

SAD phasing using iodide ions in a high-throughput structural genomics environment

Jan Abendroth · Anna S. Gardberg · John I. Robinson ·
Jeff S. Christensen · Bart L. Staker · Peter J. Myler ·
Lance J. Stewart · Thomas E. Edwards

Received: 19 November 2010 / Accepted: 14 February 2011
© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract The Seattle Structural Genomics Center for Infectious Disease (SSGCID) focuses on the structure elucidation of potential drug targets from class A, B, and C infectious disease organisms. Many SSGCID targets are selected because they have homologs in other organisms that are validated drug targets with known structures. Thus, many SSGCID targets are expected to be solved by molecular replacement (MR), and reflective of this, all proteins are expressed in native form. However, many

community request targets do not have homologs with known structures and not all internally selected targets readily solve by MR, necessitating experimental phase determination. We have adopted the use of iodide ion soaks and single wavelength anomalous dispersion (SAD) experiments as our primary method for de novo phasing. This method uses existing native crystals and in house data collection, resulting in rapid, low cost structure determination. Iodide ions are non-toxic and soluble at molar concentrations, facilitating binding at numerous hydrophobic or positively charged sites. We have used this technique across a wide range of crystallization conditions with successful structure determination in 16 of 17 cases within the first year of use (94% success rate). Here we present a general overview of this method as well as several examples including SAD phasing of proteins with novel folds and the combined use of SAD and MR for targets with weak MR solutions. These cases highlight the straightforward and powerful method of iodide ion SAD phasing in a high-throughput structural genomics environment.

Electronic supplementary material The online version of this article (doi:10.1007/s10969-011-9101-7) contains supplementary material, which is available to authorized users.

J. Abendroth · A. S. Gardberg · J. I. Robinson ·
J. S. Christensen · B. L. Staker · L. J. Stewart ·
T. E. Edwards (✉)
Emerald BioStructures, 7869 NE Day Road West,
Bainbridge Island, WA 98110, USA
e-mail: tedwards@embios.com
URL: <http://www.ssgcid.org>

P. J. Myler
Seattle Biomedical Research Institute, Seattle,
WA 98109, USA

P. J. Myler
Departments of Global Health and Medical Education
& Biomedical Informatics, University of Washington, Seattle,
WA 98195, USA

J. Abendroth · A. S. Gardberg · J. I. Robinson ·
J. S. Christensen · B. L. Staker · P. J. Myler ·
L. J. Stewart · T. E. Edwards
Seattle Structural Genomics Center for Infectious Disease,
Seattle, WA, USA

Keywords Experimental phasing · Iodide ions · SAD · Selenomethionine · Structural genomics · Structure determination

Abbreviations

MAD	Multi-wavelength anomalous dispersion
MR	Molecular replacement
NIAID	National Institute for Allergy and Infectious Diseases
PDB	Protein Data Bank
PSI	Protein structure initiative
SAD	Single wavelength anomalous dispersion
SSGCID	Seattle Structural Genomics Center for Infectious Disease

Introduction

The mission of the Seattle Structural Genomics Center for Infectious Disease (SSGCID) is to provide a blueprint for structure-guided drug design targeting NIAID class A–C infectious disease organisms [1, 2]. To meet this goal, the SSGCID plans to solve more than five hundred crystal structures of potential drug targets from infectious disease organisms over a 5 year period. The Center for Structural Genomics of Infectious Diseases (CSGID) is a companion center of SSGCID and maintains a similar mission [3]. For SSGCID, targets are selected either through an internal target selection process or are requested by members of the scientific community external to the SSGCID consortium. The majority of internally selected targets have a homolog that is a validated drug target with a known structure, inferring that in principle most targets may be solved by molecular replacement (MR). However, in practice, not all targets solve by MR due to low sequence identity, numerous sequence homology gaps, conformation changes, etc. Furthermore, other targets are selected through statistical analysis of sequence-based annotations (Cadag, E. et al. unpublished) and numerous targets requested by the scientific community do not have a homolog with a known structure, and thus MR is not possible. Given the SSGCID protein production pipeline that generates native protein samples, we pursued strategies for obtaining de novo phase information that utilize native crystals.

Dauter and co-workers described single wavelength anomalous dispersion (SAD) phasing using iodide ions

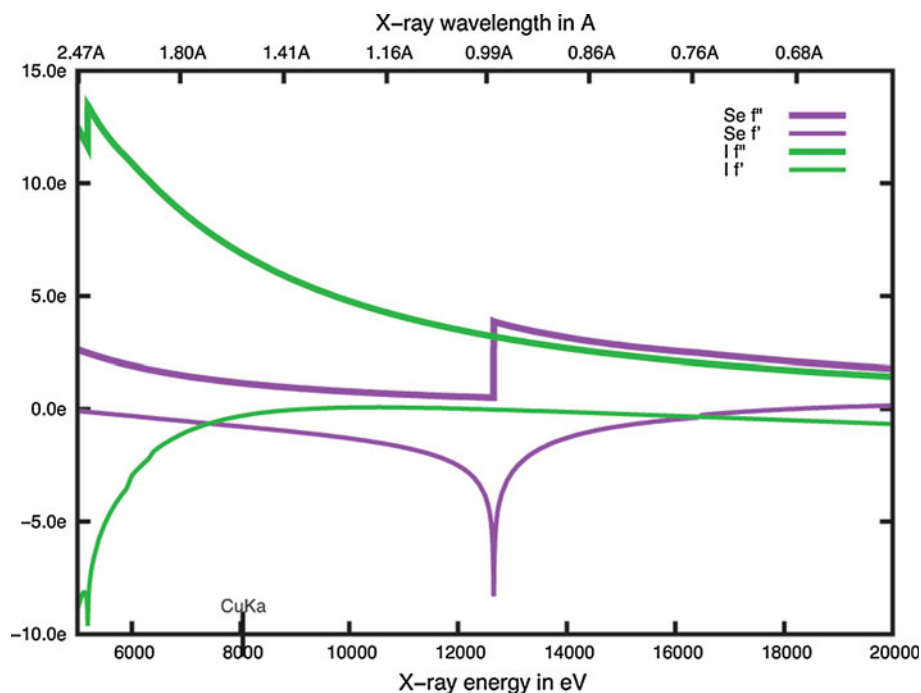
[4, 5], a method that has been successfully employed by others [6–11]. Native crystals are soaked into a solution containing high concentrations of iodide ions, data is collected in house at Cu K α radiation where the anomalous signal for iodide ions is large (Fig. 1), and the phases are estimated using a SAD experiment [12]. This method is simple, inexpensive, quick, effective, and was predicted to be “particularly suitable for high-throughput crystallographic and structural genomics projects” [8]. Here, we describe the application of iodide ion SAD phasing to a number of SSGCID targets that required experimental phase determination, resulting in sixteen new structures in 1 year. These structures are perhaps the largest collection of iodide phased structures obtained by a single scientific collaboration.

Materials and methods

Protein expression and purification

Detailed SSGCID protocols were (for example, see [7, 13, 14]) or will be published elsewhere. Here, we present a general overview of target cloning, expression, and purification. SSGCID targets were cloned using ligation independent cloning [15] from genomic DNA when available or from codon engineered synthetic genes [16, 17]. The most commonly used SSGCID expression vector (pAVA0421) encodes an N-terminal histidine affinity tag followed by the human rhinovirus 3C protease cleavage

Fig. 1 Anomalous scattering factors for iodide and selenium across the energy range used in macromolecular crystallography. The image was generated using the University of Washington X-ray Anomalous Scattering Server developed and maintained by Ethan A. Merritt (<http://skuld.bmsc.washington.edu/scatter/>) which is based on an earlier publication [59]



sequence (the entire tag is MAHHHHHHMGTLE-AQTQGPS-ORF), although other vectors were used as well. All SSGCID targets were forward and reverse sequence verified. Proteins were expressed in *E. coli* using BL21 (DE3) R3 Rosetta cells and autoinduction media [18] in a LEX bioreactor. The cells were pelleted, frozen at -80°C , and purified at one of three different purification groups, all of which used slightly variant purification protocols reflective of different equipment. Briefly, cells were re-suspended in lysis buffer, sonicated, and clarified by centrifugation. The proteins were purified initially by immobilized metal affinity chromatography. At this point in the purification protocol optional removal of the expression affinity tag was done for about 60% of all targets. The protein sample was incubated with 3C protease followed by a second nickel affinity column in which the tagless protein of interest appeared in the flow through. All protein samples used for structure determination described here were purified by size exclusion chromatography equilibrated in 20 mM HEPES pH 7.0, 300 mM NaCl, 2 mM DTT, and 5% glycerol. Fractions containing pure protein were collected, pooled, concentrated to $\sim 20\text{--}30$ mg/ml, and stored at -80°C prior to crystallization experiments.

