The influence of promiscuous metals on metalloprotein structure: Complementary techniques to separate the good, the bad, and the ugly.
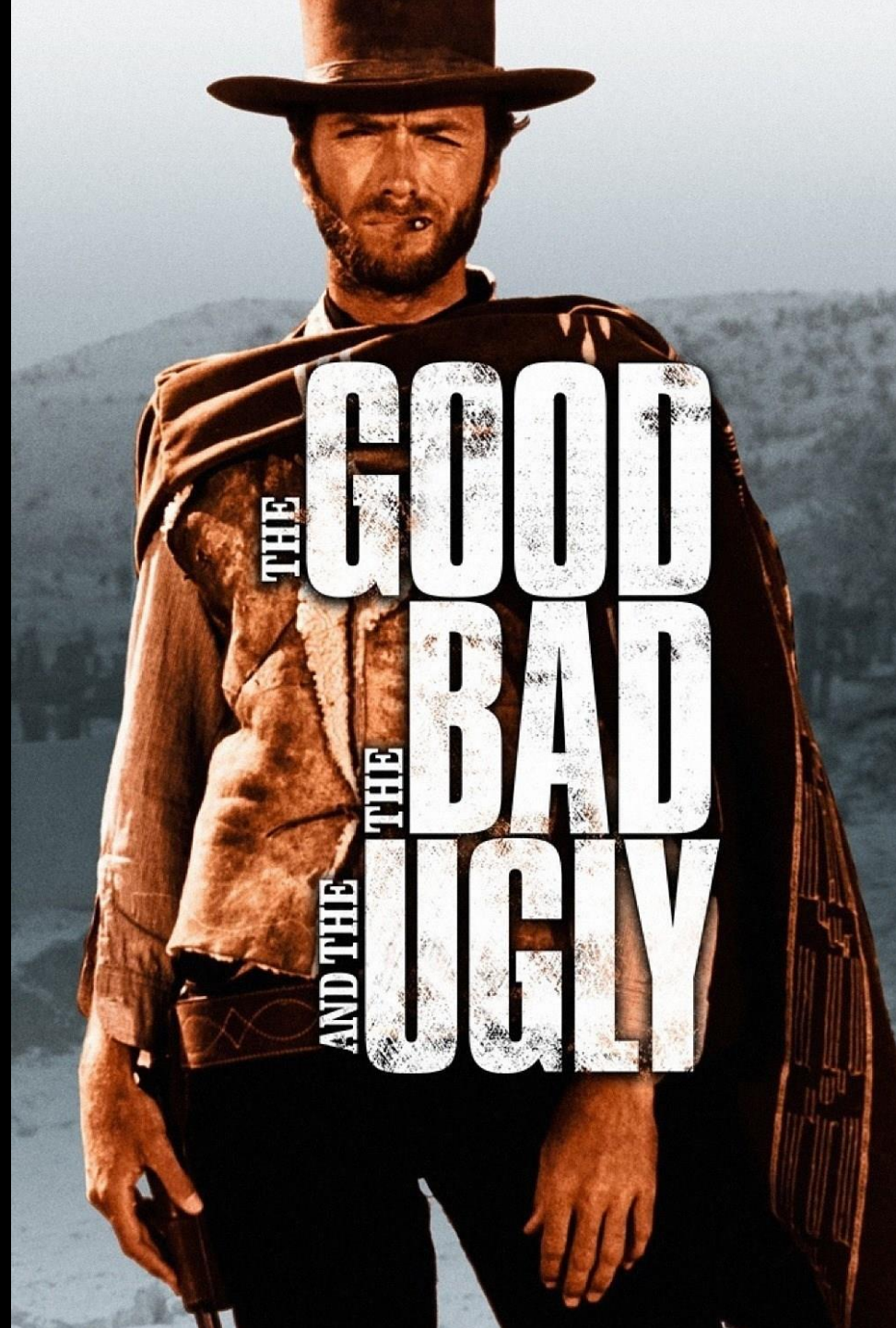
Edward H. Snell, M. Elizabeth Snell, Oliver B. Zeldin, Geoff W. Grime, and Elspeth F. Garman

Take home message:

A structure is a model.

You get out what you put in.

Think about what you did in the experiment before making conclusions

The Crystallization Screening Center at the Hauptman-Woodward Medical Research Institute

Since February of 2000 the High Throughput Crystallization Center has been screening potential crystallization conditions as a high-throughput service
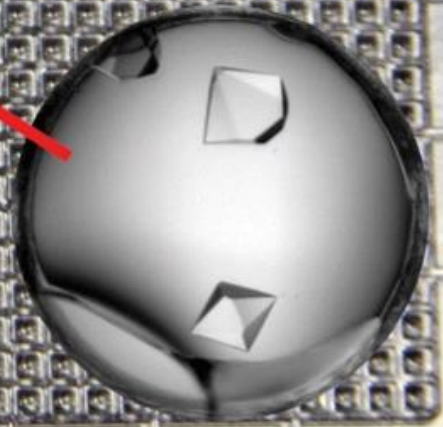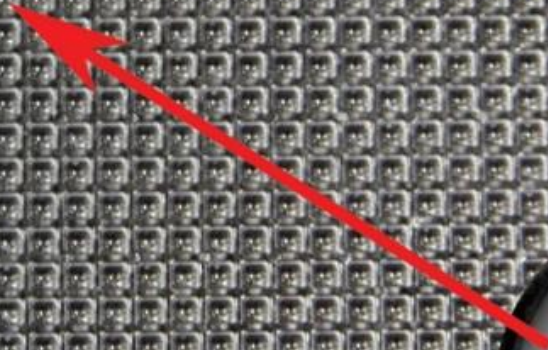
The HTS lab screens samples against three types of cocktails:

1. Buffered salt solutions varying pH, anion and cation and salt concentrations
2. Buffered PEG and salt, varying pH, PEG molecular weight and concentration and anion and cation type
3. Almost the entire Hampton Research Screening catalog.

 The HTSlab has investigated the crystallization properties of over 16,000 individual proteins archiving approximately 160 million images of crystallization experiments.

All data and in many cases, dead volume recovered samples are available

Minimize sample volume

http://Getacrystal.org

Crystallization screening, 1536 conditions including most of Hampton Research. SONICC and UV imaging.

$375 per sample

Crystal – X-ray data - Model

The Protein Data Bank

- The Protein Data Bank contains depositions for 130,365 biological macromolecules.

- Some 116,701 of those are from data derived by X-ray crystallography.

- Simple validation tests are available but a deposition can still be accepted even if a test is failed.

- How accurate are the 'structures' in the PDB?

Source of the errors in the models, if any?

- Residues have well defined geometries.

- Sequence information is well known.

