High-throughput PIXE as an essential quantitative assay for accurate metalloprotein structural analysis; development and application

Geoffrey W. Grime[a], Oliver B. Zeldin[b], Mary E. Snell[c], Edward D. Lowe[b], John F. Hunt[e], Gaetano T. Montelione[f], Liang Tong[e], Edward H. Snell[c,d]*, Elspeth F. Garman[b]*

Email: esnell@hwi.buffalo.edu, elspeth.garman@bioch.ox.ac.uk

Addresses

a. Ion Beam Centre, Advanced Technology Institute, University of Surrey, Guildford, Surrey, GU2 7XH, UK

b. Department of Biochemistry, University of Oxford, South Parks Road, Oxford, OX1 3QU, UK

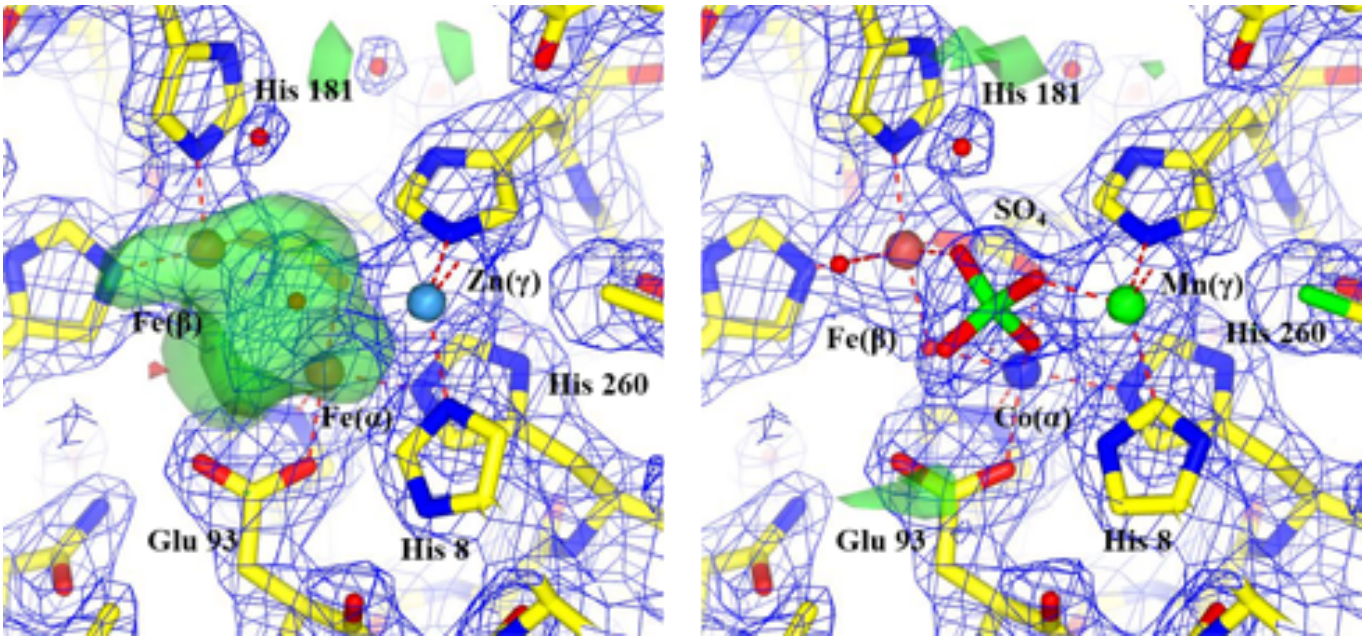
c. Hauptman-Woodward Medical Research Institute, 700 Ellicott St., Buffalo, NY 14203, USA

d. Materials Design and Innovation, SUNY Buffalo, 700 Ellicott St., Buffalo, NY 14203, USA.

e. Department of Biological Sciences, Columbia University, New York, NY, 10027, USA.

f. Department of Chemistry and Chemical Biology, Center for Biotechnology and Interdisciplinary Sciences, Rensselaer Polytechnic Institute, Troy NY 12180
USA.

Suggestion for TOC figure



Abstract

Metalloproteins comprise over one third of proteins, with approximately half of all enzymes requiring metal to function. Accurate identification of these metal atoms and their environment is a prerequisite to understanding biological mechanism. Using ion beam analysis through particle induced X-ray emission (PIXE), we have quantitatively identified the metal atoms in 30 previously structurally characterized proteins using minimal sample volume and a high-throughput approach. Over half of these metals had been misidentified in the deposited structural models. Some of the PIXE detected metals not seen in the models were explainable as artifacts from promiscuous crystallization reagents. For others, using the correct metal improved the structural models. For multi-nuclear sites, anomalous diffraction signals enabled the positioning of the correct metals to reveal previously obscured biological information. PIXE is insensitive to chemical environment, but coupled with experimental diffraction data deposited alongside the structural model, it enables validation and potential remediation of metalloprotein models, improving structural and more importantly, mechanistic knowledge.

1. Introduction

Metals are important in biology: over one third of all proteins having one or more metal ligands playing a key structural or catalytic role[1] critical for the progression of many diseases and attractive for therapeutic

intervention[2]. The correct identity and accurate stoichiometry of these ligands are vital bio-physical data for characterizing proteins and understanding mechanism, but there is currently no widely accepted standard metal assay. If the structural model is known, circumstantial evidence from the model is used, but this has been shown to be unreliable, particularly at low resolutions[3]. For models determined by X-ray crystallography or similar resolution techniques, the choice of metal made at the refinement stage affects the restraints, biasing the final structure. There are sophisticated techniques that make use of anomalous signals in the structure factors which allow for element identification[4-5] independent of PIXE data. However these require a knowledge of the expected elemental species and use multiple refinements, comparing the models produced for the different species, or the use of multiple incident X-ray wavelengths. When no structural information is available, bioinformatic approaches can be used[6], but experimental measurement is not part of routine characterization protocols.

To identify and quantitate the metal content of proteins, the limit of detection (LoD) required is a single metal atom in a 100 kD macromolecule or 500 parts per million by dry weight (ppm). Current methods for metal identification (reviewed by Hare *et al.*[7]) include wet assays[8], mass spectrometry[9] and X-ray[10] or electron based characterization[11] and imaging[12] and various combinations of these[13-14]. The former rely on detecting chemical compounds bound to metal atoms but can only detect one species at a time. Mass spectrometry is sensitive but results can be degraded by partial occupancies, glycosylation or post-translational modifications. X-ray absorption spectroscopy (XAS) requires a high sample volume and has stringent experimental requirements[15]. Electron induced X-ray emission (electron probe microanalysis) has a high LoD so that generally metals in proteins are undetectable. X-ray fluorescence has a sufficiently low LoD and spatial resolution for metal identification[16]. However, the LoD and range of elements detectable depends strongly on the primary X-ray beam energy used to induce X-ray emission, and achieving accurate quantitation is challenging.

We have developed the use of an MeV microbeam ion beam analysis (IBA) technique, proton induced X-ray emission (PIXE), in combination with Rutherford backscattering spectrometry (RBS)[17] to identify metal atoms in proteins and quantify their stoichiometry, in a high-throughput pipeline (HT-IBA). A single microPIXE analysis in a sub-nanolitre sample can detect all elements heavier than sodium with detection limits (<50 ppm) well matched to the accurate quantification of metals in proteins. Our HT-IBA methodology has been validated on a well characterized sample set of 32 metallo-proteins which revealed the promiscuous nature of metals in the protein samples used for crystallization. We examined three in detail for which the IBA detected metal differed from that in the Protein Data Bank (PDB) model, with one revealing important structural knowledge that was initially missed.

2. Experimental Methods

Samples

The 32 proteins used for the study came from the Northeast Structural Genomics (NESG) program. These samples were expressed, purified, and prepared as described elsewhere[18-19]. Purification mostly involved a Ni-NTA affinity column followed by gel filtration chromatography, although in some cases, Co-NTA affinity purification was used. Details of the sample preparation for each protein are recorded in the SPINE database[20] under the corresponding NESG Target ID listed in Table 1. Sample homogeneity was > 95% based on SDS-PAGE gels with Coomassie Brilliant Blue staining.

#	PDB ID	Uniprot or UniParc ID[1] (NESG ID)[2]	Protein description	Residues	Chains in the asymmetric unit	Metal in PDB	Resolution (Å)	Metals in PIXE (>3×LoD)	Potential metals in PIXE (1-3×LoD)	Crystallization conditions or NMR sample buffer for PDB model
Class I. PDB inconsistent with PIXE										
1	3HIX	Q8YQN0 (NsR437I)	Alr3790 protein	106	3	Mn	1.92	-	-	20% PEG 4000, 0.1 M Mn chloride, 0.1 M MES, pH 6.0
2	3JSR	Q8Z082 (NsR236)	All0216 protein	119	1	K	1.80	-	Ca	8.64 M K acetate, 0.1 M TAPS, pH 9.0
3	3ILM	Q8YQN0 (NsR437H)	Alr3790 protein[3]	141	4	Mn	2.26	-	Fe, Co	20% PEG 1000, 0.1M Mn chloride, 0.1 M MES, pH 6.0
4	3HLY	A0A0H3K5M0 (SnR135D)	Flavoprotein	161	3	Ca	2.40	-	Ca	20% PEG 8000, 0.1M Ca acetate, 0.1 M MES, pH 6.0
5	**3GGL**	**Q8A9F0 (BtR324A)**	**Putative chitobiase**	**169**	**1**	**Zn**	**3.00**	**-**	**Ca, Mn, Fe**	**0.75 M Mg formate, 0.1 M Bis-Tris, pH 7.0**
6	**3KB1**	**O28015 (GR157)**	**Nucleotide-binding proteins**	**262**	**2**	**Zn**	**2.87**	**-**	**Co**	**100 mM Na acetate (pH 4.6), 30% MPD, and 200 mM NaCl.**
7	**3DCP**	**A0A3H3KZ89 (LmR141)**	**Histidinal-phosphatase**	**283**	**3**	**Fe/Zn**	**2.10**	**Ca (3.3), Mn (0.5), Fe (1.2), Co (1.2)**	**Zn**	**15% PEG 8000, 0.17 M Na acetate, 0.01 M L-cysteine, 0.1 M MES pH 6.2**