Crystallization

Crystallization trials were set up using a rational sparse matrix approach [19] which utilized the JCSG+ and PACT crystallization screens from Emerald BioSystems or Molecular Dimensions. Sitting drop vapor diffusion crystallization trials were set up at 16°C using $0.4\text{ }\mu\text{l}$ of protein and $0.4\text{ }\mu\text{l}$ of precipitant against $80\text{ }\mu\text{l}$ of reservoir in Compact Jr 96-well crystallization plates from Emerald BioSystems. High value targets such as viral, eukaryotic, fungal or community request targets were set up in additional crystallization trials such as the ProPlex screen from Molecular Dimensions, the CSHT, Index, and Salt Rx screens from Hampton Research or the Wizard Full (I/II) and Wizard III/IV screens from Emerald BioSystems. About 9% of targets that entered crystallization trials yielded a data set with diffraction limits of $2.5\text{ }\text{\AA}$ resolution or better straight out of the primary screen. In general, targets that produced crystals that diffract to better than $\sim 3.5\text{ }\text{\AA}$ resolution but did not yield data sets suitable for structure determination were optimized using a 96-well gradient optimization screen designed and produced using the E-Wizard screen builder from Emerald BioSystems. Many targets were screened and optimized using the Microcapillary Protein Crystallization System (MPCS) [20, 21] by Emerald BioSystems. SSGCID utilized numerous salvage pathways such as in situ proteolysis [22] and seeding techniques [23]. Full crystallization conditions

for each target solved by iodide ion SAD are shown in the Supporting Information.

Iodide ion soaking

Typically, crystals were soaked for up to 5 min but occasionally as long as 2 h into a solution similar to the precipitant solution, but which was supplemented with $0.2\text{--}1$ M iodide ions (Table 1). In general, the cation used for iodide ion soaks was aligned with the cations of the crystallization solution, while the anion of the crystallization solution was replaced with iodide. Full iodide ion soaking conditions for each target solved by iodide ion SAD are shown the Supporting Information.

Data collection and structure determination

Data sets were collected in house using either a Rigaku 007-HF or Rigaku SuperBright FR-E+ X-ray generator with Osmic VariMax HF optics and a Saturn 944 or Saturn 944+ CCD detector. Diffraction images are available through the CSGID web page (www.csgid.org). Data were reduced with XDS/XSCALE [24] with the Bijvoet pairs unmerged (i.e., Friedel setting at FALSE). Sites were located using either phenix.hyss/phenix.autosol [25] or SHELXD [26]. The anomalous substructure was refined and extended, and phases were estimated using PHASER EP [27] from the CCP4 suite [28] followed by density improvement in PARROT [29]. Initial models were built using automated building in BUCCANEER [30], followed by model extension/rebuilding in ARP/wARP [31]. The model was refined using SAD refinement with optimization of the iodide ion occupancy in REFMAC [32]. Final models were produced after numerous iterative rounds of manual re-building in Coot [33] and refinement in REFMAC [32] using the merged data (i.e., Friedel setting at TRUE in XSCALE [24]). The correctness of each structure was examined, validated, and improved using Molprobit [34, 35].

Results

SAD phasing using iodide ions

In essence, SAD phasing using iodide ions is comprised of four steps. First, native crystals are soaked into a solution similar to the precipitant reservoir supplemented with iodide ions and cryoprotectant if necessary. Second, a data set is collected in house at Cu K α radiation where iodide ions have an anomalous scattering coefficient of 6.9 e^{-} (f'') (Fig. 1). Third, the iodide ion sites are located. Fourth,

Table 1 Crystal structures determined by SSGCID using iodide ion soaks and SAD phasing

PDB ID	targetDB ID ^a	Protein name ^b	Phasing resolution (Å) ^c	Precipitant	pH	Iodide	Soak time
3K9G	BobuA.01478.a	Plasmid partition protein	2.25	NaCl	6.5	1 M KI	2 × 1 h
3KM3	AnphA.00973.a	Deoxycytidine triphosphate deaminase	2.1	PEG 3350	8.1	1 M KI	1 h, 15 min
3KW3	BaheA.00339.a	Alanine racemase	2.95	PEG 3350	8.5	1 M KI	1 h
3LA9	BupsA.01663.a	BpaA trimeric autotransporter adhesin	2.05	PEG 1500	4.0	1 M KI	1 h
3LR0	BupsA.00863.i	risS periplasmic domain pH sensor at low pH	1.9	PEG 1000	4.2	1 M KI	1 h
3LR5	BupsA.00863.i	risS periplasmic domain pH sensor at neutral pH	2.3	PEG 3350	6.5	0.2 M NaI	Co-xtal
3LUZ	BaheA.00759.a	Extragenic suppressor protein suhB	2.05	PEG 400	7.5	0.75 M KI	1 h
3MD7	BrabA.11339.a	β-lactamase like protein	2.0	PEG 3350	8.6	0.25 M KI	5 min
3MEN	BupsA.10154.b	Acetylpolyamine aminohydrolase	1.9	(NH ₄) ₂ SO ₄	6.5	0.2 M NaI	1 h
3NJB	MysmA.00358.i	Enoyl-CoA hydratase	2.2	PEG 3350	8.5	0.4 M NaI	10 min
3O2E	BaboA.10365.a	Bol-A like protein	1.95	(NH ₄) ₂ SO ₄	6.5	1.0 M NaI	4 × 5 min
3OIB	MysmA.00247.b	Acyl-CoA dehydrogenase	2.1	PEG 3350	8.5	0.5 M NaI	2 min
3OL3	MysmA.17112.a	Ortholog of community request target Rv0543c	1.95	PEG 400	7.5	1 M NaI	4 min
3P96	MyavA.01155.a	Phosphoserine phosphatase SerB	2.75	PEG 6000	6.0	1 M KI	4 min
3PFD	MythA.00185.b	Acyl-CoA dehydrogenase	2.1	PEG 8000	4.2	1 M NaI	20 min
3PM6	CoimA.00345.a	Putative fructose-1,6-bisphosphate aldolase	2.5	PEG 8000	5.0	1 M NaI	15 min

^a Organism identifiers correspond to the first four letters of the targetDB ID. Anph identifies *Anaplasma phagocytophilum*, Babo identifies *Babesia bovis*, Bahe identifies *Bartonella henselae*, Bobu identifies *Borrelia burgdoerferi*, Brab identifies *Brucella melitensis* (biovar abortus), Bups identifies *Burkholderia pseudomallei*, Coim identifies *Coccidioides immitis*, Myav identifies *Mycobacterium avium*, Mysm identifies *Mycobacterium smegmatis*, and Myth identifies *Mycobacterium thermoresistibile*

^b BupsA.01663.a [7] and BupsA.00863.i were community request targets for which no homologous protein structure was available in the PDB; neither of these targets contain internal methionine residues

^c The resolution limits of the native data sets were 2.05 Å for 3KW3, 1.27 Å for 3MD7, and 2.05 Å for 3P96

experimental phases are calculated using a SAD experiment. We applied this method over a 1 year period to our structural genomics structure determination pipeline, resulting in the determination of 16 new structures from seventeen targets, a 94% success rate (Tables 1, 2; Fig. 2). Only five of the sixteen new structures followed a linear path to successful structure determination, whereas the other eleven required reiteration of one or more steps before successful structure determination.

During the first step, we typically soak protein crystals into a solution containing 1 M iodide ions, although in a number of cases, the crystals were visibly damaged upon soaking and/or diffracted poorly. Therefore, the soaking step was repeated either at lower iodide ion concentrations or in stepwise fashion across a range of iodide ion concentrations until a final concentration of 1 M was obtained. For other targets we started with a more conservative iodide concentration of 0.2 M which did not result in successful structure determination (e.g., AnphA.00973.a), whereas soaking at a higher concentration resulted in successful structure determination. In one case (PDB ID 3LR5), the protein crystallized in a condition from the PACT screen which contained 0.2 M NaI, and thus no soaking was required.