- Potential problems are:
  - Structural perturbation due to radiation damage
  - Incorrect ligand identification
  - Missing ligands
  - Just generally bad refinement
  - Crystallographic oligomer
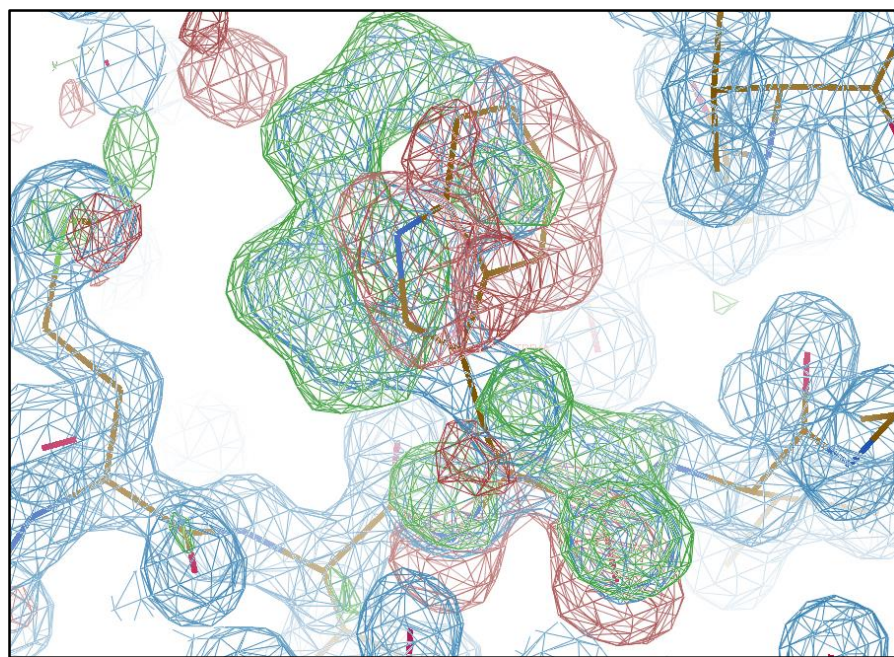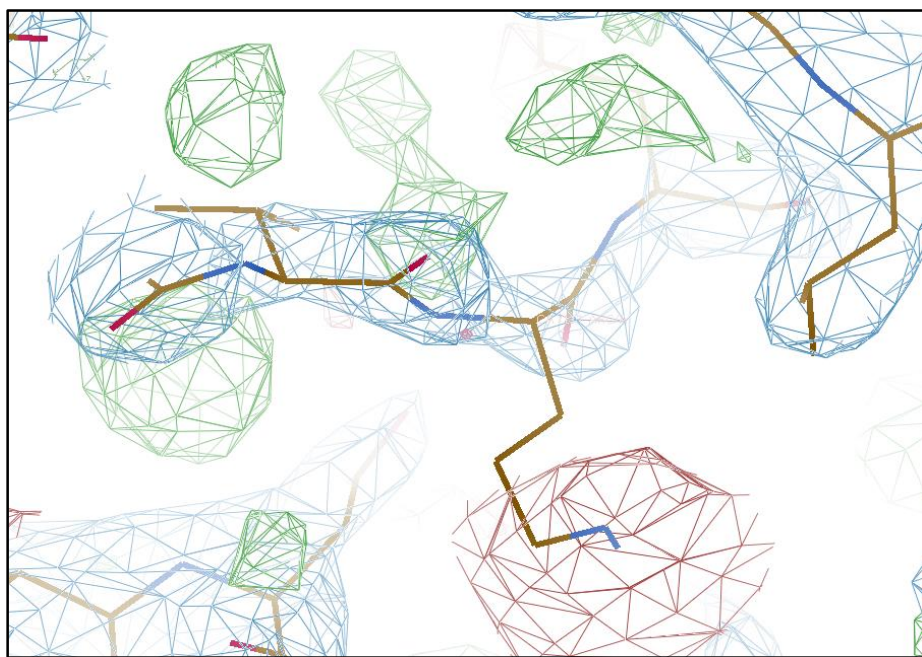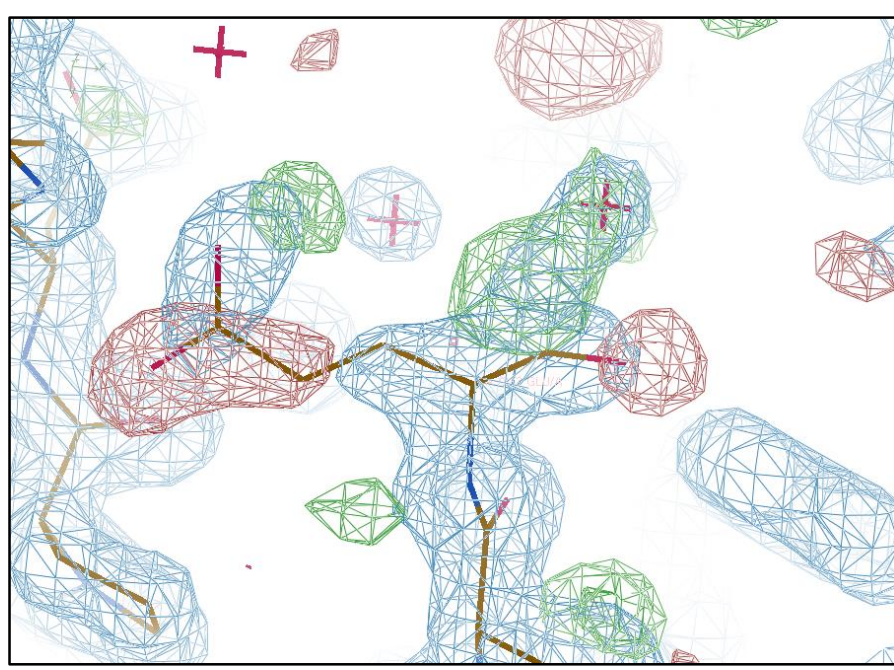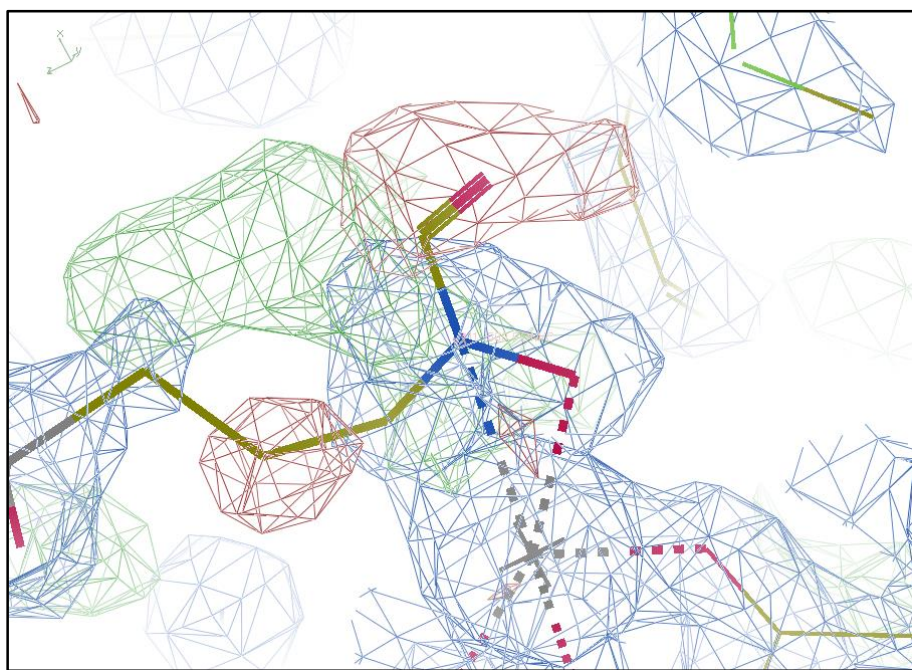
Error propagation ...

- More common than you may think

- The examples presented are in the PDB and all come from well respected structural biologists

- Despite care and diligence, errors still get through

- There are serious problems in many models yet the non-crystallographic community for the most part use these models as 'structures' on the assumption that the model accurately represents the structure

How can we over come these problems?

- Structural perturbation due to radiation damage
  - Radiation damage studies, knowledge of the chemical processes and signatures
- Incorrect ligand identification
  - Better ligand treatment during refinement
  - Careful analysis of the crystallization conditions
  - Analysis of the sample pre or post crystallization
- Missing ligands
  - Similar approaches to the above
- Just generally bad refinement
  - To paraphrase Bernard Rupp, sometimes is worthwhile to look at the map!
- Crystallographic oligomer
  - Solution scattering

How can we over come these problems?

- Structural perturbation due to radiation damage
  - Radiation damage studies, knowledge of the chemical processes and signatures
- Incorrect ligand identification
  - Better ligand treatment during refinement
  - Careful analysis of the crystallization conditions
  - **Analysis of the sample pre or post crystallization**
- Missing ligands
  - Similar approaches to the above
- Just generally bad refinement
  - To paraphrase Bernard Rupp, sometimes is worthwhile to look at the map!
- Crystallographic oligomer
  - Solution scattering

Metals in proteins

- Many proteins contain small numbers of metal atoms (estimated to be at least 30%)

  - Binding and transport of metals

  - A single metal atom helps to determine the folded shape of the molecule

X-ray crystallography measures electron density

Cannot determine Z of metal atoms

Z is often inferred indirectly from molecular modelling

Excitation scans should be used

Unknown metal atoms
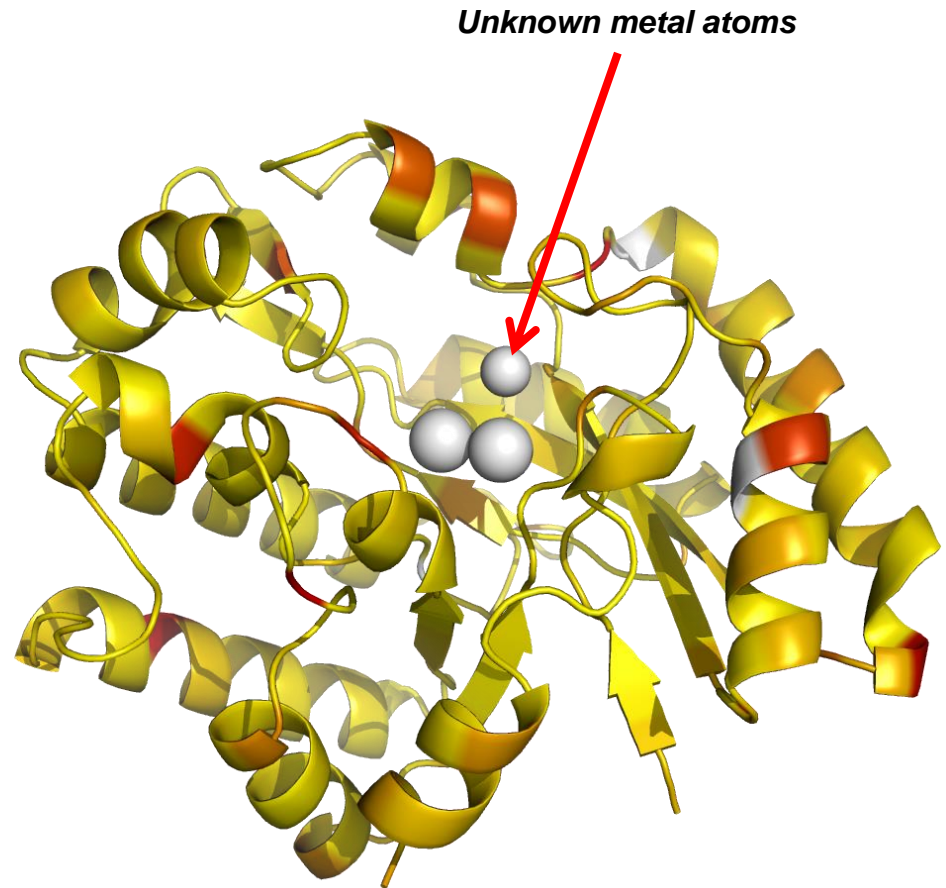
Particle induced X-ray emission

