

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/221788330>

# Enhancement of crystallization with nucleotide Ligands identified by dye–ligand affinity chromatography

ARTICLE *in* JOURNAL OF STRUCTURAL AND FUNCTIONAL GENOMICS · JANUARY 2012

DOI: 10.1007/s10969-012-9124-8 · Source: PubMed

---

CITATIONS

4

---

READS

23

7 AUTHORS, INCLUDING:



Li-Wei Hung

Los Alamos National Laboratory

59 PUBLICATIONS 11,216 CITATIONS

SEE PROFILE



Thomas Charles Terwilliger

Los Alamos National Laboratory

231 PUBLICATIONS 24,638 CITATIONS

SEE PROFILE

# Enhancement of crystallization with nucleotide ligands identified by dye-ligand affinity chromatography

Heungbok Kim · Cecelia Webster · Justin K. M. Roberts ·  
Juthamas Kositsawat · Li-Wei Hung ·  
Thomas C. Terwilliger · Chang-Yub Kim

Received: 21 September 2011 / Accepted: 6 January 2012 / Published online: 28 January 2012  
© The Author(s) 2012. This article is published with open access at Springerlink.com

**Abstract** Ligands interacting with *Mycobacterium tuberculosis* recombinant proteins were identified through use of the ability of Cibacron Blue F3GA dye to interact with nucleoside/nucleotide binding proteins, and the effects of these ligands on crystallization were examined. Co-crystallization with ligands enhanced crystallization and enabled X-ray diffraction data to be collected to a resolution of at least 2.7 Å for 5 of 10 proteins tested. Additionally, clues about individual proteins' functions were obtained from their interactions with each of a panel of ligands.

**Keywords** Characterization of proteins based on ligands · Dye-ligand affinity chromatography · Enhancement of crystallization · Ligand aided crystallization · Ligand analysis · Nucleotide ligand

## Abbreviations

ADP	Adenosine-5'-diphosphate
AMP	Adenosine-5'-monophosphate
ATP	Adenosine-5'-triphosphate
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
FAD	Flavin-adenine dinucleotide
GTP	Guanosine-5'-triphosphate
IPTG	Isopropyl- $\beta$ -D-thiogalactoside
NAD	Nicotinamide adenine dinucleotide
NADH	Nicotinamide adenine dinucleotide hydride
NADP	Nicotinamide adenine dinucleotide phosphate
NADPH	Nicotinamide adenine dinucleotide phosphate hydride
OD	Optical density
PCR	Polymerase chain reaction
PMSF	Phenylmethylsulfonyl fluoride
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
UTP	Uridine-5'-triphosphate

**Electronic supplementary material** The online version of this article (doi:10.1007/s10969-012-9124-8) contains supplementary material, which is available to authorized users.

H. Kim · T. C. Terwilliger · C.-Y. Kim (✉)  
Bioscience Division, Los Alamos National Laboratory,  
MS M888, Los Alamos, NM 87545, USA  
e-mail: cykim@lanl.gov

C. Webster · J. K. M. Roberts  
Department of Biochemistry, University of California,  
Riverside, CA 92521, USA

J. Kositsawat  
School of Medicine, Loma Linda Medical School, Loma Linda,  
CA 92350, USA

L.-W. Hung  
Physics Division, Los Alamos National Laboratory, MS D454,  
Los Alamos, NM 87545, USA

## Introduction

Obtaining diffraction-quality crystals is a major bottleneck in determining macromolecular structures using X-ray crystallography [1, 2]. Various general approaches have been suggested for improving crystallization, including systematic screening of conditions [3]; seeding with small crystals [4]; cell-free expression [5]; screening multiple starting and ending amino acid positions for expression clones [6]; identifying domains by partial proteolysis [7, 8]; crystallizing complexes of small proteins [9, 10]; performing surface engineering [11]; using crystallization

chaperones [12]; and supplementing with small-molecule ligands to increase crystallization [13] (See recent review of [14].