In the data collection step, 360° of data were collected for all data sets in order to maximize the multiplicity of the data [8, 36], which should lower the error in the measurement of Bijvoet pairs and thereby increase the accuracy in the measurement of anomalous signal [37]. During this step we discovered that the selection of the scaling resolution had an influence in at least two cases for which inclusion of all of the data did not lead to a successful structure determination. For example, cutting the data at lower resolution limits (2.5 Å) led to the successful structure determination of a putative fructose-1,6-bisphosphate aldolase from *Coccidioides immitis* (PDB ID 3PM6), whereas the data could not be solved at 2.2 Å resolution. Others have reported similar results for phase calculation and main chain building at 3 Å, followed by refinement at 2.36 Å resolution [8]. The authors of the program SHELXD [26] indicate that for anomalous-substructure searches truncation of the data to 3.0–3.5 Å resolution may be critical for some sub-structure solutions [38]. For these data sets, we did not observe any evidence for radiation damage, even for crystals collected with long exposure times.

Unlike selenomethionine-labeled samples, the number of iodide ion sites is not known a priori. Thus, locating the heavy atom sites in the third step is an iterative process using

Table 2 Crystallographic and phasing statistics for iodide SAD phased SSGCID structures

PDB ID	Phasing resolution (Å)	Space group	Multiplicity ^a	SigAno ^b	Residues in AU	Solvent content (%)	Iodide sites ^c	FOM ^d	Method
3K9G	2.25	<i>P</i> 4 ₃ 2 ₁ 2	12.5	1.43	1 × 267	51	2/12	0.45	SAD
3KM3	2.1	<i>H</i> 3	5.6	1.16	2 × 206	42	13/16	0.24	SAD
3KW3	2.95	<i>C</i> 2	3.8	1.26	2 × 372	45	4/0	0.29	SAD/MR
3LA9	2.05	<i>H</i> 3	5.4	2.08	1 × 178	29 ^e	4/9	0.49	SAD
3LR0	1.9	<i>P</i> 3 ₂ 2 ₁	11.2	2.33	1 × 123	46	6/6	0.44	SAD
3LR5	2.3	<i>P</i> 2 ₁ 2 ₁ 2 ₁	3.6	0.99	2 × 123	43	4/5	0.39	SAD
3LUZ	2.05	<i>P</i> 2 ₁	3.8	1.22	2 × 267	37	6/13	0.50	SAD/MR
3MD7	2.00	<i>C</i> 222 ₁	6.5	1.34	2 × 272	40	11/0	0.49	SAD
3MEN	1.9	<i>P</i> 2 ₁ 2 ₁ 2 ₁	6.8	1.24	4 × 341	43	24/35	0.41	SAD
3NJB	2.2	<i>I</i> 23	10.9	1.27	2 × 333	59	9/50	0.39	SAD
3O2E	1.95	<i>P</i> 4 ₁ 2 ₁ 2	10.3	1.70	1 × 90	25	5/9	0.53	SAD
3OIB	2.1	<i>C</i> 2	3.7	1.78	2 × 403	48	15/50	0.55	SAD
3OL3	1.95	<i>P</i> 2 ₁ 2 ₁ 2 ₁	6.6	1.91	2 × 103	51	9/21	0.43	SAD
3P96	2.75	<i>I</i> 222	7.7	1.70	1 × 418	53	15/0	0.41	SAD
3PFD	2.1	<i>P</i> 2 ₁	3.7	1.19	4 × 389	48	8/109	0.53	SAD/MR
3PM6	2.5	<i>P</i> 2 ₁	3.8	1.64	2 × 302	46	5/31	0.46	SAD/MR

^a Overall multiplicity for anomalous scaled data; multiplicity for merged data is ca. twofold higher

^b SigAno is the mean anomalous difference in units of its estimated standard deviation ($|F^+ - F^-|/\sigma$) [24]

^c The number of iodide sites input into Phaser EP [27], followed by the number of iodide sites in the final model. For 3KW3, 3MD7 and 3P96 only the high-resolution native data was deposited into the PDB, and thus the final model contained no iodide ion sites

^d Overall Figure of Merit (FOM) from Phaser EP [27] prior to density modification

^e The solvent content was calculated based on the full length tagged protein. This protein only crystallized using in situ proteolysis with chymotrypsin [22], and thus, the true solvent content is likely higher [7]

one or more computational programs. We found this step to be the most challenging for targets with a large asymmetric unit because the programs compare the sites from multiple solutions as a quality indicator. For example, acyl-CoA dehydrogenase from *Mycobacterium thermoresistibile* (PDB ID 3PFD) contained over one hundred iodide ions in the final model (all of which had electron density above 5σ in an anomalous difference Fourier map). Challenging the programs to find 10 sites only returned 1 or 2 sites, because it could not find the same 10 sites out of more than 100 total sites. However, challenging the programs to find a higher number of sites resulted in a successful structure determination. In contrast, crystals of the trimeric autotransporter adhesin BpaA from *Burkholderia pseudomallei* were twinned, and thus the number of sites was kept to a minimum to avoid selecting strong sites from the minor twin fraction [7]. In that case, selecting six to ten sites yielded low quality experimental electron density maps, whereas selecting only two or four sites led to successful structure determination, despite nine iodide ions in the final model. Based on the number of amino acids per iodide ion for successfully determined structures (Table 2), we recommend challenging the programs to find one iodide ion per twenty amino acids of the projected scattering mass.

The fourth step, SAD phase calculation, is also an iterative process. After density improvement, the experimental electron density maps were inspected for tertiary structure such as α -helices and β -sheets as well as solvent channels and correlation between iodide sites and the experimental electron density. Any iodide sites which did not correlate with the experimental electron density maps were eliminated and the phases were re-calculated with only the real iodide ion sites. This reiteration often improved the phase quality and consequently resulted in more extensive automated structure building. For targets with a weak MR solution from which a structure could not be determined directly, the weak MR solution could be incorporated into the SAD experiment. This combined use of SAD and MR has been successful in four cases (Table 2 and see below). In these cases, the increased number of sites identified by combined SAD/MR could be used in a subsequent SAD experiment with no MR component to calculate the phases.

Success with a diverse set of proteins, crystallization conditions, and crystal forms

The proteins used in these experiments derived from Gram negative bacteria, fungi, and eukaryotes (Table 1).