The energy of an X-ray emitted when an atomic electron undergoes an energy transition between its shell and a vacant electron site in a lower energy shell (e.g. for an M to L shell transition, sulphur gives a 2.3 keV X-ray) gives an unambiguous identification of atoms.
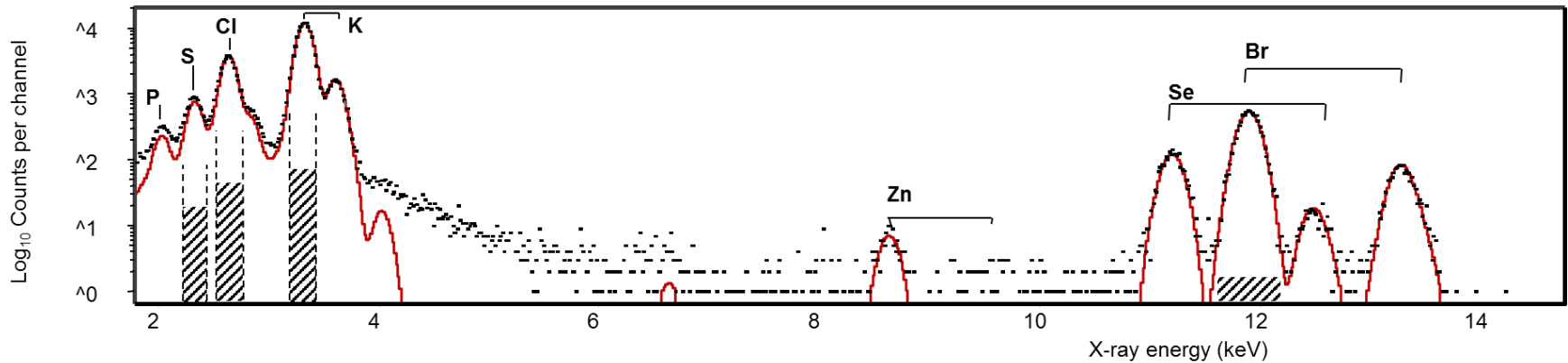
Emission of the characteristic X-rays from a sample can be induced by an incident beam of high energy protons (Particle Induced X-ray Emission: PIXE).

PIXE analysis of proteins

- Concentrations are typically 1 atom per molecule of 10 – 100 kDa (10s to 100s ppm)

- Available sample size is small (microlitres of solution)

- MicroPIXE is ideal for identifying and quantifying unknown metal atoms in proteins

- Well developed technique.

Structure 1999

Ways & Means   R291

Leaving no element of doubt: analysis of proteins using microPIXE

Elspeth Garman

Address: Laboratory of Molecular Biophysics, Department of Biochemistry, University of Oxford, Oxford OX1 3QU, UK.

E-mail: elspeth@biop.ox.ac.uk

Structure December 1999, 7:R291–R299

number. In contrast, for ionisation of atoms with electrons (EPMA) and protons (PIXE), the cross-sections for inter-actions which subsequently cause X-ray production are high for elements with low atomic numbers but decrease with increasing atomic number. Thus for identifying

UniS
University of Surrey

PIXE conference 2004

PIXE analysis of single atoms: measuring metal atom concentrations in protein molecules

G.W. Grime[a], I. Gomez-Morilla[a], D. Yates[b], E.F. Garman[b]

(a)University of Surrey, Department of Physics, Guildford U.K.
(b)University of Oxford, Department of Biochemistry, South Parks Road, Oxford, U.K.

PBMB 2005

Available online at www.sciencedirect.com

SCIENCE@DIRECT®

Progress in Biophysics & Molecular Biology

ELSEVIER   Progress in Biophysics and Molecular Biology 89 (2005) 173–205

www.elsevier.com/locate/pbiomolbio

Review

Elemental analysis of proteins by microPIXE

Elspeth F. Garman[a,*], Geoffrey W. Grime[b]

[a]Laboratory of Molecular Biophysics, Department of Biochemistry, University of Oxford, Oxford OX1 3QU, UK.
[b]Department of Physics, University of Surrey, Guildford GU2 7XH, Surrey, UK

Available online 5 November 2004

However there are issues

- Samples are prepared by manual pipetting onto foils

- Samples are analysed by manual positioning on the PIXE maps to locate the precipitated protein. Differential precipitation of buffer may require accurate positioning using elemental maps of sulphur.

- Spectra are processed manually

It is difficult to analyse more than 10 samples in a run day
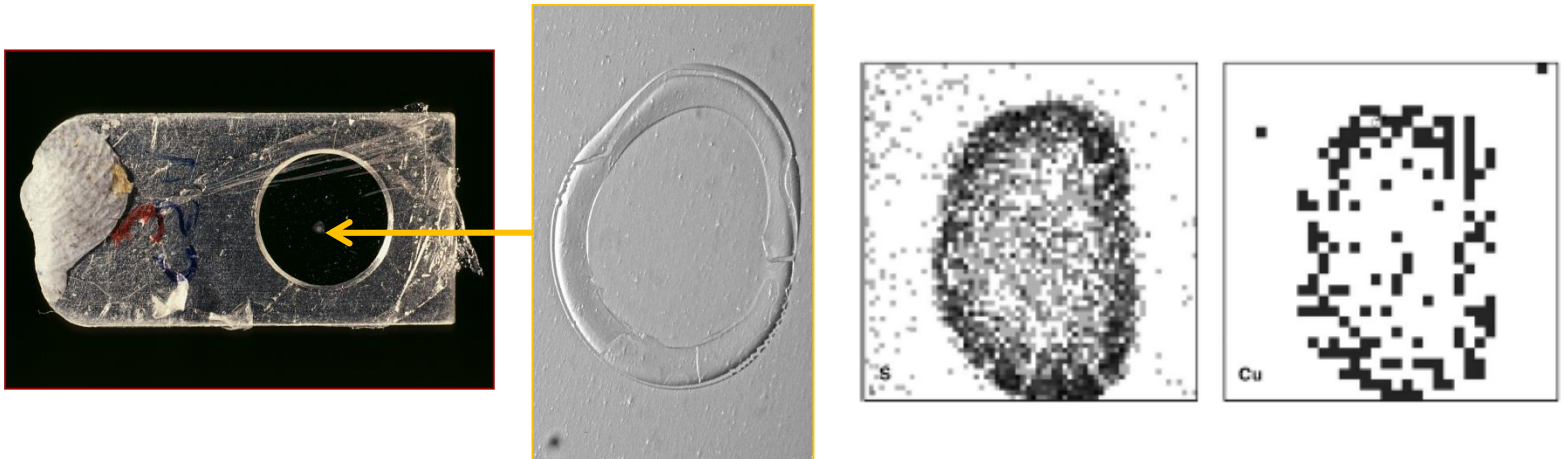
An atomic technique – can be applied to samples that are biologically 'past their sell by date'

Can we apply it in high-throughput?

Sample preparation: support film and printing
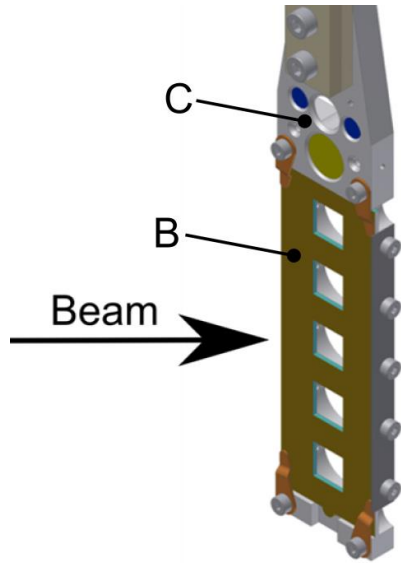
- Sample holders have the same dimensions as a standard microscope slide and are adapted for compatibility with both printer and sample stage

- Five 8 × 8 mm sample windows per slide covered in polypropylene film using a specially developed coating machine and non-instant contact adhesive

- Samples supplied as solutions in well plates

- Printed by a non-contact ArrayJet microarrayer

- Up to 144 samples per 12 × 12 array, 5 arrays per slide

High-throughput Sample Preparation

2 mm

Up to 144 samples dispensed into a 384 well plate and printed into a 12x12 array of 60 um drops with 200 um spacing.

Up to five arrays can be mounted into a single sample holder giving a total of 720 samples per slide.

Dispense samples with a non-contact microarray printer

Cross-talk?