Ligand supplementation is particularly attractive as an option for improving crystallization in cases where ligand information for the target protein can be predicted from sequence-based functional annotation [15, 16]. If the right ligand is chosen, the effect on crystallization can be dramatic [17, 18]. In practice, the success rate of this approach is limited by the accuracy of ligand prediction for target proteins. In many cases, available annotations give little or no useful information on the ligand for target proteins. Estimates of error rates in annotation of gene function range from 5% to over 80% depending on the annotation methods and the database [19, 20], and this can lead to an incorrect choice of ligands. The unknown and hypothetical proteins that account for 25–60% of genome sequences give no indications of ligand [21–23]. As crystallization can be enhanced by addition of correctly selected ligand(s), accurate ligand information would be of substantial utility in crystallization.

After the fortuitous discovery of the binding of Cibacron Blue F3GA to enzymes led to the concept of using dyes for affinity chromatography [24], several triazine dyes were applied to enzyme purification as affinity resins. One of the most popular triazine dyes, Cibacron Blue F3GA, was found to interact with a variety of enzymes such as NAD- and NADP-dependent dehydrogenases, kinases, glucose oxidase, lysozyme, albumin, catalase, ovalbumin, as well as plasma proteins [25–31].

We have made use of the interaction of Cibacron Blue F3GA dye with nucleoside and nucleotide binding proteins to identify protein–ligand interactions [32]. Taking advantage of the binding of Cibacron Blue F3GA dye to a variety of cellular enzymes, we expanded the use of this dye from protein purification to high-throughput ligand identification of resin-bound proteins by selective elution with ligands [17, 32]. The premise of our approach is that a ligand that interacts specifically with a protein can release that protein from the dye-resin competitively. Crude cell extracts are applied to the dye-resin and the proteins eluted by each of a series of potential ligands are identified by two-dimensional electrophoresis and mass spectrometry [32]. Then the specificity of interactions is further tested by examining the elution of purified recombinant proteins from the dye-resin by individual ligands [17].

In this report, we demonstrate the general application of the F3GA dye–ligand affinity chromatography method to enhancing crystallization of 9 recombinant proteins chosen from a set of 26 *Mycobacterium tuberculosis* (*Mtb*) targets that we previously found to interact with nucleoside/nucleotide ligands [32].

## Materials and methods

### Target selection, expression and purification of proteins

From our previous study with *Mtb* strain H37Rv cell extract in which we identified proteins interacting with nucleotide–ligands [32], 26 genes were selected and used in cloning. All genes in this study were amplified by PCR from a *Mtb* H37Rv genomic DNA, using Pfu polymerase (Stratagene), and the following primers: 5'-TACTTCC AATCCAATGCGAT G + N-terminal 20 nucleic acids coding region of target protein-3' (forward) and 5'-TTA TCCACTTCC AATGTTA + C-terminal 20 nucleic acids coding region of target protein-3' (reverse). The underlined bases were to generate ligation-independent cloning (LIC) sites for the pMCSG7 vector [33]. The pMCSG7 vector was digested with the SspI restriction enzyme (Promega), and the amplified and purified PCR product and a singly digested pMCSG7 vector were treated with T4 DNA polymerase (Novagen). The 50 µl of *Escherichia coli* NovaBlue cells (Novagen) was transformed with the self-annealed PCR product and pMCSG7 vector. The insert of genes in plasmids was confirmed by DNA sequencing.

*E. coli* BL21(DE3) cells were used to express the cloned genes. Cells were grown at 37°C in LB medium (Sigma) containing 100 µg/ml ampicillin, induced with 1 mM IPTG when OD600 reached 1.0, and grown at 25°C overnight in a shaking incubator set at 250 rpm. The cells were harvested and stored at −80°C. The expression of each protein was checked by SDS-PAGE [34].