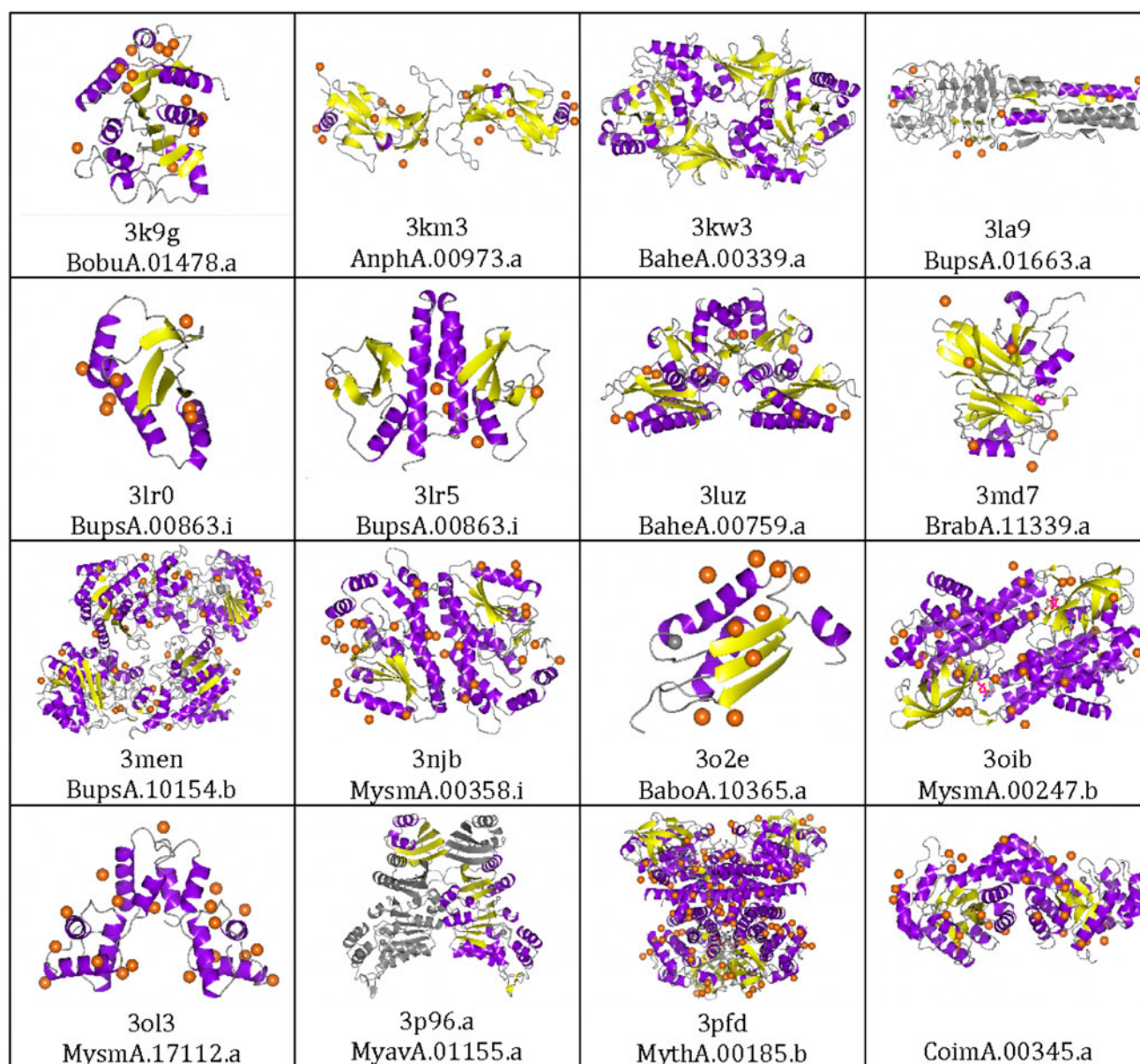


Fig. 2 Crystal structures determined by SSGCID using iodide ion soaks and SAD phasing

Examination of the crystallization and soaking conditions for the sixteen structures determined using iodide ion SAD phasing revealed a number of different precipitants from low to high molecular weight polyethylene glycols (PEG 400 to 8,000) and a variety of different salts including ammonium sulfate and sodium chloride (Table 1 and Supporting Information). These examples represent a broad distribution of precipitants common to a rational sparse matrix approach [19]. In addition, crystals grew and were soaked over a wide pH range from 4.0 to 8.6 (Table 1), spanning nearly the entire range of commonly used crystallization screens. The space groups varied from low symmetry (monoclinic, $C2$ and $P2_1$) to high symmetry

(cubic, $I23$) and from the most commonly observed space group $P2_12_12_1$ to rare space groups such as $I23$ (Table 2). The number of residues in the asymmetric unit varied from less than 100 to more than 1,500 (Table 2). The packing density [39] ranged from a solvent content of 25% ($V_m = 1.64 \text{ \AA}^3/\text{Da}$) to 59% ($V_m = 3.04 \text{ \AA}^3/\text{Da}$), spanning the range commonly observed for protein crystals. The phasing resolution ranged from 1.9 to 2.95 $\text{\AA}$ with most about 2.1 $\text{\AA}$; the native data sets for these structures ranged from 1.27 to 2.3 $\text{\AA}$ resolution (Table 1). Finally, the proteins themselves are quite varied in structure and function (Table 1; Fig. 2). For example, these structures contain both novel (3LA9 [7]) and previously observed folds. The

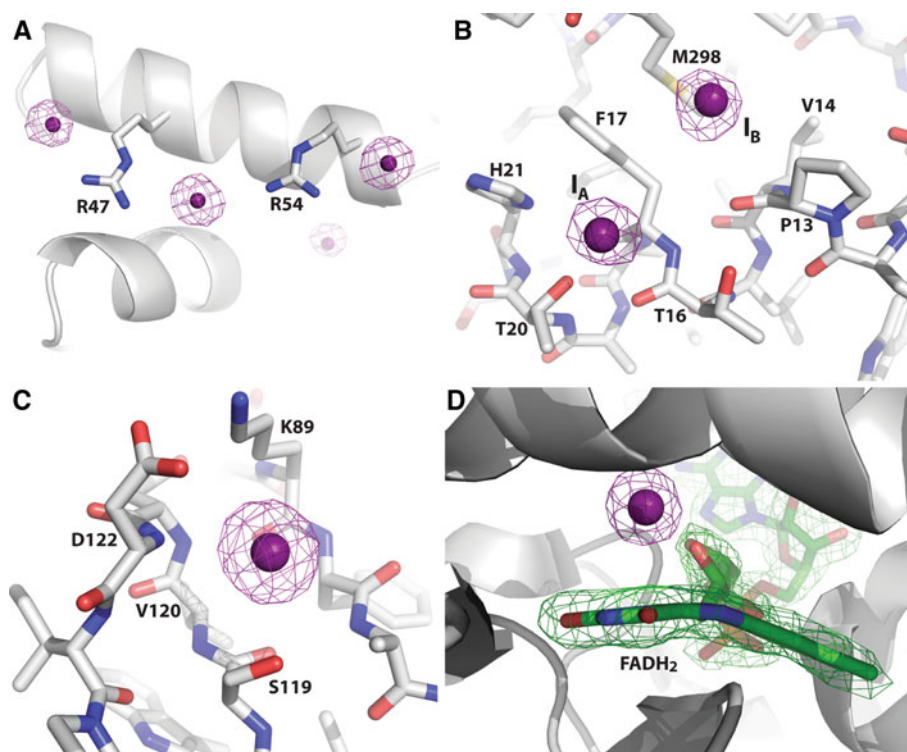


Fig. 3 Types of iodide ion binding sites. **a** Arginine-iodide ion interaction network on the surface of MymA.17112.a (PDB ID 3OL3), a putative uncharacterized protein from *Mycobacterium smegmatis* and an ortholog of community request protein MytuD.17112.a, Rv0543c from *Mycobacterium tuberculosis* (PDB ID 2KVC [60]) **b** Iodide ions binding along an α -helix of CoimA.00345.a, a putative fructose-1,6-bisphosphate aldolase from *Coccidioides immitis* (PDB ID 3PM6). Iodide ion I_A forms a possible anion-cation interaction with His21, while forming an interaction with the side chain hydroxyl of Thr20 (3.3 Å), an amide interaction with Thr16 and a hydrophobic interaction with Phe17. Iodide ion I_B forms hydrophobic interactions with the side chains of Met298 and Val14, while packing off the amide of Pro13-Val14. **c** Iodide ion

binding to the periplasmic domain of the risS pH sensor histidine kinase from *Burkholderia pseudomallei* (BupsA.00863.i, a community request target, PDB ID 3LR0). The iodide ion forms an interaction with the backbone amide nitrogen of Asp122 (3.5 Å) while forming another interaction with the side chain of Ser119 (3.4 Å) and packing against two β -sheets. **d** Iodide ion binding off reduced flavin adenine dinucleotide (FADH₂) in the crystal structure of an acyl-CoA dehydrogenase from *M. thermoresistibile* (MythA.00185.b, PDB ID 3PFD). An unbiased $|F_o| - |F_c|$ map calculated from a model lacking the cofactor is shown in green mesh contoured at 3.0 σ . For each panel iodide ions are shown as magenta spheres and an anomalous difference Fourier map is shown in magenta mesh contoured at 5.0 σ

tertiary structure ranged from all α -helical (3OL3) to nearly all β -sheet (3KM3). Different oligomeric states were observed from monomeric to tetrameric. Apo and ligand bound states were also observed. In all, these proteins, crystallization conditions, crystal forms, and diffraction properties are reflective of the SSGCID structure determination pipeline in general.