- Array of standard salt compounds.
- Verify that there is no cross talk between adjacent spots.
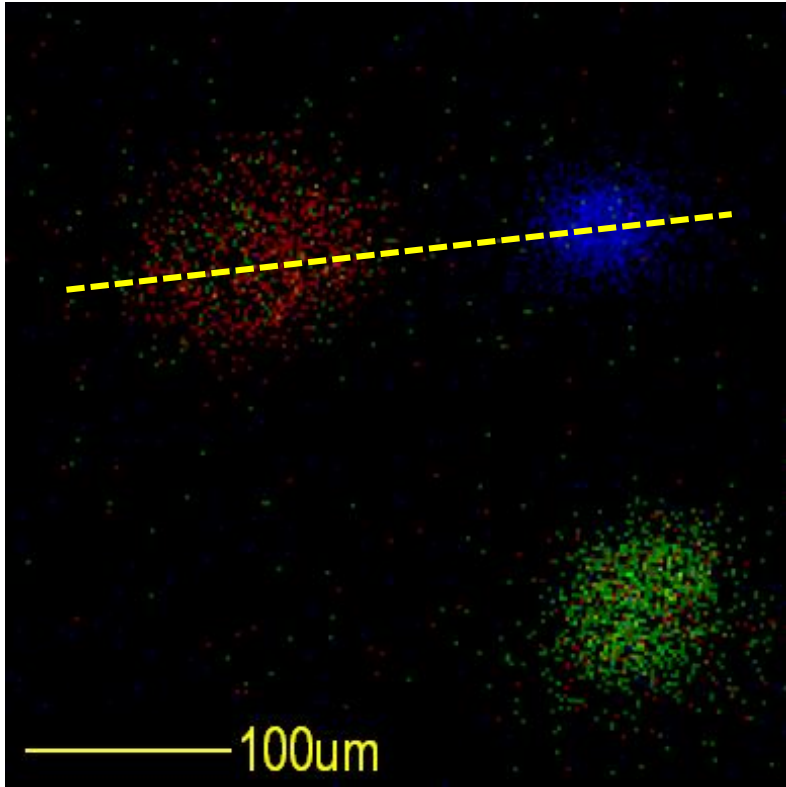
CaCl$_2$ CsI
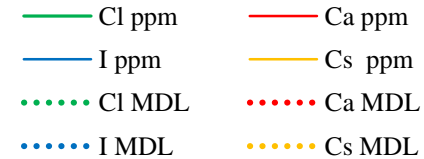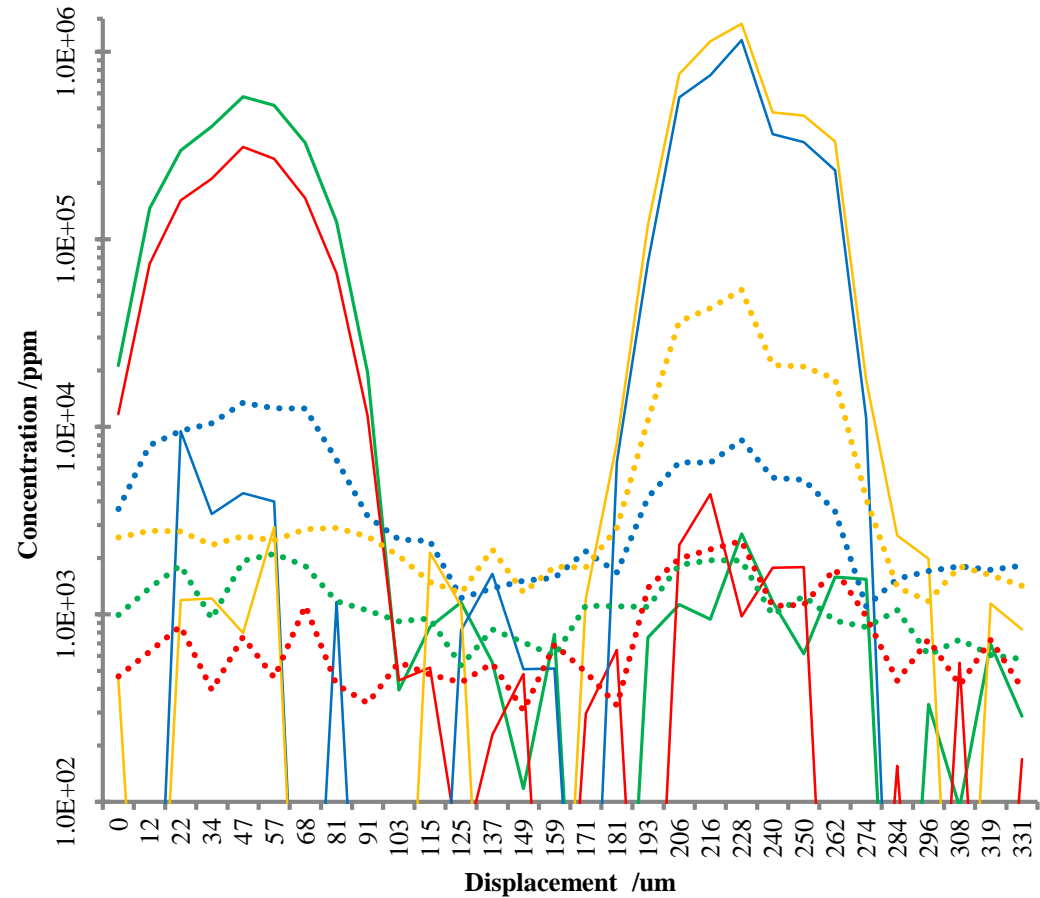
Finding the spots

1. Print '**Landing Lights**' at the corners of the array. *Spots of metal salt (e.g. KBr) which are easy to find with PIXE.*

2. Move the stage to each corner (operator control) and use a least-squares fitting routine to find the centre of the spot from the PIXE map. **(This is the only manual operation for each array)**

3. Store the stage coordinates of the corners

4. Interpolate the stage coordinates of each cell in the array. This corrects for linear geometric distortions,

S Ka1 PIXE0 (229 …

—20um   P0   S Ka1 (OM)
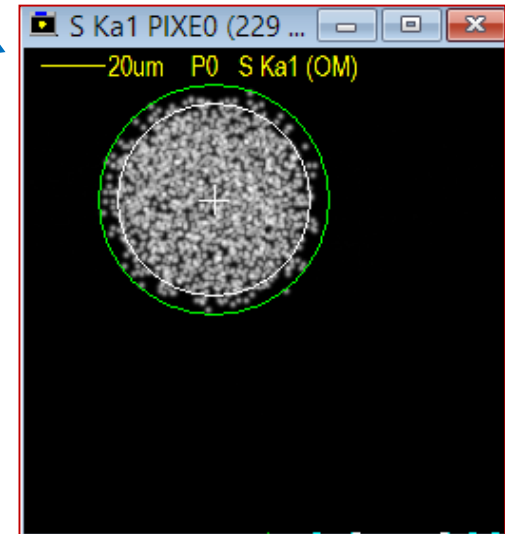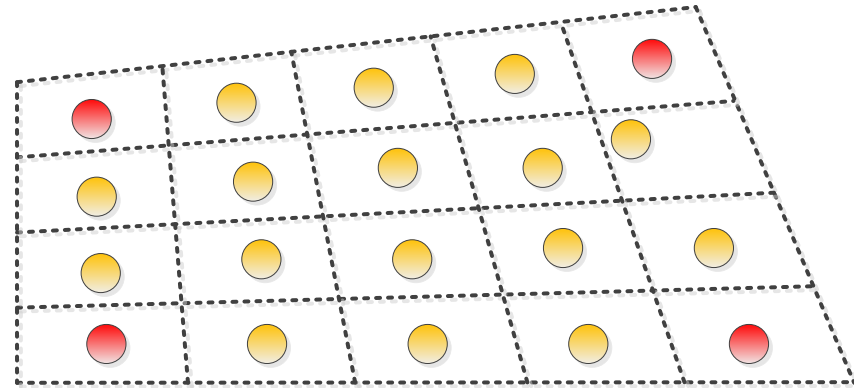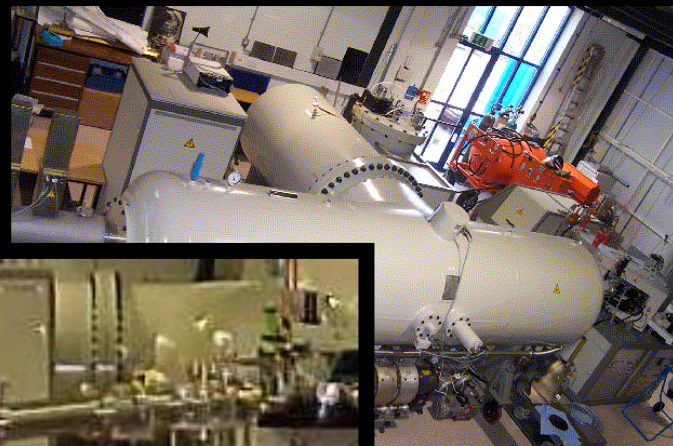
Source & Accelerator

Scanning System
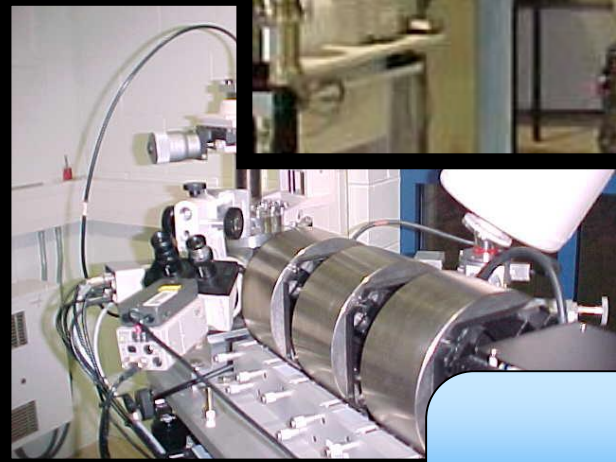
Focusing System
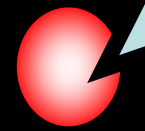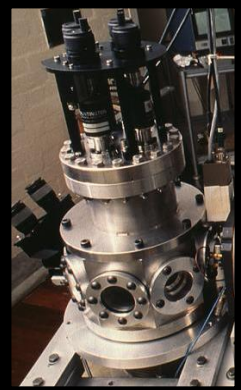
~1 μm diameter beam on target

Scanning Proton Microprobe for PIXE analysis. 2-3 MeV protons emerge from the van de Graaff accelerator and are focussed by high precision magnets onto the sample. The whole beamline is kept under vacuum.