8	3LV4	Q65D31 (BiR14)	Glycoside hydrolase, family 43 YxiA	456	2	Ca	1.70	-	Ca, Mn	18% PEG 3350, 0.2M Ca acetate, 0.1M MES, pH 6.15
Class II. Extra metals present in PIXE										
1	3HNM	Q8AAM3 (BtR319D)	Putative chitobiase	172	4	Mg^4	3.00	Ca (1.74)	-	100mM NaCl, 5mM DTT, 0.02% NaN_3, 10mM Tris-HCl (pH 7.5),
2	3LMC	A2SQK8 (MuR16)	Peptidase, zinc-dependent	210	1	Fe/Zn	2.00	Fe (0.6), Co (0.9), Ni (0.4), Zn (0.7)	-	0.1 M $Na_2MoO_42H_2O$, 0.1 M Bis-Tris propane, 12% PEG 20000
3	3IHK	Q8DVV9 (SmR83)	TPK_B1_binding domain-containing protein	218	3	Mg^4	3.00	Ca (0.5), Fe (0.1)	Ti, Co, Cu	0.1 M $LiCl_2$, 0.1 M Bis-Tris, pH 5.5, 18% PEG 3350
4	3LM8	O34664 (SR677)	Thiamine pyrophospho kinase	222	4	Mg^4	2.56	Ca (0.7), Fe (0.05)	K/Br	0.1 M KH_2PO_4, 0.1 M $NaC_2H_3O_2$, pH 5.0, 12% PEG 20000
5	3KB4	Q8YJY7 (NsR141)	Lipid binding protein	225	2	Mg^4	2.41	Mn (0.2), Fe (0.4), Ni (0.4)	Co	0.1 M citric acid, pH 5.0, 1.6 M $(NH_4)_2SO_4$
6	3E5Z	Q9RXN3 (DrR130)	Gluconolactonase, putative	296	2	Mg^4	2.01	Ca^5	-	0.1 M NaCl, 0.1 M TAPS, pH 9.0, 18% PEG 3350, $MgCl_2$ – no specified concentration
7	3E48	A0A0H3JTC8 (ZR319)	Putative nucleoside-diphosphate-sugar epimerase	289	2	Mg^4	1.60	-	Ca, Fe, Cu	0.1 M Tris-HCl, pH 9.1, 18% PEG 3350, 0.1 M $MgSO_4$
8	3DEV	Q4L745 (ShR87)	Uncharacterized protein	320	2	Mg^4	3.10	Mn (0.8), Fe (0.7)	-	0.1 M Na citrate, pH 5.2, 1.25 M Li_2SO_4, 0.5 M $(NH_4)2SO_4$
9	3K2Q	A1U3A9 (MqR88)	Pyrophosphate-fructose 6-phosphate 1-phosphotransferase	420	3	Na^4	2.48	Ca (7.1)	Fe	0.1 M Na_2MoO_4, 0.1 M Tris, pH 8.0, 20% PEG 8000
Class III. PIXE data consistent with PDB										

1	3NNG	Q5LFR2 (BfR258E)	F5/8 type C domain-containing protein	168	2	Ca	2.18	Ca (1.7)	Fe	40% PEG 4000, 0.1 M $CaCl_2$, 0.1 M Bis-Tris propane, pH 7.0
2	2KPN[5]	Q81D73 (BcR147A)	Bacillolysin	103	1	Ca	N/A	Ca (0.8)		20 mM MES, 200 mM NaCl, 5 mM $CaCl_2$, 10 mM DTT, 0.02% NaN_3, 0.05 mM DSS, 90% H_2O/10% D_2O. pH 6.5
3	3LRQ	O94972 (HR4604D)	E3 ubiquitin-protein ligase TRIM37	100	4	Zn	2.29	Zn (2.5), Fe (0.3)	Ca, Co, Cu	0.1% (w/v) MPD, 0.1% (w/v) 1,2,3-heptanetriol, 0.1% (w/v) diethylenetriaminepentakis (methylphosphonic acid), 0.1% (w/v) D-sorbitol, 0.1% (w/v) glycerol, 0.06 M HEPES, 12.5% PEG 3350
4	3NNQ	P03355 (OR3)	Gag-Pol polyprotein	114	2	Zn	2.69	Ca, Zn[6]	Fe, Ni	2.0 M $Na_2C_3H_2O_4$, 0.1 M $NaC_2H_3O_2$, pH 5.0, 0.05% Anapoe X-305
5	2K52[5]	Q58598 (MjR117B)	Uncharacterized protein	80	1	-	N/A	Ca (0.2)	Fe	20 mM NaAc, 100 mM NaCl, 5 mM $CaCl_2$, 10 mM DTT, 0.02% NaN_3, 0.02 mM DSS, 90% H_2O/10% D_2O. pH 4.5
6	3ESI	Q6CYK7 (EwR179)	FabA domain-containing protein	129	4	-	2.50	-	Ca, Fe	PEG 4000 (no specified concentration), 0.2 M $NH_4C_2H_3O_2$, 0.1 M Na citrate, pH 5.6
7	3DM3	Q58559 (MjR118E)	Replication factor A	105	3	Na[4]	2.40	-	-	0.1 M Na Citrate, 0.1 M NaCl, pH 5.0
8	3I24	Q5E3V1 (VfR176)	HIT family hydrolase	149	2	Na	1.50	-	Co	NaCl (no specified concentration), 0.2 M MES, pH 6.0,

										20% PEG 3350, pH 6.15
9	3L8M	Q49X04 (SyR86)	Putative thiamine pyrophospho kinase	212	2	Na[4]	2.40	-	Fe	RbCl (no specified concentration), 0.1 M Na Citrate, pH 4.2
10	3FOJ	Q49YI7 (SyR101A)	Rhodanese domain-containing protein	100	1	Na[4]	1.60	-	Ca, Fe, Cu	0.15 M $MgSO_4$, 0.1 M Na Citrate, 20% PEG 3350
11	4EVW	A0A0K9UNA8 (VcR193)	Nucleoside-diphosphate-sugar pyrophosphorylase	255	2	Mg[4]	1.89	-	-	40-44% MPD, 0.1 M HEPES, pH 7.5
12	2KW4[5]	Q24WL3 (DhR1A)	RNase H domain-containing protein	147	1	Mg[4]	N/A	-	Ca, Fe	20 mM MES, 200 mM NaCl, 150 mM $MgCl_2$, 5 mM $CaCl_2$, 10 mM DTT, 0.02% NaN_3, 50 μM DSS, 90% H_2O/10% D_2O, pH 6.5
13	3DJB	Q6HJG6 (BuR114)	Hydrolase, HD family	223	2	Mg[4]	2.90	-	Fe, Ni	0.1 M HEPES, pH 7.5, 40% PEG 1000, 0.1 M KNO3
Class IV. Sample too dilute for PIXE (no Se signal)										
1	3D3N	F9US10 (LpR108)	Lipase/esterase	284	2	Ca	2.50	-	K, Mn	0.1 M HEPES, pH 7.5, 5% PEG 8000, 0.1 M $Ca(C_2H_3O_2)_2$
2	3DC7	D7V7L4 (LpR109)		232	3	Mg/Na[4]	2.12	-	-	0.1 M $MgSO_4$, 0.1 M Bis-Tris, pH 5.5, 16% PEG 8000

[1]Uniprot for full-length protein.

[2]Northeast Structure Genomics Consortium ID defines the domain construct.

[3]Different constructs for the same target.

[4]Presence of sodium and magnesium could not be confirmed in the experimental configuration used here.

[5]NMR derived models: in these cases the sample buffer contained 20 mM $CaCl_2$.

[6]Selenium signal was below 3 times the limit of detection, so accurate stoichiometry could not be established.

The metals contained in the PDB model are shown along with the resolution of the structural data, the PIXE identified metals (concentration above 3× the limit of detection (LoD)) with their stoichiometry in brackets, and potential metals (concentration 1-3× the LoD) from the same protein as used for the structural determination. The

crystallization conditions used to produce the PDB model are listed based on the deposited information for the model.

Table 1. Samples studied with PDB ID, gene, protein description, number of residues and chains in the asymmetric unit, together with results discussed below.

Purified protein samples were provided to the Hauptman-Woodward High-Throughput Crystallization Screening Center (HTCSC)[21] as part of the NIH supported Protein Structure Initiative (PSI) and all resulted in crystals. The samples used for the PIXE measurements represented the material recovered from the delivery robot dead volume after crystallization screening took place and were not produced specifically for this study. They consist of the protein in the protein buffer and are not contaminated by any of the components of the crystallization cocktails. All the samples incorporated seleno-methionine (SeMet) for crystallographic phasing purposes and all were in a buffer containing 5 mM dithiothreitol (DTT), 100 mM sodium chloride, 10 mM Tris-HCl pH 7.5, and in some cases 0.02% sodium azide. The structures of all samples were determined as part of the PSI, 29 by crystallography and 3 by NMR using the same protein as produced for crystallization screening. All were deposited in the PDB together, in the case of the crystallographic models, with the structure factors and the optimized crystallization conditions. Details of the samples are provided in Table 1 which also summarizes the results. Samples ranged in molecular weight from 80 to 456 kDa and were chosen at random from a set of metalloproteins screened in the regular HTCSC pipeline.

Sample preparation

The liquid samples were prepared for high-throughput PIXE studies by the addition of 10% v/v glycerol to control their viscosity and wetting properties. They were then deposited (printed) onto a 4 μm thick Prolene (polypropylene, $(C_3H_6)_n$) backing film (TF-240, Fluxana, Germany) using a non-contact array jet printer (ArrayJet Marathon microprinter arrayer Edinburgh, UK). Two 100 pl volume drops were deposited in each position onto the film which was stretched over and then attached with contact adhesive to a 75 mm × 25 mm aluminum alloy plate with five 8 mm × 8 mm square apertures. The spots were 60 μm in diameter and can be printed in up to a 12 × 9 rectangular array with their centers positioned 200 μm apart, Figure 1(a). An array can be printed on each of the five apertures in a single holder that fits onto the microbeam target ladder, Figure 1(b), enabling up to 540 different proteins to be mounted in the sample chamber in a single loading (avoiding the need to break the vacuum).