Types of iodide ion sites

The sixteen structures described here contained over 350 iodide ion sites combined, representing a wide variety of iodide ion-protein interactions (Fig. 3). These sites can be divided into several categories. As expected, we observed binding of iodide ions near positively charged surface residues, mostly arginine, lysine and to a lesser extent histidine residues. Interestingly, we observed a number of cases where the iodide ion displaced a surface exposed

negatively charged residue relative to the native structure, in order to interact with a positively charged residue. A second type of interaction we observed was the binding of iodide ions to hydrophobic patches, such as binding proximal to proline, methionine, or aromatic side-chains. Again, in some cases, we observed displacement of a side chain to an alternative rotamer conformation, relative to the native structure, to accommodate the iodide ion in proximity to a hydrophobic region. A third type of interaction, which was one of the most frequently observed, involved interactions with backbone amides via packing off the O-C-N plane or binding to solvent exposed amide nitrogen atoms. In the latter case, the iodide ion resided approximately 3.5–3.7 Å away from the amide nitrogen atom. A fourth type of interaction was the binding of iodide ions in the vicinity of H-bond forming residues such as glutamine and asparagine residues, and to a lesser extent threonine and serine residues. Surprisingly, this category included

aspartic acid and glutamic acid residues for crystals obtained at low pH (4.0–4.2), in which the carboxylic acid side chains are expected to be protonated. Of course, many iodide ions made multiple types of interactions with the protein, such as residing next to the side chain of an arginine residue while packing off an α -helix (Fig. 3). Interestingly, we observed iodide ion binding off the uridine-like ring of reduced riboflavin adenine dinucleotide (FADH₂) in the *M. thermoresistibile* acyl-CoA dehydrogenase crystal structure solved at 2.1 Å resolution (PDB ID 3PFD, Fig. 3d), an interaction which was observed in all four protomers of the biological tetramer observed in the asymmetric unit.

Case study 1: BolA-like protein from *Babesia bovis*

Babesia bovis is a tick-borne parasitic protozoan of the phylum apicomplexa. It primarily infects cattle causing babesiosis, a ‘malaria-like’ hemolytic anemia, although occasionally infecting humans as well [40]. SSGCID target BaboA.10365.a is a BolA-like protein from *B. bovis*. Although the function is not fully understood, the expression of a BolA-like protein in *Escherichia coli* was reported to be linked to a change in morphology, suggesting a function during cell division [41]. BaboA.10365.a is a small, 86-residue protein. Sequence database searches revealed BolA-like proteins from *Plasmodium falciparum* (51% sequence identity, PDB ID 2KDN; Buchko, G.W. et al. unpublished), from *Mus musculus* (43% sequence identity, 1V9J [42]) and from *E. coli* (34% sequence identity, 2HDM [43]) as closest sequence homolog with a known structure. Unfortunately, all of these solution NMR structures proved unsuitable for MR [44]. BaboA.10365.a could be crystallized in a tetragonal space group ($P4_12_12$) with a small unit cell ($a = b = 66$ Å, $c = 35$ Å) using ammonium sulphate as the precipitant. The crystals lost order when the environment was changed too quickly, for example by introducing iodide for phasing or ethylene glycol as cryoprotectant. When the concentration of iodide and cryoprotectant were simultaneously increased in several steps, well diffracting crystals could be prepared. A highly redundant data set with resolution limits of 1.95 Å revealed strong anomalous signal that extended to full resolution (SigA_{no} 1.70, anomalous correlation coefficient 59%). Four anomalous sites were located in Phenix [25] and extended to nine anomalous sites in PHASER EP [27]. After density modification with PARROT [29], ARP/wARP [31] could build 77 residues. While considering both enantiomorphic space groups, the first model of BaboA.10365.a was obtained literally minutes after data collection and data reduction had finished. A post-hoc analysis of homologous structures (Fig. 4) using the SSM/PDBeFOLD server [45] revealed

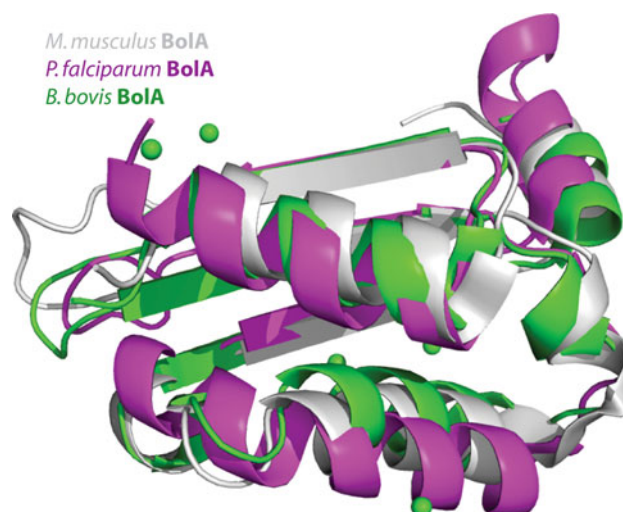


Fig. 4 Overlay of NMR solution structures of BolA-like proteins from *M. musculus* (gray, PDB ID 1V9J [42]), *P. falciparum* (magenta, Buchko, G.W. et al. unpublished) and the crystal structure of a BolA-like protein from *B. bovis* solved by iodide ion SAD (green, PDB ID 3O2E). Iodide ions are shown as green spheres. For simplicity only the ordered regions are shown

that sequence homologs have enough structural diversity, rendering them unsuitable for MR. The closest structural homolog is the BolA-like protein from *M. musculus* (1V9J, [42]) with a RMSD of 1.9 Å over 79 aligned residues. The closest sequence homolog is the BolA-like protein from *P. falciparum* (2KDN, Buchko, G.W. et al. unpublished), which is less structurally similar: RMSD of 2.6 Å over 76 aligned residues. The crystal structure of the BolA-like protein from *B. bovis* described here (PDB ID 3O2E) is the first crystal structure of a BolA-like protein, whereas all other BolA-like protein structures have been solved by solution NMR (*Plasmodium falciparum* PDB ID 2KDN, Buchko, G.W. et al. unpublished; *Mus musculus* 1V9J [42]; *E. coli* 2HDM [43]).

Case study 2: phosphoserine phosphatase SerB from *Mycobacterium avium*

Orthologs from several *Mycobacterium* species (*M. abscessus*, *M. avium*, *M. bovis*, *M. leprae*, *M. marinum*, *M. paratuberculosis*, *M. smegmatis*, *M. ulcerans*, and *M. thermoresistibile*) are used in a salvage pathway to rescue *M. tuberculosis* targets that fail at some stage of the SSGCID structure determination pipeline. Phosphoserine phosphatase SerB catalyzes the reaction of 3-phosphoserine to L-serine (EC:3.1.3.3), the final step in the biosynthesis of serine. SerB from *M. tuberculosis* (targetDB MytD.01155.a) failed to produce diffraction quality crystals, and thus the *M. avium* ortholog Ma SerB was entered into the SSGCID pipeline. Ma Ser yielded a 2.05 Å resolution native data set.

The closest homologous structure in the PDB contained 44% sequence identity over less than half the *Ma* SerB sequence with no homologous structure for the remainder of the protein. Unsurprisingly, MR failed. Crystals used for iodide phasing grew in 20% PEG 3350 and 0.2 M magnesium formate. For the first attempt, a crystal was soaked into a mixture of 75% precipitant and 25% KI-saturated ethylene glycol (final [KI] $\approx$ 0.5 M), but did not yield sufficient anomalous signal for structure solution (SigAno 1.34 for all reflections to 3.15 Å resolution). A second crystal was soaked into 20% PEG 3350, 0.1 M magnesium formate, 1.0 M KI, and 25% ethylene glycol for 4 min, followed by flash-cooling in liquid nitrogen and data collection, resulting in enhanced anomalous signal (SigAno 1.70). The structure was solved at 2.75 Å resolution using 15 iodide ion sites located using *phenix.autosol* [25], followed by automated building and refinement using the high resolution data set.

Like other SerB homologs, *Ma* SerB is a homodimer with Mg^{2+} and Cl^{-} bound in the active site. By homology, Mg^{2+} is a required cofactor, and the chloride ion occupies approximately the predicted phosphate position. Unusual for SerB enzymes, *Ma* SerB consists of three domains (Fig. 5), which is reflected in the low sequence homology of *Ma* SerB with other SerB enzymes. The first domain (residues 1–85) and second domain (residues 97–175) adopt the $\beta\alpha\beta\beta\alpha\beta$ ferredoxin fold, and there is a domain swap between the two monomers of the dimer via the long linker region between these two domains (Fig. 2). The sequence identity between domains 1 and 2 is only 25%, but they superimpose with RMSD 1.4 Å. Domain 3 (residues 182–400) represents the conserved core of the enzyme, and consists of a 6-stranded parallel beta sheet, resembling a Rossman fold with a set of extra helices and anti-parallel beta strands inserted. Other phosphoserine phosphatases overlap well on domain 3, and some, such as that from *Vibrio cholerae* (PDB ID 3N28, Patskovsky, Y. et al. unpublished) match domain 2 poorly (Fig. 5). The N-terminal domain 1 of *Ma* SerB appears to be unique.