Proton Beam
2-3 MeV
diameter ~ 1um

Si(Li) X-Ray Detector

Si Proton Detector

Vacuum

Sample in orange
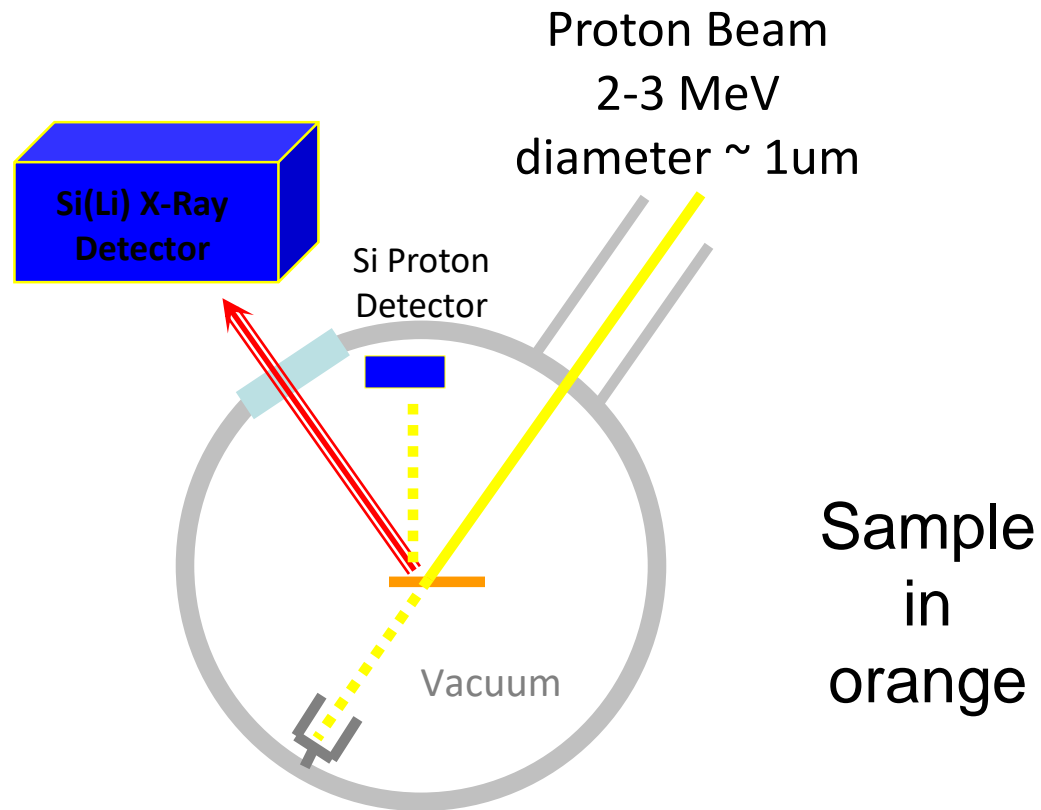
Measuring the metal content

- MicroPIXE can be used to determine the proportion of methionine substitution where no sulfur is present in the buffer.
- The concentration of an element is determined by fitting the area of the X-ray peak corresponding to the element.
- If the total number of Se atoms per protein molecule is α_{Se}, the total number of S atoms left per protein molecule is α_S, and the original number of S atoms (cysteines + methionine's) in the sequence was α then $\alpha = \alpha_S + \alpha_{Se}$ and we can write: $$\frac{\alpha_S}{\alpha_{Se}} = \frac{c_S}{c_{Se}} \frac{A_{Se}}{A_S} \frac{(\alpha - \alpha_{Se})}{\alpha_{Se}}$$

- Where A_S and A_{Se} are the atomic masses of S and Se respectively and c_S and c_{Se} are the mass concentrations.

High-Throughput PIXE

- In our case the buffer for all the samples contained Sulfur so we could not use Cys and Met as a calibration standard.
- All the proteins studied were expressed with SeMet for phasing purposes.
- The number of atoms of element Z per protein can be determined from this by

$$\alpha_Z = \frac{c_Z}{c_{Se}} \frac{A_{Se}}{A_Z} \alpha_{Se}$$

- Where A_Z and A_{Se} are the atomic masses of element Z and Se respectively and c_Z and c_{Se} are the mass concentrations determined from the PIXE spectrum.

- We do make the assumption of full Se incorporation but because we already have structural data, we can confirm this assumption.

The initial experiment

- 34 metalloprotein samples chosen from a set of samples successfully crystallized in the High-Throughput Crystallization Screening Center.
- All were SeMet samples.
- All produced crystals and a structural model.
- PIXE analysis was carried out on each sample.

- The samples used were split into four groups based on PIXE analysis
  - Those where the PDB was inconsistent with the PIXE data
  - Those where extra metals were seen in the PIXE data (but not present in the PDB)
  - Those that were consistent with the PIXE data.
  - Those that produced no signal.

Table of results

	PDB ID	Gene	Residues	Metal in PDB	Metals in PIXE (>3xLOD)	Potential metals in PIXE (1-3xLOD)
PIXE data consistent with PDB						
1		BfR258E	168	Ca	Ca (1.7)	Fe
2		BcR147A	103	Ca	Ca (0.8)	
3		HR4604D	100	Zn	Zn (2.5), Fe (0..3)	Ca, Co, Cu
4		OR3	114	Zn	Ca, Zn*	Fe, Ni*
5		LkR105	290	-	Fe (0.04)	Ca, Cu
6		MjR117B	80	-	Ca (0.2)	Fe
7		EwR179	129	-	-	Ca, Fe
8		MjR118E	105	Na◆	-	-
9		VfR176	149	Na◆	-	Co
10		SyR86	212	Na◆	-	Fe
11		SyR101A	100	Na◆	-	Ca, Fe, Cu
12		VcR193	255	Mg◆	-	-
13		DhR1A	147	Mg◆	-	Ca, Fe*
14		BuR114	223	Mg◆	-	Fe, Ni
Sample too dilute for PIXE (no S signal)						
1		LpR108	284	Ca	-	K, Mn
2		LpR109	232	Mg/Na◆	-	-

*S signal was below 3 times the limit of detection, so accurate stoichiometries could not be established.
◆Presense of sodium and magnesium could not be confirmed at the proton energies used in these experiments.

Table of results

PDB inconsistent with PIXE

PDB ID	Gene	Residues	Metal in PDB	Metals in PIXE (>3xLOD)	Potential metals in PIXE (1-3xLOD)
1	BiR14	456	Ca	-	Ca, Mn
2	NsR437I	106	Mn	-	-
3	SnR135D	161	Ca	-	Ca
4	LmR141	283	Fe/Zn	Ca (3.3), Mn (0.5), Fe (1.2), Co (1.2)	Zn
5	NsR236	119	K	-	Ca
6	NsR437H	141	Mn	-	Fe, Co
7	SoR237	137	Na	Co (0.7), Zn (0.7)	Fe, Ni
8	BtR324A	169	Zn	-	Ca, Mn, Fe*
9	GR157	262	Zn	-	Co

Extra metals present in PIXE

PDB ID	Gene	Residues	Metal in PDB	Metals in PIXE (>3xLOD)	Potential metals in PIXE (1-3xLOD)
1	MuR16	210	Fe/Zn	Fe (0.6), Co (0.9), Ni (0.4), Zn (0.7)	-
2	MqR88	420	Na◆	Ca (7.1)	Fe
3	SR677	222	Mg◆	Ca (0.7), Fe (0.05)	K/Br
4	DrR130	296	Mg◆	Ca*	-
5	BtR319D	172	Mg◆	Ca (1.74)	-
6	ShR87	320	Mg◆	Mn (0.8), Fe (0.7)	-
7	SmR83	218	Mg◆	Ca (0.5), Fe (0.1)	Ti, Co, Cu
8	NsR141	225	Mg◆	Mn (0.2), Fe (0.4), Ni (0.4)	Co
9	ZR319	289	Mg◆	-	Ca, Fe, Cu

- More than half of the proteins analysed were inconsistent with their entry in the PDB!
- This highlights a deep problem in identifying metal constituents of proteins.

Problem

- Of the 34 samples analyzed, 9 were inconsistent with the PDB results, 9 had extra metals present, 18 were consistent, and 2 were unsuitable for analysis due to low protein concentration on the sample.

- In total, 18 of the 32 analyzable samples (56%) were not correctly or fully described in the PDB deposition.

