. (2000). Crystal structure of the human acyl protein thioesterase I from a single X-ray data set to 1.5 Å. Structure *8*, 1137–1146.
14. Chen, L., et al., and Wang, B.C. (1991). Crystal structure of a bovine neurophysin II dipeptide complex at 2.8 Å determined from the single-wavelength anomalous scattering signal of an incorporated iodine atom. Proc. Natl Acad. Sci. USA *88*, 4240–4244.
15. Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. *276*, 307–326.
16. Weeks, C.M., and Miller, M. (1999). The design and implementation of SnB v2.0. J. Appl. Crystallogr *32*, 120–124.
17. de La Fortelle, E., and Bricogne, G. (1997). Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. Methods Enzymol. *276*, 472–494.
18. Cowtan, K.D., and Zhang, K.Y.J. (1999). Density modification for macromolecular phase improvement. Prog. Biophys. Mol. Biol. *72*, 245–270.
19. Perrakis, A., Morris, R., and Lamzin, V.S. (1999). Automated protein model building combined with iterative structure refinement. Nat. Struct. Biol. *6*, 458–463.
20. Boggon, T.J., and Shapiro, L. (2000). Screening for phasing atoms in protein crystallography. Structure *8*, R143–R149.
21. Esnouf, R.M. (1999). Further additions to MolScript version 1.4, including reading and contouring of electron-density maps. Acta Crystallogr. D *55*, 938–940.