Case study 3: putative fructose-1,6-bisphosphate aldolase from *Coccidioides immitis*

Coccidioides immitis is a pathogenic fungus which causes the potentially fatal systemic disease coccidioidomycosis [46], also known as Valley Fever because it was thought to originate in the San Joaquin Valley in California. Many open reading frames from the genome of *Coccidioides immitis* [47] are annotated as putative uncharacterized protein, but contain sequence homology to proteins of known function. One such case is the SSGCID target CoimA.00345.a, which is annotated as a putative

```

MNSPPKVSVLITVTGVDQPGVTATLFEVLSRHGVELLNVEQVVIRHRLTL
GVLVCCPADVADGPALRHDVEAAIRKVLDSIERSDDVPIIREPSTHTI
MSLDALTTLPIKHTALLNRFPETRF
100
FVLGRPITAAFGAVAREVAALGVNIDLRGVSDYFVIGLELRVSVPPGA
VTQLAKKRASWIVFGHYLTPAQFEDMDFETNRFNAILDMWKVGRYEVALM
DEALRTALNRVSSEEHVDVAVEDYTLERRAKRLIVFDVDSLTVGGEVTEM
DGELTSEHETILKALELDYARIQDVPDLTKPLIVLDMDSIAIQIECADE
200
LAAKAGAEQVAAITDAAMRGELDEACSLQORVATLAGLBATVIDEVAGQ
IAKLAGVGEEVAEVTERRAMQGELDEQSLRLRVSKLKDAFEQIILSQVRET
LEIMP GARTTLRLTLRLG YACGVV SGGFRRIIEPIAEEIMLDYVAANELE
LPLMPELPVLVATLHAFGWKVAIASGGFTYFSDYLKEQLSLDYAQSNTLE
300
IVDGTTLTGRVVGPIIDRAGKATALREFAQRAGVPMACQTVAVGDDGANDIDM
IVSGKLTGQVILGEVVSQAQTKADILTLTAQQYDVEIHN TVAVGDDGANDLVM
LAAAGLGIAFNAKPALREVA DASLSHPYLDTVLFLLGVTGRGEIEAADAID
MAAAGLVAYHAKPKVEAKQTAVRFAGLGGVVCILSALVAQQKLSWKS
400
GEVRRVEIPPE
KEGHHHHHH

```

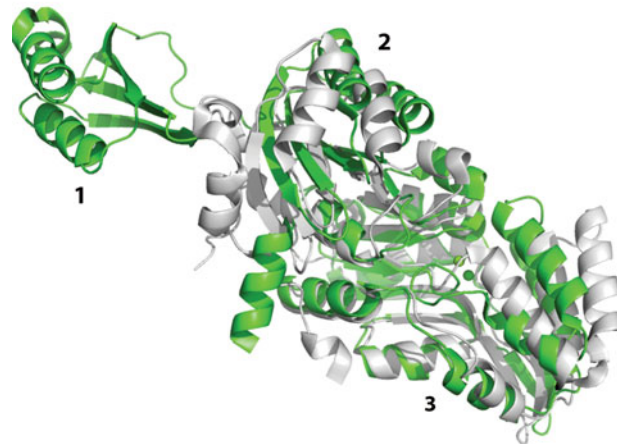


Fig. 5 Sequence alignment and crystal structures of phosphoserine phosphatase SerB from *Vibrio cholerae* (bottom sequence, gray ribbons, PDB ID 3N28, Patskovsky, Y. et al. unpublished) and *Mycobacterium avium* solved by iodide ion SAD (top sequence, green ribbons, PDB ID 3P96). For simplicity, only one monomer of the biological dimer is shown in each case. Domains 1, 2, and 3 of *Ma* SerB correspond to residues 1–85, 97–175, and 182–400, respectively

uncharacterized protein. CoimA.00345.a shares 35% sequence identity with 11% gaps over 70% of its sequence with fructose-1,6-bisphosphate aldolase from *Giardia lamblia* [48]. We obtained a 2.05 Å resolution native data set of CoimA.00345.a. Despite numerous attempts, we were unable to solve this target via MR, although we noted weak but plausible rotation (RFZ 6.0) and translation scores (TFZ 3.8) in Phaser MR [27]. Soaking with 0.5 M NaI for 2 min yielded a 2.4 Å resolution data set with weak anomalous signal (SigAno 0.91, 14% anomalous correlation) from which we were unable to solve the structure using a SAD experiment.

Another data set was obtained after soaking with 1 M NaI for 15 min. Attempts to solve the structure at 2.2 Å resolution failed, despite modest anomalous signal (SigAno 1.39, 52% anomalous correlation overall; 0.86, 7% in the last shell). The resolution limits were trimmed to 2.5 Å resolution (SigAno 1.64, 62% anomalous correlation overall; 0.96, 17% in the last shell), and despite locating five sites, again we were unable to solve the structure. Combining the weak MR solution and the anomalous sites [49] in Phaser EP [27], we obtained a total of 31 heavy atom sites (3 of which were incorrect), a FOM of 0.46, and clearly interpretable experimental electron density maps. Although the two fructose-1,6-bisphosphate aldolases from the eukaryote *G. lamblia* and the fungi *C. immitis* clearly share the same fold (Fig. 6), there are numerous gaps and deletions in the protein sequence and structure as well as considerable conformational heterogeneity, which results in an RMSD of 1.79 Å for aligned residues, a value which exceeds the common threshold for success by MR [44].

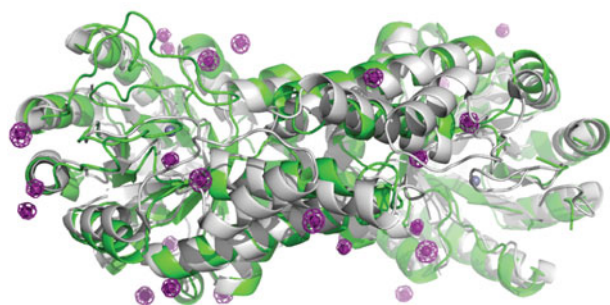
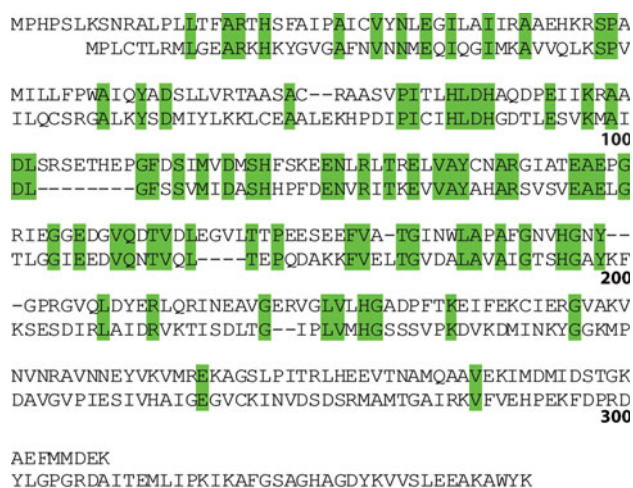


Fig. 6 Sequence alignment and crystal structures of fructose-1,6-bisphosphate aldolases from *Giardia lamblia* (bottom sequence, gray ribbons, PDB ID 2ISV [48]) and *Coccidioides immitis* solved by combined iodide ion SAD and MR (top sequence, green ribbons, PDB ID 3PM6). Iodide ions are shown as magenta spheres and the catalytic zinc ions are shown as gray spheres. Anomalous difference Fourier maps are shown in magenta mesh contoured at 5.0 σ

Discussion

Experimental phasing in a structural genomics environment

The most common method for de novo structure determination utilizes replacement of methionine residues with selenomethionine and multiwavelength anomalous dispersion (MAD) or SAD experiments [50]. For success, this method requires the completion of several steps. First, the protein of interest must contain internal methionine residues or have them engineered into the protein sequence. Second, a protein sample must be expressed using minimal media supplemented with selenomethionine, a process that often results in lower protein yields than native protein expression. Third, the protein sample must be purified and crystallized, a process that often requires re-optimization relative to the native protein due to the increased lipophilicity of selenomethionine labeled proteins. Finally, this method requires synchrotron radiation, preferably at a tunable beamline, although several monochromatic beamlines such as ALS 5.0.1 and 5.0.3 have recently been adjusted to an energy setting reflective of the selenium K edge [51]. As a consequence of the numerous additional steps required to solved structures by selenomethionine SAD/MAD phasing, there is typically a lag time of several months or more between obtaining native crystals that diffract to high resolution and solving the structure.