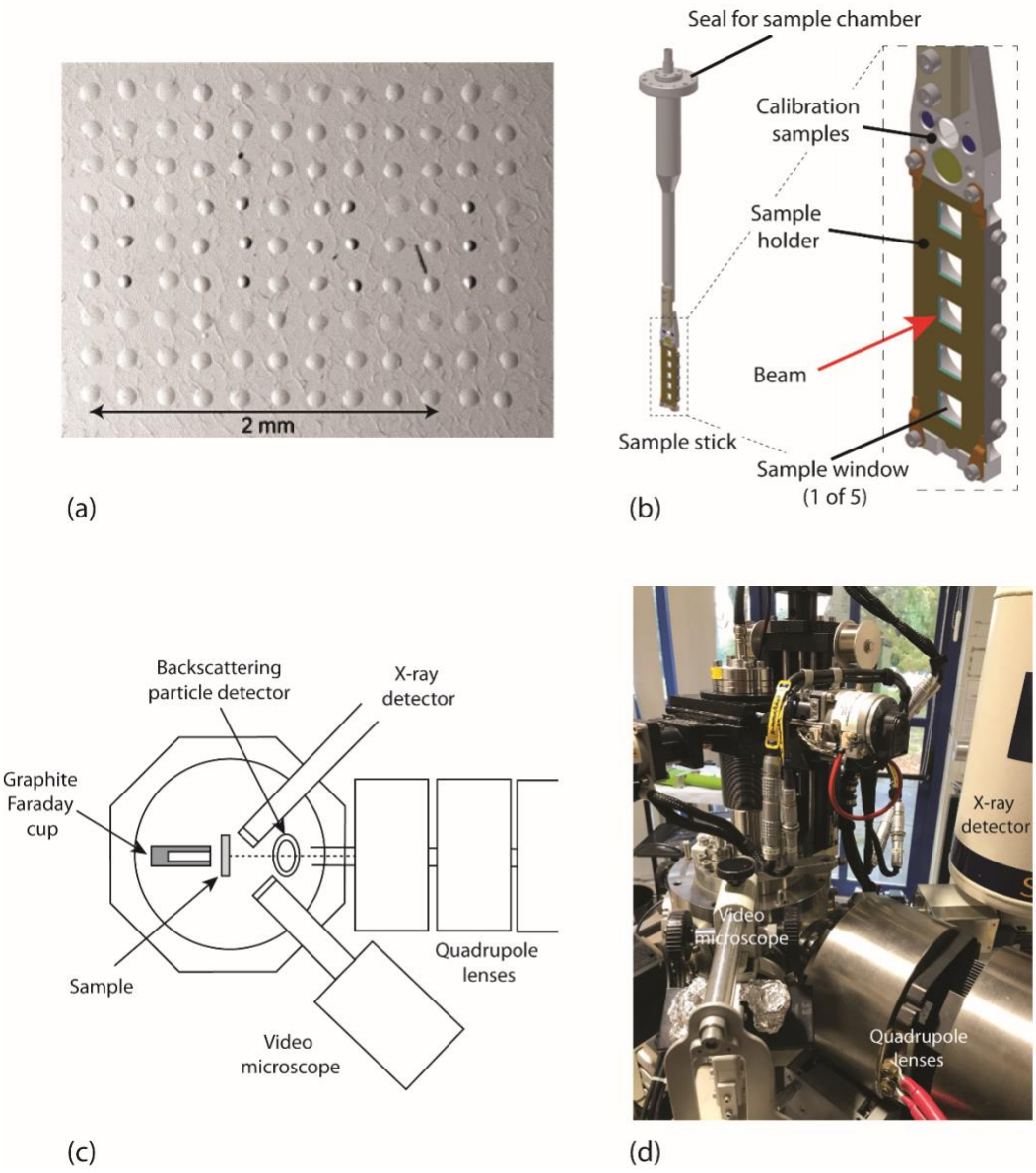


Figure 1. Key components of the experimental setup showing (a) the sample drop array, (b) a schematic of the high-throughput sample stick and sample holder, (c) a schematic of the experimental chamber and (d) a photograph of the microbeam end station at the University of Surrey Ion Beam Centre, Guildford, UK.

Data collection

The PIXE analysis was carried out using the microbeam line on the 2MV Tandetron accelerator (HVEE Corporation, NL) installed in the Stephens Laboratory of the University of Surrey Ion Beam Centre[22]. Mono-energetic protons (available energies 1.5 to 4 MeV) are focused to a diameter of < 2 μm using a magnetic quadrupole triplet lens type OM-150 (Oxford Microbeams Ltd, UK)[23]. The beam is transported under high vacuum and analysis takes place in an evacuated target chamber. The focused beam can be magnetically scanned over areas of up to 2 mm square. The beam induced X-rays are detected using a lithium drifted silicon detector (SGX Sensortech, UK) with active area 80 mm^2 and 140 eV resolution at an X-ray energy of 5.9 keV, mounted at a central angle of 135° to the beam direction in the horizontal plane and with a sample to detector distance which can be varied between 20 mm and 70 mm. Elastically backscattered protons (RBS) are detected using a PIPS charged particle detector (Canberra, now Mirion Technologies, USA) with an active area of 300 mm^2 mounted 52 mm from the sample at a central angle of 128° to the beam direction in the vertical plane. X-ray photon and particle energies are converted to amplitude-modulated voltage pulses and recorded using the OMDAQ data acquisition system (Oxford Microbeams Ltd., UK). OMDAQ also controls the magnetic beam scanning and the positioning of the motorized sample stage, allowing the recorded data pulses to be coordinated with the beam position to create spatially resolved elemental maps. A schematic and a photograph of the chamber are shown in Figures 1(c) and (d).

The HT-IBA experiments reported here were carried out using a 2.5 MeV proton beam focused to a diameter of 2 μm with a current of ~500 pA measured in a spectroscopically pure graphite Faraday cup mounted behind the samples. The X-ray detector was fitted with a 130 μm beryllium foil absorber to avoid spectrum degradation due to recoiling protons entering the detector. This also blocks the low energy X-rays emanating from Na and Mg which could not be detected under these experimental conditions. However, these elements can be measured by removing the beryllium filter and using a lower primary beam energy[24].

Due to imperfections of the printer and variations in the properties of the liquid solutions, the spots in the printed array are subject to irregularities in both shape and position. Additionally, there can be small misalignments between the axes of the printer, the sample stage motion, and the beam scanning system, so the array may appear distorted when it is mapped by PIXE. To overcome this, the array is assumed to be divided into quadrilateral cells, each containing a single spot. The array is printed with a metal salt spot, e.g. KCl, in each corner cell (fiducial marks) which are easily visible using PIXE mapping. The fiducial spots are located manually using PIXE and the centroid coordinates determined from the most intense PIXE signal using a least-squares circular feature location algorithm (CFL) implemented in OMDAQ. Using the recorded coordinates of the corner cells, the center of each cell of the array is found automatically by linear

interpolation and the sample stage is then moved to visit each cell in sequence. This procedure corrects for any linear distortion of the array pattern. To correct for spot printing errors within the cells and to optimize the utilization of beam time, an automated procedure is used in which the region occupied by each spot within the cell is located by PIXE mapping of the entire area of that cell. The CFL algorithm is then used to restrict the beam scanning area to the actual extent of the spot for the analysis exposure, thus reducing the beam time required for that spot by up at a factor of 4 and providing a significant improvement in sample throughput.

The accumulated data are monitored in real time to determine when to terminate the analysis of each spot. A Gaussian fitting procedure in OMDAQ periodically calculates the background-corrected counts (N), and hence the relative counting error ($1/\sqrt{N}$), of each of the X-ray peaks of interest in the PIXE spectrum (defined by the user). The data collection on that spot is terminated when the ratios of the element peaks of interest to the internal standard element in the sample (see below) can be determined to a user-specified accuracy, Δ i.e. $\sqrt{1/N_x + 1/N_S} < \Delta$, where x and S are the unknown and standard elements. Sulfur is normally used as the internal calibration element[25-26] but in this work, the standard sample buffer contained additional sulfur (5 mM DTT), therefore we used the selenium (as SeMet) already incorporated into the samples as the internal standard. The RBS spectrum is monitored simultaneously to ensure that the total number of detected backscattered protons is sufficient to allow the sample matrix composition and thickness to be well determined, enabling accurate correction of the PIXE data for sample matrix effects (X-ray absorption and incident particle energy loss within the sample). For each set of samples supplied to the Arrayjet printer, an input file provided by the user contains the position and identity of each spot in the printed array, the elements of interest, the number of selenium (or sulfur) atoms per macromolecule (known from the primary sequence), and the desired accuracy of the determination, Δ. This procedure is applied sequentially to the individual samples listed in the input file, automatically locating the target spot and then dwelling on the spot until the specified analytical accuracy has been achieved. Checks are in place at each stage to identify and exclude cases such as missing array spots or spots which are so badly misaligned that they cannot be uniquely located. Data collected from each spot are in the form of a 2048-channel PIXE spectrum, a 2048-channel RBS energy spectrum and a set of experimental parameters. In this experiment the run time on each spot was 5-10 minutes, so that all 32 samples could be measured in a 6 hour unattended overnight period.