	PDB ID	Gene	Residues	Metal in PDB	Metals in PIXE (>3xLOD)	Potential metals in PIXE (1-3xLOD)	Crystallization conditions
PDB inconsistent with PIXE							
1		BiR14	456	Ca	-	Ca, Mn	18% PEG 3350, 0.2M Ca acetate, 0.1M MES, pH 6.15
2		NsR437I	106	Mn	-	-	20% PEG 4000, 0.1M Mn chloride, 0.1M MES, pH 6.0
3		SnR135D	161	Ca	-	Ca	20% PEG 8000, 0.1M Ca acetate, 0.1M MES, pH 6.0
4		LmR141	283	Fe/Zn	Ca (3.3), Mn (0.5), Fe (1.2), Co (1.2)	Zn	15% PEG 8000, 0.17 M sodium acetate, 0.01 M L-cysteine, 0.1 M MES pH 6.2
5		NsR236	119	K	-	Ca	8.64 M K acetate, 0.1 M TAPS, pH 9.0
6		NsR437H	141	Mn	-	Fe, Co	20% PEG 1000, 0.1M Mn chloride, 0.1M MES, pH 6.0
7		SoR237	137	Na	Co (0.7), Zn (0.7)	Fe, Ni	NaCl 200 mM, MES PH6, PEG 3350 20%, pH 6.15
8		BtR324A	169	Zn	-	Ca, Mn, Fe*	0.75M Mg Formate, 0.1M Bis-Tris, pH 7.0
9		GR157	262	Zn	-	Co	100 mM Na Acetate (pH 4.6), 30% MPD, and 200 mM NaCl.

Model in the PDB containing a metal from the crystallization cocktail and not protein

Model in the PDB containing an incorrect metal

	PDB ID	Gene	Residues	Metal in PDB	Metals in PIXE (>3xLOD)	Potential metals in PIXE (1-3xLOD)	Crystallization conditions
PDB inconsistent with PIXE							
1		BiR14	456	**Ca**	-	Ca, Mn	18% PEG 3350, 0.2M **Ca** acetate, 0.1M MES, pH 6.15
2		NsR437I	106	**Mn**	-	-	20% PEG 4000, 0.1M **Mn** chloride, 0.1M MES, pH 6.0
3		SnR135 D	161	**Ca**	-	Ca	20% PEG 8000, 0.1M **Ca** acetate, 0.1M MES, pH 6.0
4		LmR141	283	Fe/Zn	Ca (3.3), Mn (0.5), Fe (1.2), Co (1.2)	Zn	15% PEG 8000, 0.17 M sodium acetate, 0.01 M L-cysteine, 0.1 M MES pH 6.2
5		NsR236	119	**K**	-	Ca	8.64 M **K** acetate, 0.1 M TAPS, pH 9.0
6		NsR437 H	141	**Mn**	-	Fe, Co	20% PEG 1000, 0.1M **Mn** chloride, 0.1M MES, pH 6.0
7		SoR237	137	**Na**	Co (0.7), Zn (0.7)	Fe, Ni	**Na**Cl 200 mM, MES PH6, PEG 3350 20%, pH 6.15
8		BtR324A	169	Zn	-	Ca, Mn, Fe*	0.75M Mg Formate, 0.1M Bis-Tris, pH 7.0
9		GR157	262	Zn	-	Co	100 mM Na Acetate (pH 4.6), 30% MPD, and 200 mM NaCl.

Model in the PDB containing a metal from the crystallization cocktail and not protein

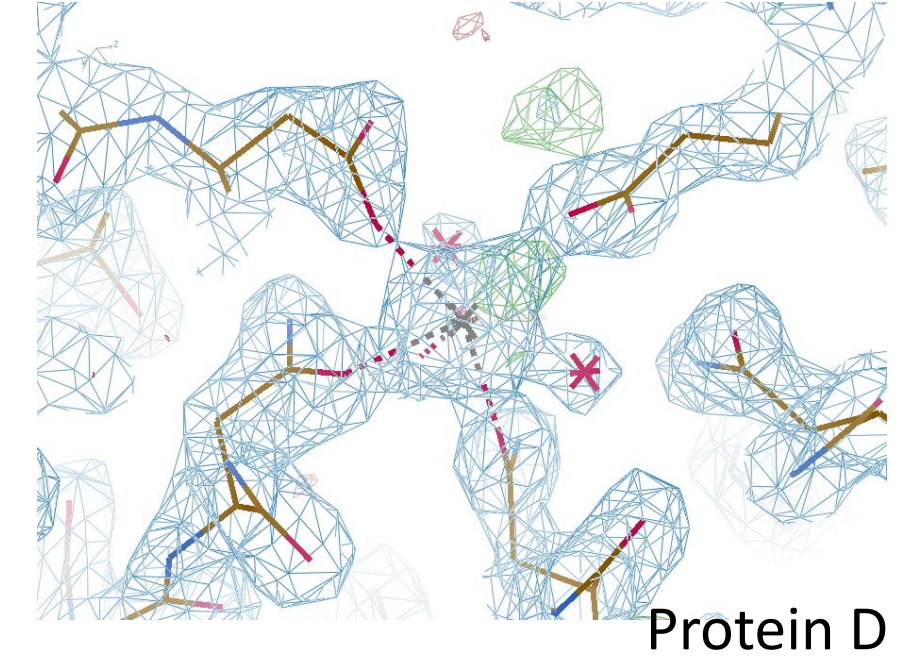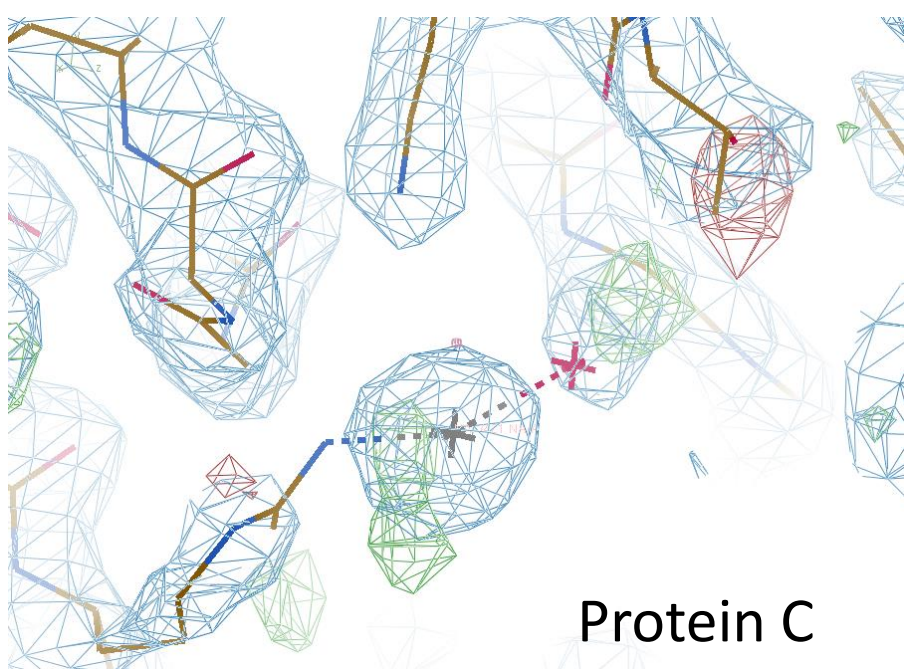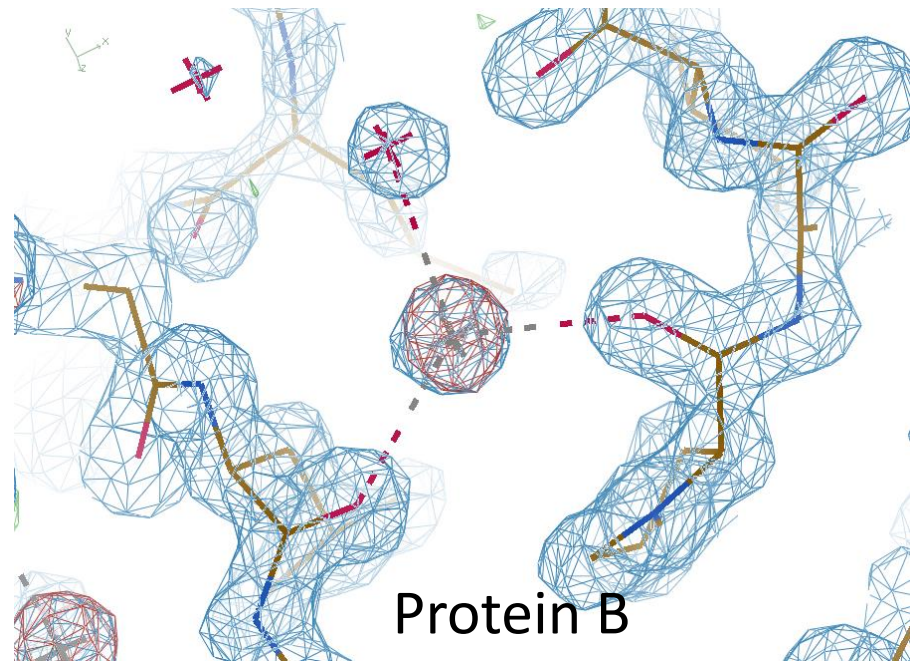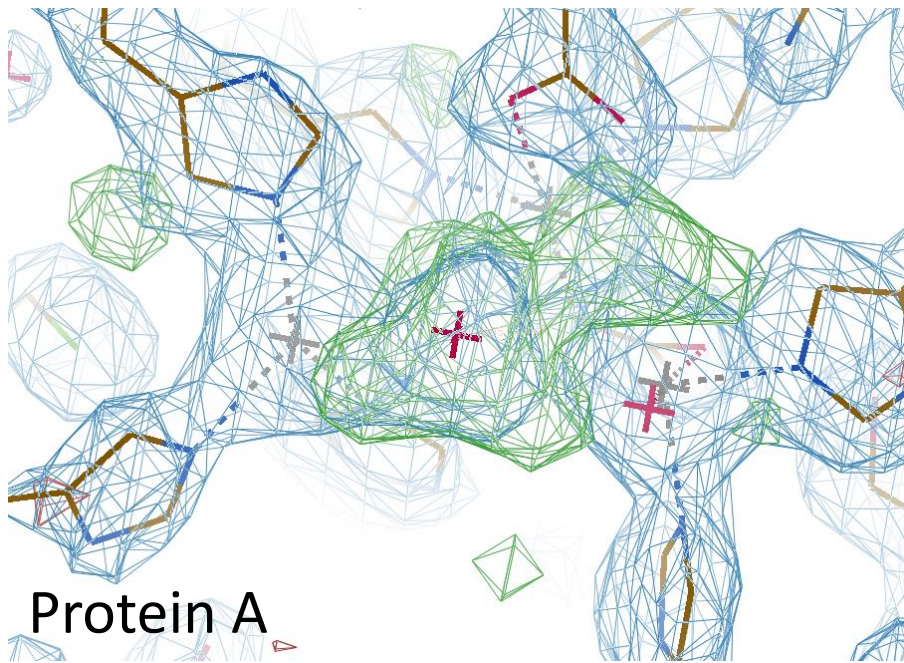Model in the PDB containing an incorrect metal