One of the major goals of the Protein Structure Initiative (PSI) was the elucidation of novel protein folds [52]. Thus, for many PSI centers all of the protein samples destined for crystallography were expressed as selenomethionine-labeled proteins [53] with the intention of solving the structures by traditional selenomethionine-based SAD or MAD experiments. Although thousands of structures have been solved using this method [50], there is additional cost in reagents and time associated with solving structures via this method in comparison with structures that may be solved by MR. Given the PSI major goal of obtaining novel protein folds, the additional cost is certainly justifiable. Rather than investigate new folds, the goal of the SSGCID is to determine structures of potential drug targets from infectious disease organisms. Given that most protein structures are expected to be solved by MR, all proteins are expressed in native form in an effort to lower costs and improve success rates; native proteins generally express at higher levels and consequently crystallize at higher rates than selenomethionine-labeled proteins. However, in practice not all structures are solved by MR, and thus the phases for these targets must be determined experimentally. We could have adopted the selenomethionine-based method; however, the SSGCID pipeline generates native crystals, and thus, we decided to proactively explore

phasing options that use native crystals rather than waiting months to obtain selenomethionine-labeled crystals.

There are a number of methods for heavy atom labeling of native crystals and structure determination by isomorphous replacement and/or anomalous scattering methods [54]. Methods developed within the past 10 years include radiation induced phasing [55], covalent iodination [56], non-covalent binding of an iodinated “magic triangle” [57], and bromine or iodide ion soaking and SAD or single isomorphous replacement with anomalous scattering (SIRAS) [4, 5]. For SSGCID, we wanted to use a method that did not require toxic compounds such as mercury or platinum, and which would be applicable over a wide range of SSGCID targets. Therefore, we selected the use of iodide ion soaking and SAD experiments [4, 5], which was predicted to be particularly suitable for structural genomics projects [8]. Over the past year, we applied this method to seventeen structural genomics targets and determined sixteen new structures. This method failed in only one case (URE3-BP from *Entamoeba histolytica*, targetDB ID EnhiA.01648.a, a community request target [58]), for which MR, selenomethionine-based SAD/MAD, bromide ion SAD/MAD, La^{3+} SIRAS, K_2PtCl_4 SIRAS, HgCl_2 SIRAS, tungstate SAD, Cs^+ SAD, and sulfur SAD have also failed thus far.

Why SAD phasing using iodide ions works

There are many reasons for the high success rate of structure determination using iodide ion soaks and SAD phasing by SSGCID. From the technical vantage point, the modern in house X-ray generators, optics and detectors used by SSGCID (see section “Materials and methods”) has had a significant impact on the success of this method. Using previous generation in house X-ray equipment, for many targets it would have taken 10 days to collect a full 360° data set required for high multiplicity, whereas the current generation in house equipment (approaching 10^{11} Xrays/s mm^2) rivals second generation synchrotrons in flux and data collection times, ranging from 10 min to a few hours. However, some synchrotron beamlines such as ALS 5.0.2 [51] have nearly as much flux at low energy where the anomalous signal for iodide is high (e.g., 1.54 Å) as at higher energy (e.g., 1 Å). Thus, this method is not exclusive to in house data collection, although one must consider the effects of radiation decay on the data collection parameters [37]. In addition to improved hardware, current software such as Phenix [25] and Phaser EP [27] has had a significant impact on the success by making the identification of anomalous sites rapid and accurate.

From the chemical vantage point, there are two major reasons for the success of this method by SSGCID. First,

SSGCID targets are purified in moderately high salt concentrations (0.3 M NaCl), implying that SSGCID proteins that crystallize were selected for stability in moderately high salt concentrations. Thus, soaking into 0.2–1 M iodide ions is unlikely to dramatically damage many of these crystals. Second, at high concentrations, soft ions such as iodides bind weakly to numerous sites on the surface of proteins. These sites include binding to positively charged residues, hydrophobic sites, amides, protonated residues, etc. These types of interactions form regardless of crystallization conditions including precipitant identity and pH, or even dramatically different tertiary or quaternary protein structure. These two features coupled with the high anomalous signal of iodide at a wavelength of 1.5418 Å ($f'' = 6.9 \text{ e}^-$ for iodide), which is higher than the theoretical value of selenium at synchrotron radiation ($f'' = 3.8 \text{ e}^-$ at 0.97946 Å; white line effects at the selenium peak may push f'' higher), are keys to the success of this method.

Conclusions

SAD phasing with iodide ions was applied to a diverse set of structural genomics protein targets from bacterial, fungal, and eukaryotic organisms, and over a wide range of crystallization conditions common to most sparse matrix approaches. The net result is rapid, effective, low cost structure solution that addresses a bottleneck otherwise created by the secondary preparation of selenomethionine-labeled crystals for a structural genomics pipeline that normally generated native protein samples and crystals. Although this method is not new, this is the first time it has been used in a high-throughput structural genomics environment. As a result, we obtained sixteen new structures in 1 year, providing a wealth of information with regard to the structural underpinnings of iodide ion SAD phasing. Moreover, these cases demonstrate the general applicability of this method for de novo structure determination.