Data processing

Determining accurate metal stoichiometry. Under certain conditions, ion beam analysis has been demonstrated to have a quantitative accuracy of 1% with traceability to primary standards[27], but this

requires very careful calibration of many experimental parameters such as detector solid angle and beam charge. In our method we eliminate many of these factors by measuring concentration ratios to a known 'standard' element within the same spectrum. Thus the stoichiometry of element x in a macromolecule is given by:

$$N_x = \frac{C_x}{C_S} \times \frac{M_S}{M_x} \times N_S$$

where N_x is the number of atoms of element x per protein, C_x and C_S are respectively the concentration of x and the known element to be used as the internal standard (extracted from the PIXE spectrum), M_x and M_S are respectively the atomic mass of x and the standard, and N_S is the number of standard atoms per molecule, in this case the selenium incorporated into the samples. An average 75-80% seleno-methionine substitution is reported in the sample production pipeline[18] and the crystallographic structural data examined for the samples studied here indicated complete substitution judging by the observed electron density, an observation that is not unexpected[28]. For the NMR cases, we cannot examine the electron density, The 75-80% seleno-methionine substitution sets a lower boundary and establishes the degree of any potential experimental error. The difference in quantification of the number of metals measured when assuming 100% substitution versus 80% is insignificant for our purposes, e.g. for 3DCP, protein 7 in Table 1, 100% substitution yields 3.28 Ca, 0.48 Mn, 1.20 Fe and 1.20 Co compared to 3.33 Ca, 0.51 Mn, 1.15 Fe and 1.15 Co for 80% substitution. Therefore, we used the number of fully substituted methionine residues in the protein for N_S.

The elemental ratios C_x/C_S are obtained from the PIXE spectrum using the commercially available GUPIX software package that embodies a physical model of the X-ray emission process to simulate the observed spectrum and a least squares fitting algorithm to vary the elemental concentrations in order to optimize the fit to the measured spectrum[29]. The only instrumental calibration routinely required is to determine the variation of X-ray detector efficiency with energy at the start of each run period. This is achieved by analyzing a lead-glass standard reference material (BCR-126A, European Commission, Joint Research Centre (JRC), Geel, Belgium), shown as 'calibration sample' in Figure 1(b), and adjusting the parameters in the model of the efficiency to optimize the accuracy of determination for all detected elements[30]. GUPIX applies corrections based on the thickness and composition of the sample matrix and these are determined from the simultaneously collected RBS spectrum. The RBS spectra were processed using a least squares fitting algorithm to vary the elemental concentrations in a physical model to obtain the concentrations and depth distribution of the sample matrix. This algorithm is implemented in the OMDAQ software and uses non-Rutherford proton elastic scattering cross sections obtained from the IBANDL database[31] maintained by the International Atomic Energy Agency.

Having determined the C_x/C_S ratio, the stoichiometry of the elements of interest can be calculated from the known number of atoms of the reference element, as above. The uncertainty and LoD of the measurements depend strongly on the specifics of each case but typically the stoichiometry can be determined with a standard error typically less than 0.1 atoms per molecule and with an LoD less than 0.01 atoms per molecule. This has been verified over many years from our previous low throughput work[25-26].

Optimization steps / Optimizing the method. To establish this technique we eliminated a number of factors that could affect the accuracy of the results including mixing of adjacent samples, beamstrike on other components, and potential contamination. Mixing of sample material between adjacent spots can be caused by memory effects and mistargeting in the inkjet printer nozzles. This possibility was investigated by printing an array of different metal salt solutions using the procedures detailed above, and scanning the beam vertically and horizontally across the salt array. No mixing of samples was observed above the LoD (data not shown). This check also confirmed that there was no cross talk due to halo on the focused proton beam. Beamstrike on other components in the target chamber, particularly the downstream graphite Faraday cup, could cause spurious signals in the measured spectra. This was monitored by regularly retracting the sample from the beam and scanning the beam over an area significantly larger than a single spot. The source of any rogue signal was identified and rectified. Signals from contaminants in the support film would also confound the PIXE results. The Prolene film used here was selected for its elemental purity after an extensive survey and testing of many possible thin support films[32].

Several different printing parameters were tested to optimize the suitability of the printed array for IBA. These included the number of inkjet drops per spot, the spacing of the spots, and the amount of glycerol added to the sample solutions. The conditions used above were those found to be the optimum compromise between increasing the amount of analyte, and hence the PIXE yield, and ensuring well-formed spots.

Structural modeling and refinement: For all 32 samples, a model had previously been deposited in the PDB, and the associated experimental information was downloaded. To remove subjectivity from the refinement process and ensure comparison using the same processing algorithms, the models were re-refined using PDB_REDO[33], version 7.25. The resulting model and electron density was examined using COOT[34] and any positive or negative difference density in the metal site of the *Fo-Fc* map above a 3 sigma level examined. If warranted by the electron density, the metal was replaced with that indicated by the PIXE data. The model was then re-refined using PDB_REDO and the process repeated. When refinement reached convergence, minimal manual modelling took place in COOT centered on the metal region[34]. The *R* and R_{free} were noted, with $R = \frac{\sum||F_{obs}|-|F_{calc}||}{\sum|F_{obs}|}$ where F_{obs} are the structure factors observed in the X-ray

experiment and F_{calc} are those calculated from the resulting model. *R* values were calculated using 95% of the reflections, with the computation of R_{free} using the remaining 5%.

For the tri-nuclear 3DCP model (a putative histidinol phosphatase, hisK, from *Listeria monocytogenes*) a further step was taken. From the electron density map and PIXE data it was not immediately clear which individual metal was in each of the three sites. The peak height of the electron density associated with each metal site was measured after refinement without the metals present using the CCP4 program Peakmax[35]. Since all data sets had an anomalous signal due to the presence of selenium, the predicted magnitude of the anomalous signal for potential metals, zinc, cobalt, iron, manganese, and calcium could be calculated for the incident X-ray wavelength of the structure determination (0.979 Å). This step was carried out for 3DCP and these calculated values were normalized to the selenium peak, and then successively to each of the metals in order of molecular weight (see Table 2). The experimentally measured peak heights in each of the tri-nuclear metal sites in the three individual protein chains (9 sites in total) were compared with the calculated ratios separately for each chain to determine the closest match (Table 3). This takes into account overall thermal factor differences between protein chains, but assumes that the relative differences between the sites are maintained within each individual protein chain. As the three metal sites are physically interconnected in each chain this seems a reasonable assumption.

	Z	f'' at 0.979 Å	f'' normalized to					
			Se	Zn	Co	Fe	Mn	Ca
Se	34	3.843	1.00					
Zn	30	2.478	0.64	1.00				
Co	27	1.715	0.45	0.69	1.00			
Fe	26	1.500	0.39	0.55	0.87	1.00		
Mn	25	1.303	0.34	0.53	0.76	0.87	1.00	
Ca	20	0.565	0.15	0.23	0.33	0.38	0.43	1.00

Table 2 Calculated anomalous scattering from metal elements at 0.979 Å X-ray wavelength normalized successively according to the atomic number, Z.

3DCP	Position	Measured f'' peak	Normalized measured f''	Nearest match with theory (see Table 2)	Difference (theory – measured)	Old metal (PDB)	PIXE metal
Chain A	α	9.59	1.00	1.00		Fe	Co
	β	6.83	0.71	0.76	0.05	Fe	Mn

	γ	6.81	0.71	0.76	0.05	Zn	Mn
Chain B	α	8.42	1.00	1.00		Fe	Co
	β	7.50	0.89	0.87	0.02	Fe	Fe
	γ	7.6	0.90	0.87	0.03	Zn	Fe
Chain C	α	9.11	1.00	1.00		Fe	Co
	β	7.70	0.85	0.87	0.02	Fe	Fe
	γ	6.28	0.69	0.76	0.07	Zn	Mn

The value of the nearest theoretical match to the normalized measured f'' is noted. The original metal in the PDB model is indicated, along with the one identified by comparing the normalized anomalous signal to those calculated here, as well as those found in the PIXE experiments.

Table 3. Metal sites in the three chains of 3DCP with the measured anomalous peak heights (f'') and their normalized values.

The individual metal ions so determined from the match in Table 3 were modeled and the structure refined with them in place. The program Phenix[36] was used to refine the occupancies of the Mn sites based on the stoichiometry indicated by the PIXE data. The resulting occupancy was applied to these sites and the refinement procedure with PDB_REDO noted above followed. Additional detail revealed in the improved electron density that resulted was further modeled and refined in an iterative manner.

3. Results and Discussion

PIXE and RBS spectra were recorded and processed for all samples. Figure 2 shows a PIXE spectrum for a putative histidinol phosphatase, hisK, from *Listeria monocytogenes*, PDB id 3DCP. Table 1 summarizes the sample information from the original PDB model (OM) and the PIXE experimental results.