For purification of expressed proteins, frozen cells were thawed on ice and resuspended in lysis buffer (20 mM Tris–HCl pH 8.0, 200 mM NaCl, 1 mM PMSF, 1 mg/ml DNase, 1 mM MgCl<sub>2</sub>). Lysates were sonicated and then centrifuged with 3,000 g at 4°C for 30 min. The supernatant was filtered through a 0.45 µm pore membrane (Stericup, Millipore) and loaded on a Ni–NTA superflow affinity column (Qiagen). After being washed with buffer A (20 mM Tris–HCl pH 8.0, 200 mM NaCl), the target protein was eluted by buffer B (buffer A plus 500 mM imidazole). To remove the contaminants, eluted fractions were further purified on a Superdex-75 gel filtration column (GE Healthcare Inc.) using buffer C (10 mM Tris–HCl pH 8.0, 150 mM NaCl, and 1 mM DTT). A centrifugal concentrator (Millipore) was used to concentrate the pooled protein fractions to 5–15 mg/ml, as measured by Bradford reagent (Bio-Rad). Protein purity was confirmed by SDS-PAGE and densitometry. A summary of progress from solubility test to X-ray data collection is included in Supplementary Table 1.

**Table 1** Estimation of retention of recombinant *Mtb* proteins on Cibacron Blue F3GA resin shown in Fig. 1

Proteins	Resin-bound portion of loaded protein (%) <sup>a</sup>	Eluted portion of bound protein (%) <sup>b</sup>							
		NAD	NADH	NADP	NADPH	AMP	ADP	ATP	GTP
NAD(P) transhydrogenase	100	0	24	0	0	0	0	0	0
Short-chain-type dehydrogenase/reductase	100	0	0	74	99	0	0	0	0
Short-chain-type oxidoreductase	100	0	0	0	15	0	0	0	0
6-phosphogluconate dehydrogenase	66	0	31	0	2	0	1	0	0
Secreted L-alanine dehydrogenase	100	0	55	0	0	0	0	11	0
Methylmalonate-semialdehyde dehydrogenase	63	0	13	0	5	8	0	0	0
5-methyltetra-hydropteroyltriglutamate-homocysteine methyltransferase	86	5	11	3	29	17	9	17	42
3-hydroxy-isobutyrate dehydrogenase	100	40	74	0	15	26	26	41	26

<sup>a</sup> The resin-bound portion of protein was estimated as a percentage of the total protein loaded by subtracting the flow-through portion from the loaded amount. Amounts of protein loaded and obtained in the flow-through were estimated from densitometry of stained SDS gels as in Fig. 1, as detailed in Supplementary Table 2

<sup>b</sup> The eluted portion of bound protein was calculated by dividing the amount of protein eluted with each ligand by the amount of resin bound protein

### Nucleotide ligand analysis by dye-ligand affinity chromatography

To evaluate recombinant proteins for their specific ligand-binding properties, we followed a modified version of the protocol described by Kim et al. [17]. Briefly, individual purified proteins were diluted to 1–2 mg/ml in column buffer (CB; 50 mM potassium phosphate, pH 7.5, 1 mM MgCl<sub>2</sub> and 2 mM DTT), and 100 µg protein was adsorbed to multiple 50 µl aliquots of F3GA resin (Bio-RAD) in 2 ml spin-columns (Costar, Fisher Scientific). Aliquots (F3GA resin + proteins) were vortexed for 1 h at 4°C for binding, followed by recovery of unbound protein (flow-through fraction) and washing of the resin (4 × 0.4 ml washes with CB); spin-columns were centrifuged for 30 s at 10,000 g to recover fractions and change solutions. Individual spin-columns containing resin-bound proteins were then incubated (as for protein binding) with 50 µl, 1 mM ligand in CB, and the eluate fraction was recovered by centrifugation. The ligands used in this experiment were NAD, NADH, NADP, NADPH, AMP, ADP, ATP, and GTP (FAD was used additionally for the conserved hypothetical protein). Aliquots of initial protein, flow-through, and eluate fractions were diluted with 1:1 volume ratio with 2 × SDS sample buffer, and 15 µl was loaded in equal proportion (equivalent to 1 µg input protein) on 10% SDS-PAGE. For quantitative evaluation of the interaction of each ligand with protein, stained gel-bands were scanned by densitometry (GS-800, Bio-RAD). We calculated a “densitometric trace” of each gel-band by integration of the absorbance of the stained band using the software, Quantity One 4.3.1. Each protein’s resin-bound portion was calculated by subtracting the portion of flow-through

from the loaded protein. The protein portion eluted by each ligand was calculated as a percentage of the resin-bound protein. (Table 1 and Supplementary Table 2). We tested 33 of the 72 ligand–protein interactions examined in this work three times to estimate the uncertainty in our measurements of percentage bound to the dye column. The range of the percentage bound was less than ±15% from the mean in each case, and averaged ~10%.