Acknowledgments This research was funded under Federal Contract No. HHSN272200700057C from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services. We wish to thank the entire SSGCID team, especially members of the Crystal Core group, Darren Begley, Doug Davies, and Robin Stacy for critical assessment of the manuscript, and Banumathi Sankaran (Berkeley Center for Structural Biology) for collection of a number of high resolution native data sets.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Myler PJ, Stacy R, Stewart L, Staker BL, Van Voorhis WC et al (2009) The Seattle Structural Genomics Center for Infectious Disease (SSGCID). *Infect Disord Drug Targets* 9:493–506
- Van Voorhis WC, Hol WG, Myler PJ, Stewart LJ (2009) The role of medical structural genomics in discovering new drugs for infectious diseases. *PLoS Comput Biol* 5:e1000530
- Anderson WF (2009) Structural genomics and drug discovery for infectious diseases. *Infect Disord Drug Targets* 9:507–517
- Dauter M, Dauter Z (2007) Phase determination using halide ions. *Methods Mol Biol* 364:149–158
- Dauter Z, Dauter M, Rajashankar KR (2000) Novel approach to phasing proteins: derivatization by short cryo-soaking with halides. *Acta Crystallogr D Biol Crystallogr* 56:232–237
- Abendroth J, Mitchell DD, Korotkov KV, Johnson TL, Kreger A et al (2009) The three-dimensional structure of the cytoplasmic domains of EpsF from the type 2 secretion system of *Vibrio cholerae*. *J Struct Biol* 166:303–315
- Edwards TE, Phan I, Abendroth J, Dieterich SH, Masoudi A et al (2010) Structure of a *Burkholderia pseudomallei* trimeric auto-transporter adhesin head. *PLoS One* 5:e12803
- Yogavel M, Gill J, Mishra PC, Sharma A (2007) SAD phasing of a structure based on cocrystallized iodides using an in-house Cu K α X-ray source: effects of data redundancy and completeness on structure solution. *Acta Crystallogr D Biol Crystallogr* 63:931–934
- Yogavel M, Gill J, Sharma A (2009) Iodide-SAD, SIR and SIRAS phasing for structure solution of a nucleosome assembly protein. *Acta Crystallogr D Biol Crystallogr* 65:618–622
- Yogavel M, Khan S, Bhatt TK, Sharma A (2010) Structure of D-tyrosyl-tRNA^{Tyr} deacylase using home-source Cu K α and moderate-quality iodide-SAD data: structural polymorphism and HEPES-bound enzyme states. *Acta Crystallogr D Biol Crystallogr* 66:584–592
- Kostrewa D, Winkler FK, Folkers G, Scapozza L, Perozzo R (2005) The crystal structure of PfFabZ, the unique beta-hydroxyacyl-ACP dehydratase involved in fatty acid biosynthesis of *Plasmodium falciparum*. *Protein Sci* 14:1570–1580
- Dauter Z, Dauter M, Dodson E (2002) Jolly SAD. *Acta Crystallogr D Biol Crystallogr* 58:494–506
- Buchko GW, Robinson H, Abendroth J, Staker BL, Myler PJ (2010) Structural characterization of *Burkholderia pseudomallei* adenylate kinase (Adk): profound asymmetry in the crystal structure of the ‘open’ state. *Biochem Biophys Res Commun* 394:1012–1017
- Yamada S, Hatta M, Staker BL, Watanabe S, Imai M et al (2010) Biological and structural characterization of a host-adapting amino acid in influenza virus. *PLoS Pathog* 6:e1001034
- Aslanidis C, de Jong PJ (1990) Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic Acids Res* 18:6069–6074
- Lorimer D, Raymond A, Walchli J, Mixon M, Barrow A et al (2009) Gene composer: database software for protein construct design, codon engineering, and gene synthesis. *BMC Biotechnol* 9:36
- Raymond A, Lovell S, Lorimer D, Walchli J, Mixon M et al (2009) Combined protein construct and synthetic gene engineering for heterologous protein expression and crystallization using gene composer. *BMC Biotechnol* 9:37
- Studier FW (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* 41:207–234
- Newman J, Egan D, Walter TS, Meged R, Berry I et al (2005) Towards rationalization of crystallization screening for small- to medium-sized academic laboratories: the PACT/JCSG+ strategy. *Acta Crystallogr D Biol Crystallogr* 61:1426–1431
- Gerdts CJ, Elliott M, Lovell S, Mixon MB, Napuli AJ et al (2008) The plug-based nanovolume microcapillary protein crystallization system (MPCS). *Acta Crystallogr D Biol Crystallogr* 64:1116–1122
- Gerdts CJ, Stahl GL, Napuli A, Staker B, Abendroth J et al (2010) Nanovolume optimization of protein crystal growth using the microcapillary protein crystallization system. *J Appl Cryst* 43:1078
- Wernimont A, Edwards A (2009) In situ proteolysis to generate crystals for structure determination: an update. *PLoS One* 4:e5094
- Thakur AS, Robin G, Guncar G, Saunders NF, Newman J et al (2007) Improved success of sparse matrix protein crystallization screening with heterogeneous nucleating agents. *PLoS One* 2:e1091
- Kabsch W (2010) Xds. *Acta Crystallogr D Biol Crystallogr* 66:125–132
- Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW et al (2010) PHENIX: a comprehensive python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66:213–221
- Sheldrick GM (2008) A short history of SHELX. *Acta Crystallogr A* 64:112–122
- McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC et al (2007) Phaser crystallographic software. *J Appl Crystallogr* 40:658–674
- Collaborative Computational Project, Number 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* 50:760–763
- Cowtan K (2010) Recent developments in classical density modification. *Acta Crystallogr D Biol Crystallogr* 66:470–478
- Cowtan K (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr D Biol Crystallogr* 62:1002–1011
- Langer G, Cohen SX, Lamzin VS, Perrakis A (2008) Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat Protoc* 3:1171–1179
- Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53:240–255
- Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60:2126–2132
- Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ et al (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* 35:W375–W383
- Chen VB, Arendall WB III, Headd JJ, Keedy DA, Immormino RM et al (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* 66:12–21
- Cianci M, Helliwell JR, Suzuki A (2008) The interdependence of wavelength, redundancy and dose in sulfur SAD experiments. *Acta Crystallogr D Biol Crystallogr* 64:1196–1209
- Dauter Z (2010) Carrying out an optimal experiment. *Acta Crystallogr D Biol Crystallogr* 66:389–392
- Schneider TR, Sheldrick GM (2002) Substructure solution with SHELXD. *Acta Crystallogr D Biol Crystallogr* 58:1772–1779
- Matthews BW (1968) Solvent content of protein crystals. *J Mol Biol* 33:491–497
- Hunfeldt KP, Hildebrandt A, Gray JS (2008) Babesiosis: recent insights into an ancient disease. *Int J Parasitol* 38:1219–1237
- Santos JM, Freire P, Vicente M, Arraiano CM (1999) The stationary-phase morphogene *bolA* from *Escherichia coli* is induced by stress during early stages of growth. *Mol Microbiol* 32:789–798

42. Kasai T, Inoue M, Koshiba S, Yabuki T, Aoki M et al (2004) Solution structure of a BolA-like protein from *Mus musculus*. *Protein Sci* 13:545–548
43. Tuinstra RL, Peterson FC, Elgin ES, Pelzek AJ, Volkman BF (2007) An engineered second disulfide bond restricts lymphotactin/XCL1 to a chemokine-like conformation with XCR1 agonist activity. *Biochemistry* 46:2564–2573
44. Chen YW, Dodson EJ, Kleywegt GJ (2000) Does NMR mean “not for molecular replacement”? Using NMR-based search models to solve protein crystal structures. *Structure* 8:R213–R220
45. Krissinel E, Henrick K (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr* 60:2256–2268
46. Hector RF, Laniado-Laborin R (2005) Coccidioidomycosis—a fungal disease of the Americas. *PLoS Med* 2:e2
47. Sharpton TJ, Stajich JE, Rounsley SD, Gardner MJ, Wortman JR et al (2009) Comparative genomic analyses of the human fungal pathogens *Coccidioides* and their relatives. *Genome Res* 19:1722–1731
48. Galkin A, Kulakova L, Melamud E, Li L, Wu C et al (2007) Characterization, kinetics, and crystal structures of fructose-1, 6-bisphosphate aldolase from the human parasite, *Giardia lamblia*. *J Biol Chem* 282:4859–4867
49. Roversi P, Johnson S, Lea SM (2010) With phases: how two wrongs can sometimes make a right. *Acta Crystallogr D Biol Crystallogr* 66:420–425
50. Walden H (2010) Selenium incorporation using recombinant techniques. *Acta Crystallogr D Biol Crystallogr* 66:352–357
51. Morton S, Glossinger J, Smith-Baumann A, McKean JP, Trame C et al (2007) Recent major improvements to the ALS sector 5 macromolecular crystallography beamlines. *Sync Rad News* 20:23–30
52. Terwilliger TC, Berendzen J (1999) Exploring structure space. A protein structure initiative. *Genetica* 106:141–147
53. Stols L, Millard CS, Dementieva I, Donnelly MI (2004) Production of selenomethionine-labeled proteins in two-liter plastic bottles for structure determination. *J Struct Funct Genomics* 5:95–102
54. Joyce MG, Radaev S, Sun PD (2010) A rational approach to heavy-atom derivative screening. *Acta Crystallogr D Biol Crystallogr* 66:358–365
55. Ravelli RB, Leiros HK, Pan B, Caffrey M, McSweeney S (2003) Specific radiation damage can be used to solve macromolecular crystal structures. *Structure* 11:217–224
56. Miyatake H, Hasegawa T, Yamano A (2006) New methods to prepare iodinated derivatives by vaporizing iodine labelling (VIL) and hydrogen peroxide VIL (HYPER-VIL). *Acta Crystallogr D Biol Crystallogr* 62:280–289
57. Beck T, Krasauskas A, Gruene T, Sheldrick GM (2008) A magic triangle for experimental phasing of macromolecules. *Acta Crystallogr D Biol Crystallogr* 64:1179–1182
58. Gilchrist CA, Baba DJ, Zhang Y, Crasta O, Evans C et al (2008) Targets of the *Entamoeba histolytica* transcription factor URE3-BP. *PLoS Negl Trop Dis* 2:e282
59. Brennan S, Cowan PL (1992) A suite of programs for calculating x-ray absorption, reflection and diffraction performance for a variety of materials at arbitrary wavelengths. *Rev Sci Instrum* 63:850
60. Buchko GW, Phan I, Myler PJ, Terwilliger TC, Kim YC (2011) Inaugural structure from the DUF3349 superfamily of proteins, *Mycobacterium tuberculosis* Rv0543c. *Arch Biochem Biophys* 506(2):150–156