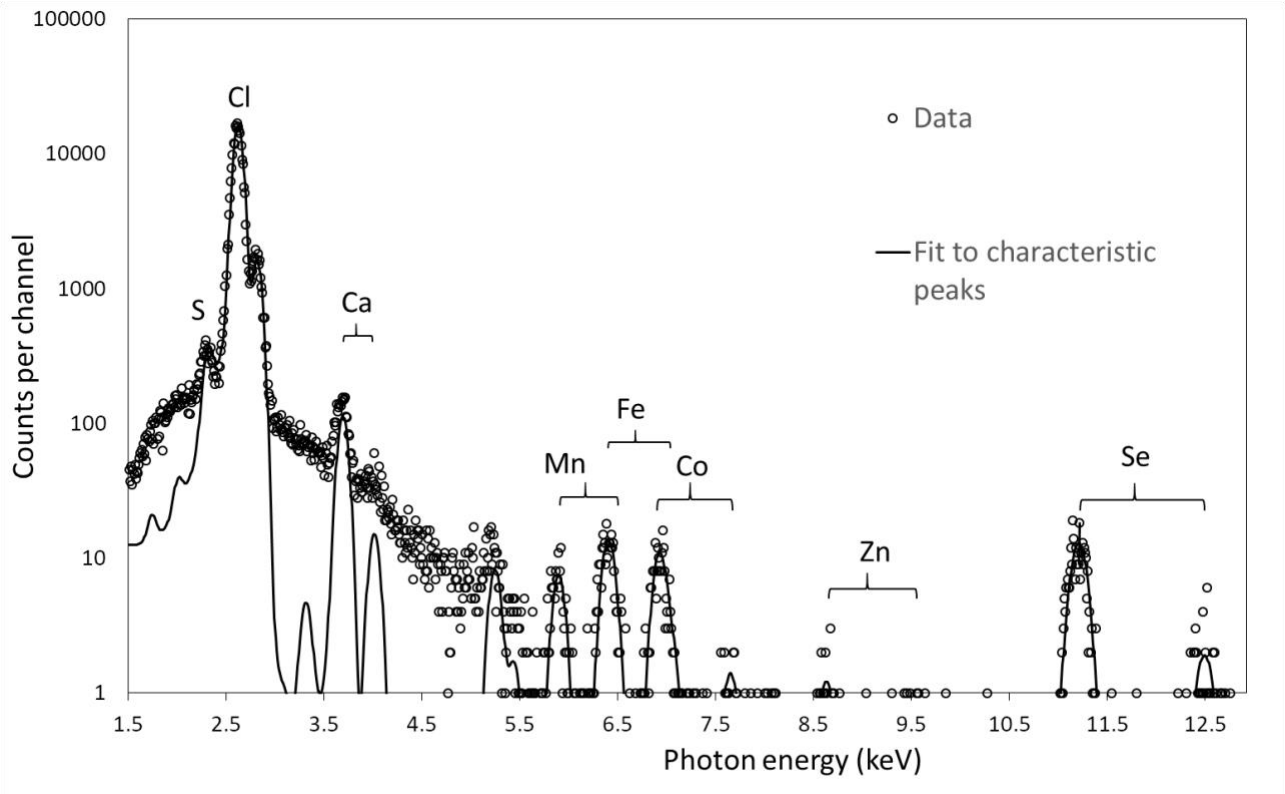


Figure 2. PIXE recorded spectrum for 3DCP, hisK, showing the data, and the fit to the individual characteristic X-ray peaks from GUPIX[29]*. The counts are on a log scale to illustrate the strong signal-to-noise in the experimental spectrum across the range of elements detected. In this spectrum each element gives two peaks resulting from the K_α and K_β electronic transitions.*

In Table 1, the 32 samples are grouped into four classes: class I, where the OM and PIXE result were inconsistent (8 samples); class II, where extra metals not in the OM were identified (9); class III, where the OM metals and PIXE data agreed (13); and class IV, where the protein sample was too dilute for successful PIXE measurement (2).

A total of 53% of the samples (classes I and II) had metals present that were inconsistent with the OM, 25% of them being completely incorrect. In class I, 5/8 samples had a metal in the OM not detected above the LoD by PIXE. However, the metal present in the OM could be explained by promiscuous metals in the crystallization conditions. The remaining three were selected for full structural refinement using the PIXE metals to replace the original assignments. Extra metals in 2/9 class II samples could be explained by the

crystallization conditions. For 12/17 samples showing metals by PIXE that were inconsistent with the OMs, the metal content could not be explained by the crystallization process and must have been present in the sample provided for crystallization screening.

For the three Class I deposited models that were studied further, the PDB_REDO based refinement statistics are shown in Table 4. When the PIXE identified metal was incorporated, PDB_REDO gave improved *R* and R_{free} for 3GGL and 3DCP. For 3KB1, the *R* and R_{free} were slightly higher than for the OM. In the case of 3DCP the *R* and R_{free} were reduced from 19.3 and 21.2% to 16.2 and 18.8%, and when refinement of the Mn occupancy was allowed in order to reflect the PIXE stoichiometry, the *R* and R_{free} further reduced to 15.5 and 18.6% respectively. The PDB_REDO models and electron density with the original metal and PIXE metal for 3GGL and 3KB1 are shown in Figures 3 and 4 respectively, generated using CCP4mg[37] with F_o-F_c maps shown as semitransparent solid surfaces.

Protein	Resolution (Å)	PDB_REDO		PIXE metal	
		R (%)	R_{free}(%)	*R* (%)	R_{free}(%)
3GGL	3.00	20.0	25.1	17.4	23.7
3KB1[1]	2.82	21.4	26.5	21.7	28.4
3DCP	2.10	19.3	21.2	15.5	18.6

[1]In the case of 3KB1, PDB-REDO appeared to inflate the *B*-factors of two ligands in the structure which led to un-modeled density and a higher *R* and R_{free} than would have been expected.

Table 4. Refinement statistics for three Class I models where the metals in the PDB were not present in the PIXE signal or obviously obtained through the crystallization cocktail.

For a C-terminal domain (277-440) of a putative chitobiase from *Bacteroides thetaiotaomicron,* PDB 3GGL, Figure 3, the F_o-F_c map shows that the OM contains lower electron density than that for a Zn atom. For the refinement against the PIXE metals, Mn and Fe, negative density was eliminated. The resulting model was redeposited in the PDB as entry 6OE2 and has an Mn coordination number of 5 compared to 4 for the OM Zn. There were no other major structural changes. The model refines equally well with an iron in the metal site, with only slightly higher *R* and R_{free} values than for Mn. Unfortunately, for this sample the Se PIXE signal was too weak for accurate stoichiometric ratios to be obtained. While the PIXE data detected no Zn above the LoD, the inability of the model refinement to distinguish between Mn and Fe in the 3 Å diffraction data illustrates the challenge in identifying metals from the electron density alone.

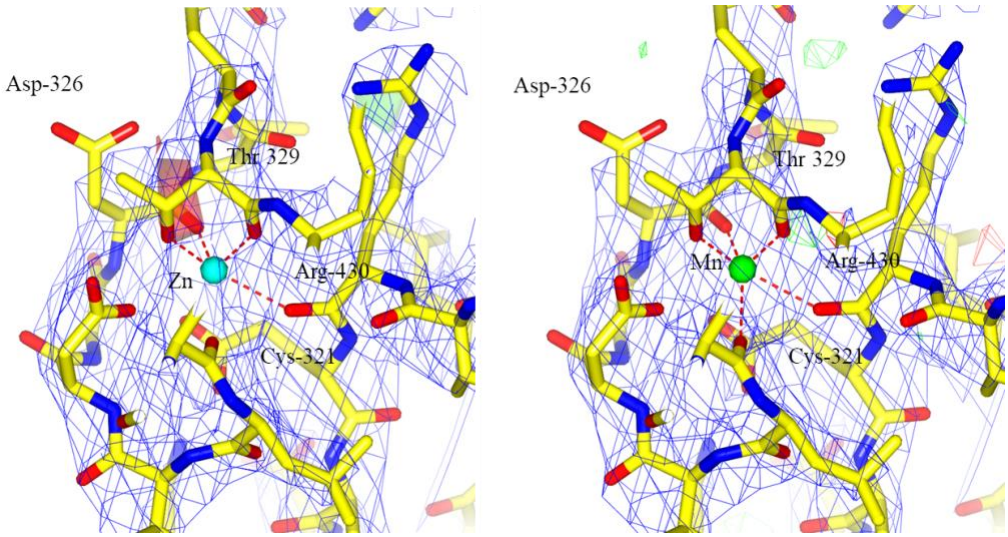


Figure 3. Structural models with electron density maps of the metal site in 3GGL. On the left, the PDB-REDO results are shown with the original metal, Zn, modelled. On the right, the metal identified by PIXE, Mn, is shown. The blue $2F_o$-F_c map is contoured at 2σ and the F_o-F_c maps are shown as a semitransparent surface contoured at 3σ and -3σ in green and red respectively. In each case selected residues are identified. Bonds to the metal ions are also illustrated. Zn and Mn have coordination numbers of 4 and 5 respectively.

For the nucleotide-binding protein AF_226 in complex with ADP from *Archaeoglobus fulgidus*, PDB 3KB1, Figure 4, the Zn built into the OM is not indicated by the electron density. On replacing this by the PIXE Co, the F_o-F_c map excess electron density is eliminated. The Co is bound to Cys 196 and 199 from chain B and Cys 196 from chain A, but the bond distance of 2.33 Å from Co to Cys 199 of chain A is outside the 2.25 Å distance expected[38] although within the standard deviation suggested for Co donor distances, especially given the limited accuracy of bond distances determined at 2.9 Å resolution. The Co binding site is interchain which may help explain why it was classed as a potential Co by PIXE, with a 1-3 × LoD, rather than a definite Co had the signal been stronger with a metal in each chain. As with 3GGL, there are no other major structural differences between the original and the revised model (PDB 6OBY).

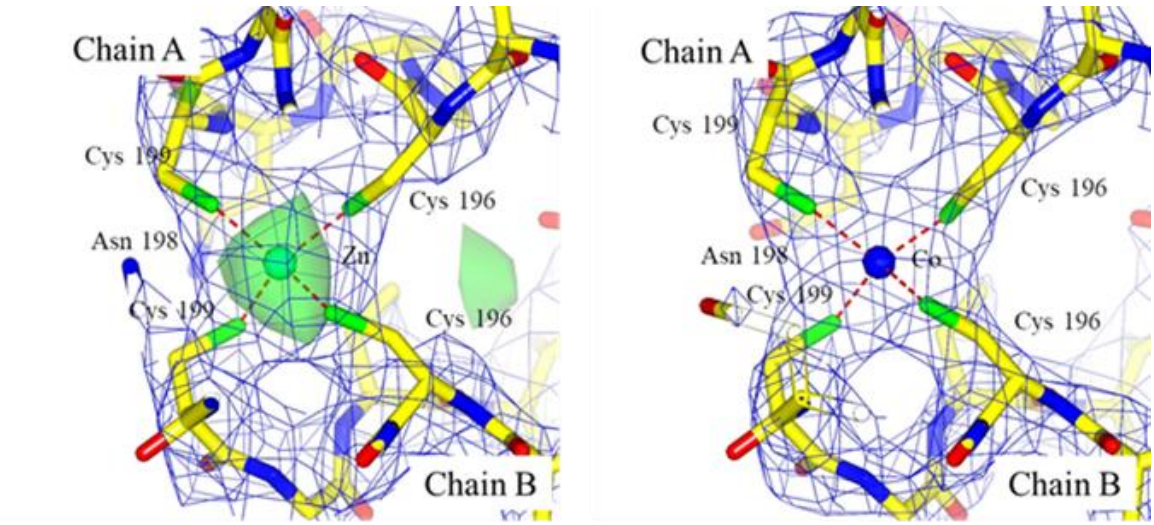


Figure 4. Structural models with electron density maps of the metal site in 3KB1. On the left, the PDB_REDO results are shown with the original metal, Zn, modelled. On the right, the metal identified by PIXE, Co, is shown. The blue $2F_o$-F_c map is contoured at 2σ and the F_o-F_c maps are shown as a semitransparent surface contoured at 3σ and -3σ as in Figure 3. In each case select residues are identified. Bonds to the metal ions are also illustrated.