	PDB ID	Gene	Residues	Metal in PDB	Metals in PIXE (>3xLOD)	Potential metals in PIXE (1-3xLOD)	Crystallization conditions
Extra metals present in PIXE							
1		MuR16	210	Fe/Zn	Fe (0.6), Co (0.9), Ni (0.4), Zn (0.7)	-	0.1 M Na$_2$MoO$_4$*2H2O, 0.1 M Bis-Tris propane, 12% PEG 20000
2		MqR88	420	Na♦	Ca (7.1)	Fe	0.1 M Na$_2$MoO$_4$, 0.1 M Tris, pH 8.0, 20% PEG 8000
3		SR677	222	Mg♦	Ca (0.7), Fe (0.05)	K/Br	0.1 M KH$_2$PO$_4$, 0.1 M NaC$_2$H$_3$O$_2$, pH 5.0, 12% PEG 20000
4		DrR130	296	Mg♦	Ca*	-	0.1 M NaCl, 0.1 M TAPS, pH 9.0, 18% PEG 3350, MgCl$_2$
5		BtR319D	172	Mg♦	Ca (1.74)	-	None given
6		ShR87	320	Mg♦	Mn (0.8), Fe (0.7)	-	0.1 M Na citrate, pH 5.2, 1.25 M Li$_2$SO$_4$, 0.5 M (NH$_4$)$_2$SO$_4$
7		SmR83	218	Mg♦	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
8		NsR141	225	Mg♦	Mn (0.2), Fe (0.4), Ni (0.4)	Co	0.1 M citric acid, pH 5.0, 1.6 M (NH$_4$)$_2$SO$_4$
9		ZR319	289	Mg♦	-	Ca, Fe, Cu	0.1 M Tris-HCl, pH 9.1, 18% PEG 3350, 0.1 M MgSO$_4$

Model in the PDB containing an extra misidentified metal

Protein A

Protein B

Protein C

Protein D

Metric	Percentile Ranks	Value
Rfree		0.229
Clashscore		12
Ramachandran outliers		0.6%
Sidechain outliers		3.1%
RSRZ outliers		0.5%

Worse — Better

■ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution
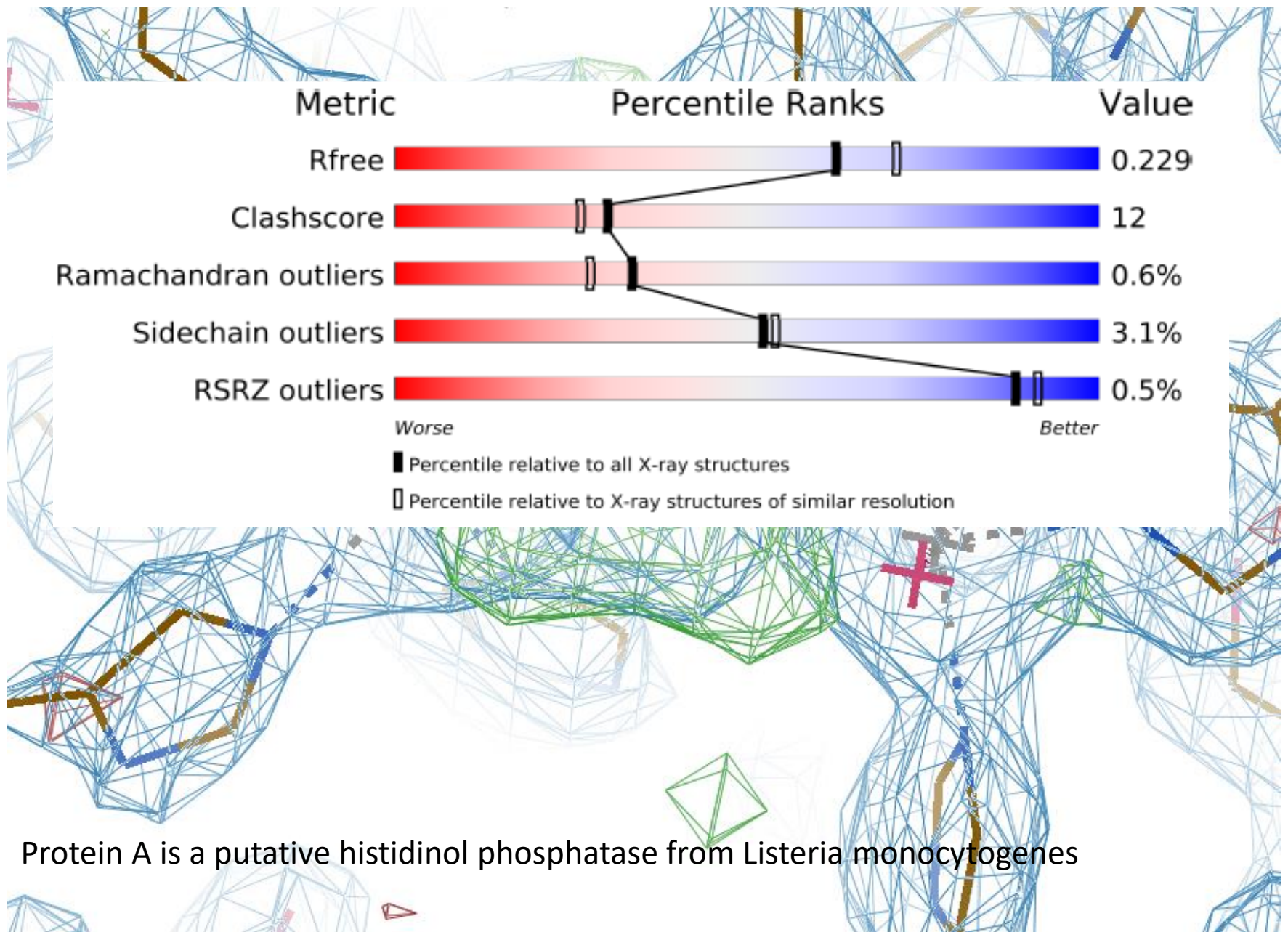
Protein A is a putative histidinol phosphatase from Listeria monocytogenes