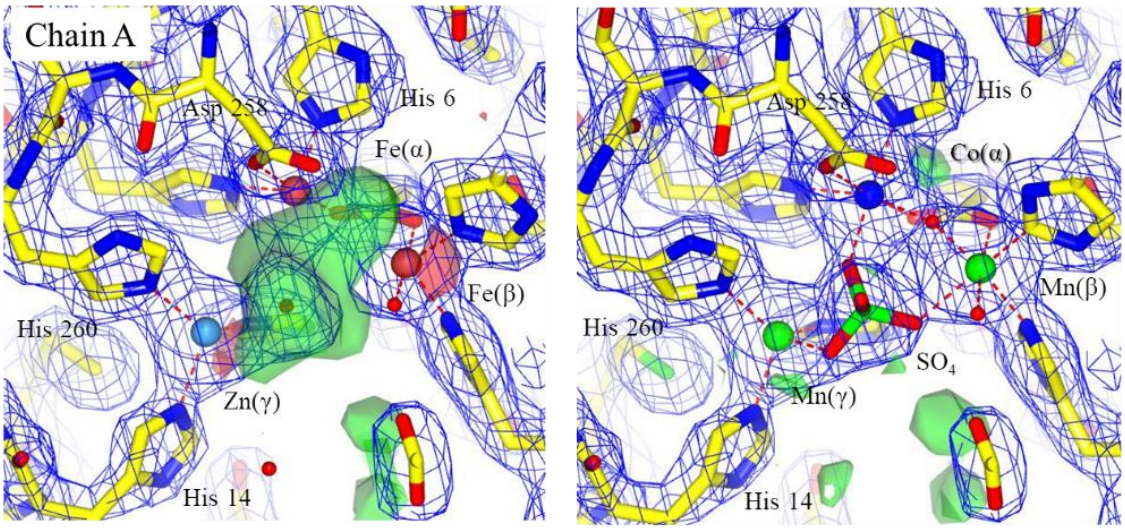
Protein 3DCP is a putative histidinol phosphatase, hisK, from *Listeria monocytogenes*. The OM of 3DCP contained excess electron density in the F_o-F_c map for the metal sites in all three of its chains, Figure 5, left. A related protein, L-histidinol phosphate phosphatase from *Lactococcus lactis* (HPP), is a member of the polymerase and histidinol phosphatase (PHP) family of proteins[39], which have tri-nuclear metal centers. They are part of the amidohydrolase superfamily (AHS) which can have either bi- or tri-nuclear metal centers. The two common metal sites are defined as α and β, with a third, γ, being seen in the PHP family.

By combining the PIXE information with the anomalous X-ray scattering signal present in the deposited structure factor data, and positional data from the 3DCP OM, the correct metals could be assigned to each position in the tri-nuclear site. The calculated values of the absorptive component, f'', at 0.979 Å incident X-ray wavelength for Se, Zn, Co, Fe, Mn, and Ca, normalized to each metal sequentially according to their atomic number are shown in Table 2. Se, present in all the proteins studied, was used to phase all the crystallographic OMs. The measured electron density peak height in the maps for each of the three tri-nuclear metal sites in 3DCP was compared to the calculated f'', and the PIXE indicated metals were placed, Table 3. In chain A, the measured peak heights are only consistent with Co, and Mn, in chain B they are compatible with Co and Fe, and in chain C they are consistent with Co, Fe, and Mn. The PIXE data yield a stoichiometry of 1.2 Fe, 1.2 Co and 0.5 Mn per protein molecule, totaling 2.9 metal atoms for the three sites

in each chain. The Mn occupancy refined to 0.83, 0.78 and 0.86 for the two A chain and one C chain sites respectively. It is likely, given the apparent promiscuity of the β and γ sites, that there could be a small proportion of Fe and Co in these sites. The resolution of the X-ray data does not allow accurate refinement of these partial occupancies, and it would require anomalous X-ray data at specific wavelengths to confirm this postulate.

The recombinant HPP enzyme expressed in *E. coli*. has a propensity for binding iron that reduces the catalytic activity. The structure of HPP was obtained from an iron-free protein expression system and crystallized with an excess of Zn^{2+}. Three structures were deposited with the enzyme containing (i) Zn^{2+} and a sulfate ion (PDB id 4GC3), (ii) Zn^{2+}, an L-histidinol, and a phosphate ion (4GYF), and (iii) Zn^{2+} L-histidinol and arsenate (4GK8)[39]. Zn is modelled in all three metal sites and at least for the sulfate and phosphate associated structures, displays a similar signature to that seen in the 3DCP OM.

3DCP (hisK) is structurally related to HPP[39] and has identical metal binding residues except for His 14 which is an Asp in HPP. 3DCP also has differences in three of the four residues interacting with the substrate, with Phe 36 replacing Tyr, Lys 87 replacing Arg, and Phe 88 replacing Tyr. In the 3DCP OM, electron density in the active site is attributed to water and there is a mixture of Fe and Zn in the tri-nuclear metal site (Figure 5 left). Re-refinement using the existing X-ray data coupled with the PIXE and anomalous X-ray signal analyses shows some of the Fe positions are correct, but that Co and Mn are present instead of Zn (Figure 5 right).



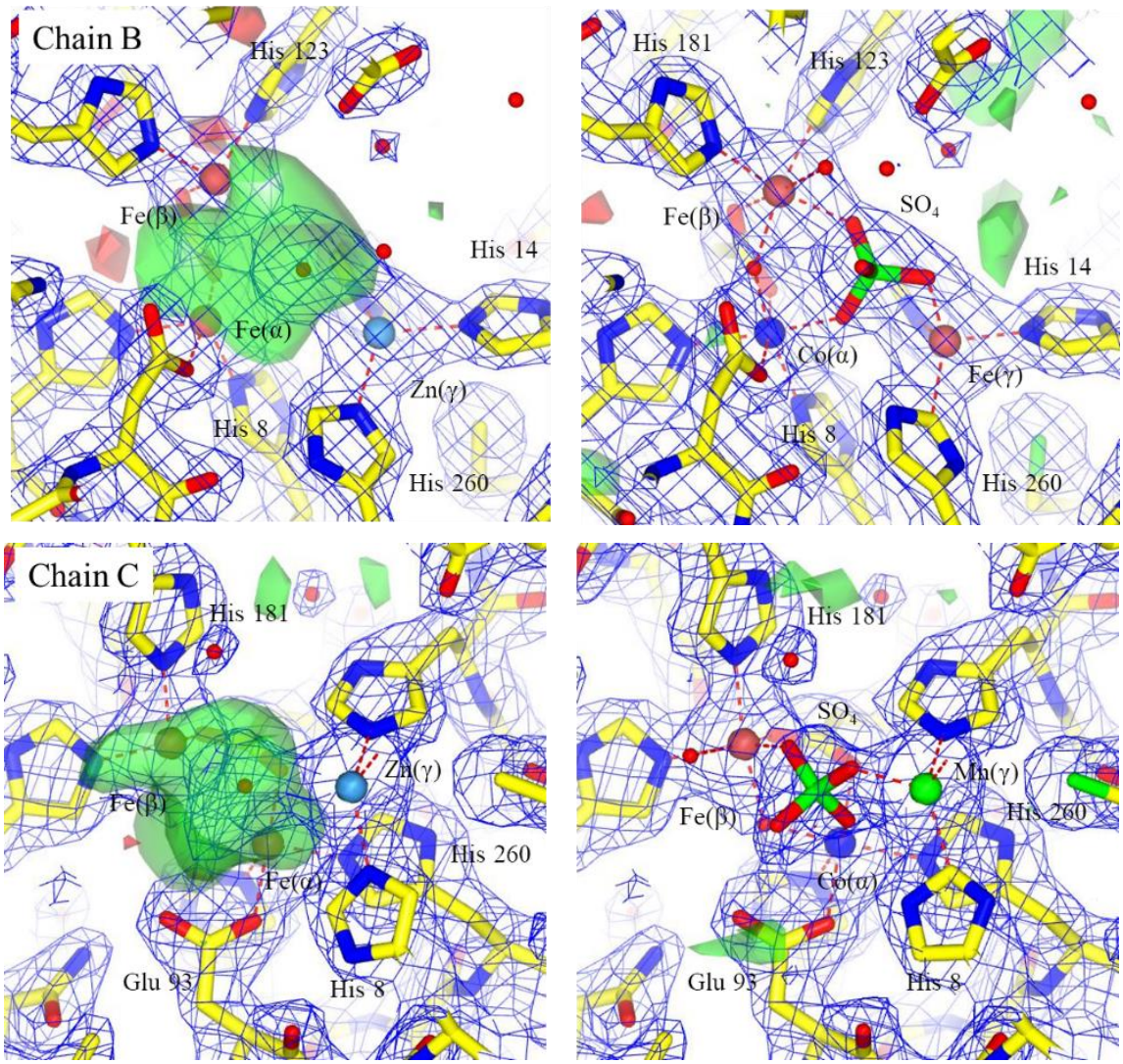


Figure 5. Structural models with electron density maps of the metal sites of all three chains of 3DCP. In the left column, the PDB_REDO results are shown with the original metals modelled. In the right column, the metals identified by PIXE and positioned using the f'' analysis (see text) are shown. The blue $2F_o$-F_c map is contoured at 2σ and the F_o-F_c maps are shown as a semitransparent surface contoured at 3 and -3σ in green and red respectively. In each case selected residues are identified. The metal ions are colored according to their identity and bonds to them illustrated. On the right, with the correct metals identified and positioned, an SO_4 ion is clearly resolved in the electron density.

HPP has an L-histidinol and a phosphate in the active site (4GYF), presumably from the hydrolyzation of the L-histidinol phosphate. In our revised model, 3DCP has a sulfate in a similar location to the phosphate in 4GYF. The presence of the sulfate in 3DCP is somewhat unsurprising given that L-cysteine was present in the crystallization solution. Under these crystallization conditions, 3DCP appears to be acting as a

sulfatase, cleaving the L-cysteine, and leaving the sulfate in a similar position and with binding distances to the sulfate as seen in HPP, PDB entry 4GC3. In the 3DCP structure there is additional electron density associated with the active site that supports the presence of L-cysteine, but not at a level to model it confidently. The revised model was redeposited in the PDB as entry 6NLR. Without extensive biochemical studies the biological relevance of the new model is unclear, we can only comment that when the correct metals were identified and modeled, a ligand is identified in the active site that was unseen in the OM.

Metal identification is particularly important for NMR structural analysis, since the bound metals are usually not directly observed in the NMR experiment. PIXE showed that all three NMR samples contained Ca (Table 1), which is unsurprising since they were in buffers containing 20 mM $CaCl_2$. Although proteins 2KPN and 4EVW were refined with bound Ca and Mg, respectively, identified by independent biochemical studies, the observation of Fe in 2KW4 and 2K52 NMR samples is interesting and unexpected.

It is unsurprising that metalloproteins incorporate metals present in the crystallization screen. There is likely to be a significant correlation between metalloprotein structure, the metal ligand, and the crystallization conditions used. Additionally, in the cases presented here, Ni and Co columns were used in the protein purification procedures, possibly providing potentially non-physiological sources for these two metals. Metal incorporation is commonly determined through a balance between metal availability in the cell, and the relative affinity of the metal to the enzyme. The Irving-Williams series[40] ranks the relative stability of complexes formed by divalent metal ions in the order $Mg^{2+} < Mn^{2+} < Fe^{2+} < Co^{2+} < Ni^{2+} < Cu^{2+} > Zn^{2+}$. The cellular concentrations of ions in the resting cell in general correlate inversely with the series with K, Mg, Mn, Fe, Co, Ca, Ni, Zn and Cu having concentrations of approximately 10^{-1}, 10^{-3}, 10^{-6}, 10^{-6}, 10^{-6}, 10^{-7}, 10^{-9}, 10^{-12} and 10^{-15} M respectively[41-42]. The metal binding preferences of most metalloproteins do not match their metal requirements. It is estimated that one third obtain their metal through metal delivery systems with the majority involving preassembled metal co-factors. The other 70% are presumed to compete for metals from buffered metal pools[41]. *In vivo*, there is competition for metal binding to other metalloproteins, and in some cases compartmentalized protein formation allows acquisition of a specific metal[43]. The metal concentrations associated with crystallization are significant compared to cellular concentrations and challenge existing metals that are often present during crystallization.

Bioinformatic approaches are available to check metals in structural models, and these methods have been reviewed[44] by the authors of a relevant software tool, CheckMyMetal. This server was used to examine all the models studied here. While useful in detecting irregularities, it was not able to separate the populations of metalloproteins experimentally flagged by PIXE. CheckMyMetal analyzes all metal-ligand bonds in the

model. It recommended alternate metals for five of the 28 metal bonds in the class I cases (from four models), four of the 28 bonds in class II (from three models), and one of the 44 bonds in class III (from one model). In no case did the suggested metal agree with the experimental PIXE identified metal. The bioinformatic analysis, while useful to direct attention to outliers from a well-defined set of metalloprotein models, did not provide the same degree of clarity as the experimental PIXE measurements.

Knowledge of the precise metal present is vital for the correct biological interpretation. Some enzymes are only catalytically active with metals that have limited bioavailability in the cell, and they acquire these preferentially over others[45-46]. Information on the presence or absence of a native metal can avoid an incorrect mechanistic understanding of the native system. However, from the experimental perspective, without a very detailed study of the metal environment of protein systems, the origin of some of the metals in the native protein cannot be traced. Metals found in protein crystals can be incorporated at many stages other than crystallization, e.g. during the expression, purification, and concentration where unexpected metal contamination could be present[47]. The true identity of metals in native proteins may remain uncertain.

4. Conclusions

We have developed a high-throughput PIXE approach and used it to define the identity and stoichiometry of metals present in a range of 30 metalloproteins. In over 50% of the randomly chosen metalloproteins, PIXE detected metals that were not in the PDB OM. The extra metal content in nine of the proteins could be explained from the crystallization conditions, but nine others contained metals that were originally misidentified. Two out of the three proteins in the latter group that were studied in detail displayed minimal structural changes when the correct metal was used to re-refine the new model against the original data. The third, hisK, was more dramatic. Building the correct metals into the model's multinuclear metal site revealed a previously unidentified ligand important for the understanding of mechanism. For the hisK multinuclear metal site, the anomalous signal in the deposited diffraction structure factor data were used to accurately place the PIXE identified metals. With information from PIXE alone, the correct metal identity could guide the collection of the appropriate anomalous X-ray signal. The overall fold of the protein was not impacted by this information, but subtle changes in the metal site, and even the presence of ligands not originally identified, can have a profound impact on understanding the overall mechanism, especially given the importance of metal ions in many biological systems.

It is salutary that over a third of the PIXE tested proteins contained the incorrect metal in the OM. The samples were taken from the High Throughput Crystallization Screening Center from those submitted as part of the protein structure initiative and are generally representative of those deposited in the PDB. An enormous number of researchers use PDB structures without knowledge of the potential fundamental errors that may be present. In 2017 there were on average 1.86 million structural downloads per day from the PDB in the US alone and over 200 data resources access and use PDB data. At the time of writing over 38% of the PDB models contain a metal[48]. Extrapolating our results (~50% misidentified or unidentified metals), over 350,000 models downloaded per day may contain incorrect or additional metals, with profound implications for those using the models as accurate representations of the structure. High throughput PIXE as implemented here could theoretically provide verification for all metalloproteins in the PDB in less than a year, since approximately 5-10 minutes per protein sample is required to collect enough PIXE data to characterize the metal content. An unattended overnight run could measure between 84 and 168 samples. By developing HT PIXE analysis methods, further information is added to structural knowledge thus improving the quality of the macromolecular models to reveal new detail and biological function in both crystallographic and NMR cases. The method presented here is needed to give accurate metal identification, stoichiometry, and position for metalloprotein systems, providing a more robust foundation to understand mechanism.

Acknowledgements

OBZ was funded by a studentship through the University of Oxford Engineering and Physical Sciences Research Council Life Science Interface Doctoral Training Centre. We thank Vladimir Palitsin for crucial technical support at the Surrey Ion Beam Centre and Nigel Saunders for early use of an Arrayjet printer at Oxford and Brunel Universities. EHS is supported by NSF grant 1231306.

References

1. Shi, W.; Chance, M. R., Metalloproteomics: forward and reverse approaches in metalloprotein structural and functional characterization. *Curr Opin Chem Biol* **2011,** *15* (1), 144-8.
2. Chen, A. Y.; Adamek, R. N.; Dick, B. L.; Credille, C. V.; Morrison, C. N.; Cohen, S. M., Targeting Metalloenzymes for Therapeutic Intervention. *Chem Rev* **2019,** *119* (2), 1323-455.
3. Zheng, H.; Chruszcz, M.; Lasota, P.; Lebioda, L.; Minor, W., Data mining of metal ion environments present in protein structures. *J Inorg Biochem* **2008,** *102* (9), 1765-76.
4. Mueller-Dieckmann, C.; Panjikar, S.; Schmidt, A.; Mueller, S.; Kuper, J.; Geerlof, A.; Wilmanns, M.; Singh, R. K.; Tucker, P. A.; Weiss, M. S., On the routine use of soft X-rays in macromolecular crystallography. Part IV. Efficient determination of anomalous substructures in biomacromolecules using longer X-ray wavelengths. *Acta Crystallogr D Biol Crystallogr* **2007,** *63* (Pt 3), 366-80.