Re-Refining Protein A

	R_{work}	R_{free}	RMS(bonds)	RMS(angles)	Clash	Ram-fav	Ram-out	Rot-out
PDB								
	0.193	0.212	0.008	1.2	11.97	96.07	0.61	
Re-refined								
	0.1847	0.2143	0.0031	0.744	1.9	96.81	0.61	2.82
Metal	Metals replaced with Co, Fe and Mn, PO_4 added in active site. Ca added in places							
	18.08	21.111	0.003	0.707	1.1	97.3	0	0
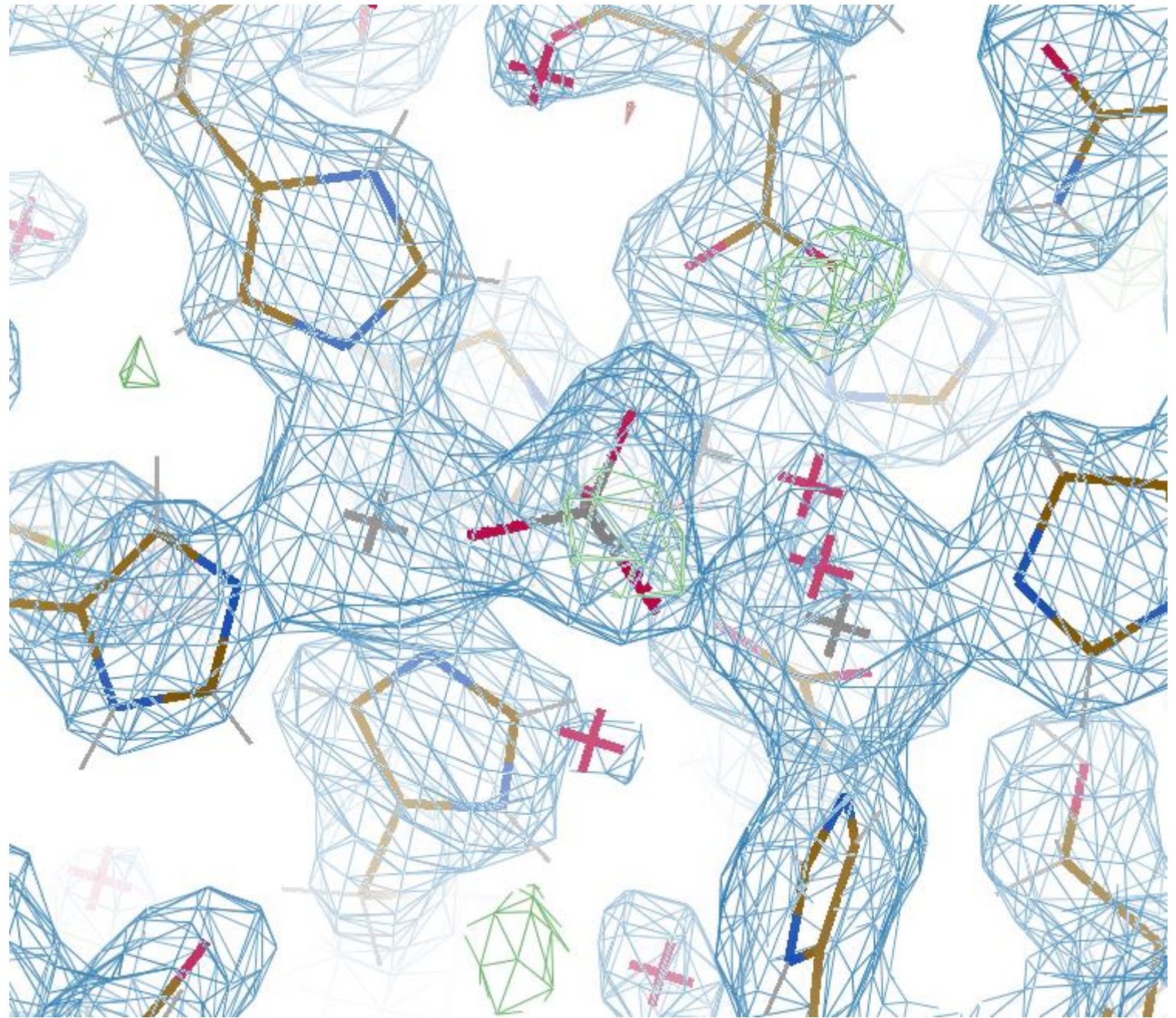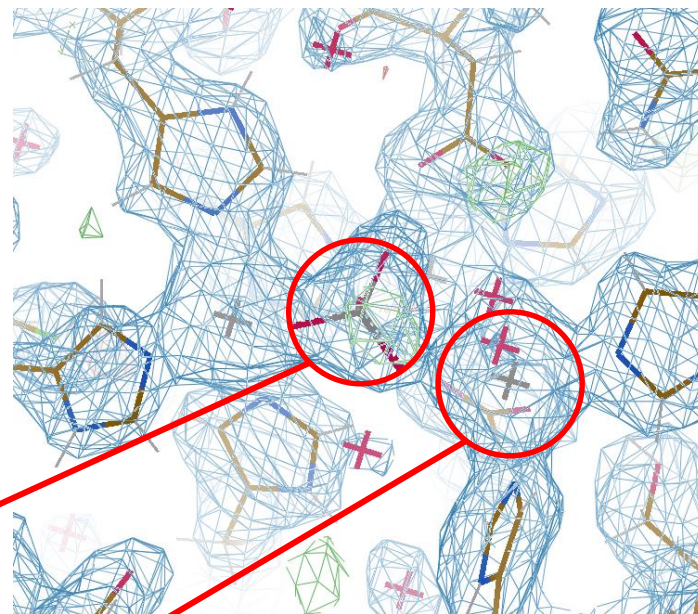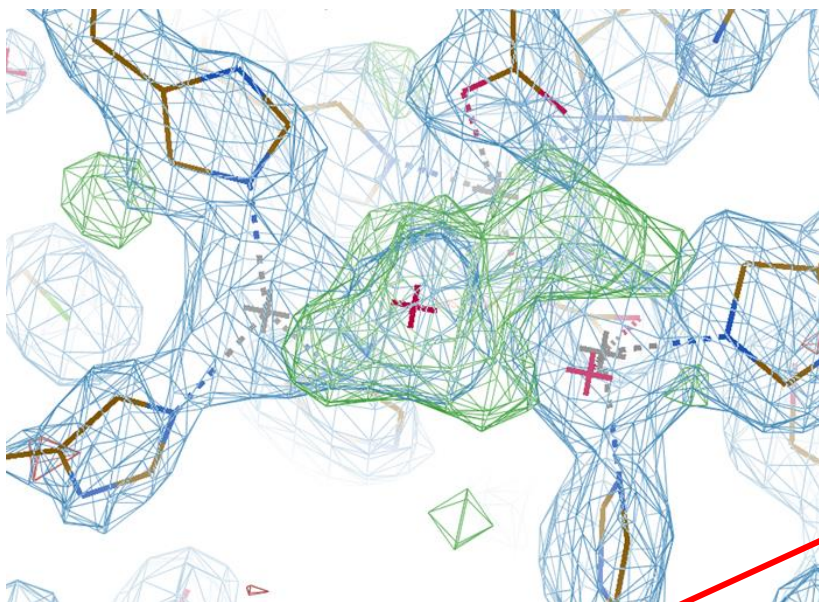
PO$_4$

Fe

Metals in new structure, Fe, Mn, Co cluster

A closely
related protein

Escherichia coli contained a mixture of iron and zinc in the active
site and had a catalytic efficiency of $\sim 10^3$ M^{-1} s^{-1}. Expression of
the protein under iron-free conditions resulted in the production of an enzyme with a 2 order of magnitude improvement in
catalytic efficiency and a mixture of zinc and manganese in the active site. Solvent isotope and viscosity effects demonstrated that
proton transfer steps and product dissociation steps are not rate-limiting. X-ray structures of HPP were determined with sulfate,
L-histidinol phosphate, and a complex of L-histidinol and arsenate bound in the active site. These crystal structures and the
catalytic properties of variants were used to identify the structural elements required for catalysis and substrate recognition by the
HPP family of enzymes within the amidohydrolase superfamily.

S *Supporting Information*

ABSTRACT: L-Histidinol phosphate phosphatase (HPP) cata-
lyzes the hydrolysis of L-histidinol phosphate to L-histidinol and
inorganic phosphate, the penultimate step in the biosynthesis of L-
histidine. HPP from the polymerase and histidinol phosphatase
(PHP) family of proteins possesses a trinuclear active site and a
distorted $(\beta/\alpha)_7$-barrel protein fold. This group of enzymes is
closely related to the amidohydrolase superfamily of enzymes.
The mechanism of phosphomonoester bond hydrolysis by the
PHP family of HPP enzymes was addressed. Recombinant HPP
from *Lactococcus lactis* subsp. *lactis* that was expressed in
Escherichia coli contained a mixture of iron and zinc in the active
site and had a catalytic efficiency of $\sim 10^3$ M^{-1} s^{-1}. Expression of

the protein under iron-free conditions resulted in the production of an enzyme with a 2 order of magnitude improvement in
catalytic efficiency and a mixture of zinc and manganese in the active site. Solvent isotope and viscosity effects demonstrated that
proton transfer steps and product dissociation steps are not rate-limiting. X-ray structures of HPP were determined with sulfate,
L-histidinol phosphate, and a complex of L-histidinol and arsenate bound in the active site. These crystal structures and the
catalytic properties of variants were used to identify the structural elements required for catalysis and substrate recognition by the
HPP family of enzymes within the amidohydrolase superfamily.

Metal content
measured with
an inductively
coupled mass
spectrometer

Accurate Metal identification is important

- The original structure contained Fe and Zn.
- The revised structure shows the phosphate and Co.
- The phosphate and tri-nuclear metal center are important for mechanism.

- Where they from crystallization? In some cases we don't know due to incomplete crystallization information.

Important notes about the technique

- Because PIXE is an elemental analysis the sample does not have to be in any preserved state.
- Samples from years ago can be used to collect experimental data.
- The number and ratio of different metals (or other atoms) per protein molecule can be determined.
- Not discussed today, but the data reveals signatures in protein models coupled with the deposited X-ray data that identify suspect metals.

Summary

- Crystallization analysis and elemental analysis have great potential in improving structural models.
- This improvement is needed as our limited study shows a an error rate of greater than 50%.
- Experimentally identifying errors defines signatures of those same errors in other structural models.
- The work leads to a potential quality control mechanism to identify suspect structural models.
- It also allows native metals (at least from expression) to be distinguished from opportune ones.

The Team

Elspeth Garman, Geoffrey Grime, Elizabeth Snell, and Oliver Zeldin

Special thanks to the 'Pixie'

Thank you and questions?

esnell@hwi.buffalo.edu