5. Liu, Q.; Liu, Q.; Hendrickson, W. A., Robust structural analysis of native biological macromolecules from multi-crystal anomalous diffraction data. *Acta Crystallogr D Biol Crystallogr* **2013,** *69* (Pt 7), 1314-32.
6. Andreini, C.; Bertini, I.; Rosato, A., Metalloproteomes: A Bioinformatic Approach. *Accounts of Chemical Research* **2009,** *42* (10), 1471-1479.
7. Hare, D. J.; New, E. J.; de Jonge, M. D.; McColl, G., Imaging metals in biology: balancing sensitivity, selectivity and spatial resolution. *Chem Soc Rev* **2015,** *44* (17), 5941-58.
8. Finney, L.; Chishti, Y.; Khare, T.; Giometti, C.; Levina, A.; Lay, P. A.; Vogt, S., Imaging metals in proteins by combining electrophoresis with rapid x-ray fluorescence mapping. *ACS Chem Biol* **2010,** *5* (6), 577-87.
9. Robinson, C. V., Mass spectrometry: From plasma proteins to mitochondrial membranes. *Proc Natl Acad Sci U S A* **2019,** *116* (8), 2814-2820.
10. Fruncillo, S.; Trande, M.; Blanford, C. F.; Astegno, A.; Wong, L. S., A Method for Metal/Protein Stoichiometry Determination Using Thin-Film Energy Dispersive X-ray Fluorescence Spectroscopy. *Anal Chem* **2019,** *91* (18), 11502-11506.
11. Elad, N.; Bellapadrona, G.; Houben, L.; Sagi, I.; Elbaum, M., Detection of isolated protein-bound metal ions by single-particle cryo-STEM. *Proc Natl Acad Sci U S A* **2017,** *114* (42), 11139-11144.
12. Zhang, R.; Li, L.; Sultanbawa, Y.; Xu, Z. P., X-ray fluorescence imaging of metals and metalloids in biological systems. *Am J Nucl Med Mol Imaging* **2018,** *8* (3), 169-188.
13. Ugarte, M.; Grime, G. W.; Lord, G.; Geraki, K.; Collingwood, J. F.; Finnegan, M. E.; Farnfield, H.; Merchant, M.; Bailey, M. J.; Ward, N. I.; Foster, P. J.; Bishop, P. N.; Osborne, N. N., Concentration of various trace elements in the rat retina and their distribution in different structures. *Metallomics* **2012,** *4* (12), 1245-54.
14. Smart, K. E.; Kilburn, M. R.; Salter, C. J.; Smith, J. A. C.; Grovenor, C. R. M., NanoSIMS and EPMA analysis of nickel localisation in leaves of the hyperaccumulator plant Alyssum lesbiacum. *International Journal of Mass Spectrometry* **2007,** *260*, 107-114.
15. Ascone, I.; Strange, R., Biological X-ray absorption spectroscopy and metalloproteomics. *J Synchrotron Radiat* **2009,** *16* (Pt 3), 413-21.
16. Handing, K. B.; Niedzialkowska, E.; Shabalin, I. G.; Kuhn, M. L.; Zheng, H.; Minor, W., Characterizing metal-binding sites in proteins with X-ray crystallography. *Nat Protoc* **2018,** *13* (5), 1062-1090.
17. Grime, G. W., High-Energy Ion Beam Analysis. In *Reference Module in Chemistry, Molecular Sciences and Chemical Engineering.* , Reedijk, J., Ed. Elsevier: Waltham, MA, USA, 2014.
18. Xiao, R.; Anderson, S.; Aramini, J.; Belote, R.; Buchwald, W. A.; Ciccosanti, C.; Conover, K.; Everett, J. K.; Hamilton, K.; Huang, Y. J.; Janjua, H.; Jiang, M.; Kornhaber, G. J.; Lee, D. Y.; Locke, J. Y.; Ma, L. C.; Maglaqui, M.; Mao, L.; Mitra, S.; Patel, D.; Rossi, P.; Sahdev, S.; Sharma, S.; Shastry, R.; Swapna, G. V.; Tong, S. N.; Wang, D.; Wang, H.; Zhao, L.; Montelione, G. T.; Acton, T. B., The high-throughput protein sample production platform of the Northeast Structural Genomics Consortium. *J Struct Biol* **2010,** *172* (1), 21-33.
19. Acton, T. B.; Xiao, R.; Anderson, S.; Aramini, J.; Buchwald, W. A.; Ciccosanti, C.; Conover, K.; Everett, J.; Hamilton, K.; Huang, Y. J.; Janjua, H.; Kornhaber, G.; Lau, J.; Lee, D. Y.; Liu, G.; Maglaqui, M.; Ma, L.; Mao, L.; Patel, D.; Rossi, P.; Sahdev, S.; Shastry, R.; Swapna, G. V.; Tang, Y.; Tong, S.; Wang, D.; Wang, H.; Zhao, L.; Montelione, G. T., Preparation of protein samples for NMR structure, function, and small-molecule screening studies. *Methods Enzymol* **2011,** *493*, 21-60.
20. Goh, C. S.; Lan, N.; Echols, N.; Douglas, S. M.; Milburn, D.; Bertone, P.; Xiao, R.; Ma, L. C.; Zheng, D.; Wunderlich, Z.; Acton, T.; Montelione, G. T.; Gerstein, M., SPINE 2: a system for collaborative structural proteomics within a federated database framework. *Nucleic Acids Res* **2003,** *31* (11), 2833-8.
21. Luft, J. R.; Snell, E. H.; Detitta, G. T., *Expert Opin Drug Discov* **2011,** *6*, 465-480.
22. Simon, A.; Jeynes, C.; Webb, R. P.; Finnis, R.; Tabatabaian, Z.; Sellin, P. J.; Breese, M. B. H.; Fellows, D. F.; van den Broek, R.; Gwilliam, R. M., The new Surrey ion beam analysis facility,. *Nuclear*

Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms **2004**, 405-409.
23. Grime, G. W.; Dawson, M.; Marsh, M.; McArthur, I. C.; Watt, F., The Oxford submicron nuclear microscopy facility. *Nucl. Instrum. Methods B* **1991,** *54*, 52-63.
24. Graham, D. L.; Lowe, P. N.; Grime, G. W.; Marsh, M.; Rittinger, K.; Smerdon, S. J.; Gamblin, S. J.; Eccleston, J. F., MgF3—as a transition state analog of phosphoryl transfer. *Chem. Biol.* **2002,** *9*, 375–38.
25. Garman, E., Leaving no element of doubt: analysis of proteins using microPIXE. *Structure with Folding & Design* **1999,** *7* (12), R291-R299.
26. Garman, E. F.; Grime, G. W., Elemental analysis of proteins by microPIXE. *Progress in Biophysics and Molecular Biology* **2005,** *89* (2), 173-205.
27. Jeynes, C.; Barradas, N. P.; Szilágyi, E., Accurate Determination of Quantity of Material in Thin Films by Rutherford Backscattering Spectrometry. *Anal. Chem.* **2012,** *84*, 6061–6069.
28. Hendrickson, W. A.; Horton, J. R.; LeMaster, D. M., Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J.* **1990,** *9(5)*, 1665–1672.
29. Campbell, J. L.; Boyd, N. I.; Grassi, N.; Bonnick, P.; Maxwell, J. A., The Guelph PIXE software package IV. *Nuclear Instruments & Methods in Physics Research Section B-Beam Interactions with Materials and Atoms* **2010,** *268* (20), 3356-3363.
30. Gomez-Morilla, I.; Simon, A.; Simon, R.; Williams, C. T.; Kiss, A. Z.; Grime, G. W., An evaluation of the accuracy and precision of X-ray microanalysis techniques using BCR-126A glass reference material,. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, **2006,** *249* (1-2), 897-902.
31. Abriola, D.; Barradas, N. P.; Bogdanovic-Radovic, I.; Chiari, M.; Gurbich, A. F.; Jeynes, C.; Kokkoris, M.; Mayer, M.; Ramos, A. R.; Shi, L.; Vickridge, I., Development of a reference database for Ion Beam Analysis and future perspectives. *Nuclear Instruments & Methods in Physics Research Section B-Beam Interactions with Materials and Atoms* **2011,** *269* (24), 2972-2978.
32. Southworth-Davies, R. J.; Leath, K.; Grime, G. W.; Garman, E. F. In *The Characterisation of a Contaminant-free Support Film for MicroPIXE Analysis of Biological Samples*, XI International Conference on PIXE and its Analytical Applications, Puebla, Mexico, Miranda, J.; de Lucio, R. O. G., Eds. UNAM, Mexico: Puebla, Mexico, 2007; pp C2-1-4.
33. Joosten, R. P.; Long, F.; Murshudov, G. N.; Perrakis, A., The PDB_REDO server for macromolecular structure model optimization. *IUCrJ* **2014,** *1* (Pt 4), 213-20.
34. Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K., Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **2010,** *66* (Pt 4), 486-501.
35. Winn, M. D., An overview of the CCP4 project in protein crystallography: an example of a collaborative project. *J Synchrotron Radiat* **2003,** *10* (Pt 1), 23-5.
36. Adams, P. D.; Afonine, P. V.; Bunkoczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L. W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H., PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D* **2010,** *66*, 213-221.
37. McNicholas, S.; Potterton, E.; Wilson, K. S.; Noble, M. E., Presenting your structures: the CCP4mg molecular-graphics software. *Acta Crystallogr D Biol Crystallogr* **2011,** *67* (Pt 4), 386-94.
38. Harding, M. M.; Nowicki, M. W.; Walkinshaw, M. D., Metals in protein structures: a review of their principal features. *Crystallography Reviews* **2010,** *16* (4), 247-302.
39. Ghodge, S. V.; Fedorov, A. A.; Fedorov, E. V.; Hillerich, B.; Seidel, R.; Almo, S. C.; Raushel, F. M., Structural and mechanistic characterization of L-histidinol phosphate phosphatase from the polymerase and histidinol phosphatase family of proteins. *Biochemistry* **2013,** *52* (6), 1101-12.
40. Irving, H.; Williams, R. J. P., Order of Stability of Metal Complexes. *Nature* **1948,** *162* (4123), 746-747.

41. Foster, A. W.; Osman, D.; Robinson, N. J., Metal Preferences and Metallation. *J Biol Chem* **2014,** *289* (41), 28095-28103.
42. Dudev, T.; Lim, C., Competition among Metal Ions for Protein Binding Sites: Determinants of Metal Ion Selectivity in Proteins. *Chem Rev* **2014,** *114* (1), 538-556.
43. Waldron, K. J.; Rutherford, J. C.; Ford, D.; Robinson, N. J., Metalloproteins and metal sensing. *Nature* **2009,** *460* (7257), 823-830.
44. Zheng, H. P.; Chordia, M. D.; Cooper, D. R.; Chruszcz, M.; Muller, P.; Sheldrick, G. M.; Minor, W., Validation of metal-binding sites in macromolecular structures with the CheckMyMetal web server. *Nature Protocols* **2014,** *9* (1), 156-170.
45. Ragsdale, S. W., Nickel-based Enzyme Systems. *J Biol Chem* **2009,** *284* (28), 18571-5.
46. Kobayashi, M.; Shimizu, S., Cobalt proteins. *Eur J Biochem* **1999,** *261* (1), 1-9.
47. Maret, W., Metalloproteomics, metalloproteomes, and the annotation of metalloproteins. *Metallomics* **2010,** *2*, 117-125.
48. Putignano V.; Rosato A.; Banci L.; C., A., MetalPDB in 2018: a database of metal sites in biological macromolecular structures. . *Nucleic Acids Res.* **2018,** *46*, D459-D464.