### Crystallization and X-ray data collection of proteins without and with ligands

Proteins were crystallized without and with ligand(s) that had been identified with our dye-ligand interaction approach. For initial screening, hanging drops (1 µl protein–ligand solution + 1 µl reservoir solution) were set up in 24-well plates using crystal screen 1 and 2 (Hampton Research). To compare the effect of ligand(s) on crystallization under identical conditions, two or more drops that include one drop without ligand and other drop(s) with ligand(s) were set up on a cover-slip of each well depending on the number of identified ligands for each protein by dye-ligand affinity chromatography. If crystals were observed in multiple wells, we chose the well of best crystal based on size and morphology and optimized its crystallization condition, if possible, by fine-tuning each component until crystals with dimensions of at least 50 µm × 50 µm × 50 µm were obtained. For co-crystallization with ligand, each protein (0.2–0.5 mM in buffer C) was mixed with the identified ligand at a molar ratio of protein: ligand, 1:2.5 (and 1:5 to see the effect of enhancement of crystallization with more ligand, but no clear difference in enhancement of crystallization was

**Table 2** Statistics on X-ray data collection of four proteins, for which ligands information improved crystallization

Data collection	NAD(P) trans-hydrogenase	Short-chain-type dehydrogenase/reductase		6-phosphogluconate dehydrogenase	Secreted L-alanine dehydrogenase	
	NADH	NADP	NADPH	NADH	NADH	ATP
Space group	P2 <sub>1</sub>	P2 <sub>1</sub> 2 <sub>1</sub> 2	P2 <sub>1</sub> 2 <sub>1</sub> 2	P4 <sub>2</sub> 2	R32	R32
Unit cell parameters (esd's)						
a	65.144 (0.001)	160.450 (0.001)	159.395 (0.001)	108.280 (0.001)	88.154 (0.001)	89.047 (0.001)
b	151.841 (0.001)	159.952 (0.001)	160.801 (0.001)	108.280 (0.001)	88.154 (0.001)	89.047 (0.001)
c (Å)	76.306 (0.001)	86.189 (0.001)	172.583 (0.001)	144.938 (0.002)	291.390 (0.003)	290.981 (0.004)
$\alpha, \beta, \gamma$ (degree)	$\alpha = \gamma = 90,$ $\beta = 114.38$ (0.001)	$\alpha = \beta = \gamma = 90$	$\alpha = \beta = \gamma = 90$	$\alpha = \beta = \gamma = 90$	$\alpha = \beta = 90,$ $\gamma = 120$	$\alpha = \beta = 90,$ $\gamma = 120$
Resolution limits <sup>a</sup> (Å)	50.00–2.28 (2.32–2.28)	50.00–2.00 (2.03–2.00)	50.00–2.00 (2.03–2.00)	50.00–2.70 (2.75–2.70)	50.00–2.00 (2.03–2.00)	50.00–2.40 (2.43–2.40)
Wavelength (Å)	0.9774	1.0000	0.9202	1.0000	1.0000	1.0000
No. of Unique Reflections	59,257	148,901	298,159	24,403	29,900	17,977
No. of Total Reflections	209,793	1,088,722	2,214,370	167,780	177,183	106,091
R <sub>merge</sub> <sup>b</sup> (%)	6.4 (28.0)	10.6 (93.7)	9.5 (63.3)	15.9 (78.5)	5.7 (58.9)	6.5 (34.9)
Completeness (%)	96.6 (85.0)	98.7 (97.9)	100.0 (100)	100.0 (99.7)	100.0 (99.9)	99.6 (99.4)
$\langle I/\sigma(I) \rangle$	13.1 (3.6)	9.3 (2.0)	8.5 (2.8)	5.8 (2.0)	16.4 (2.6)	16.6 (4.4)
Redundancy	3.5 (3.0)	7.3 (6.4)	7.4 (7.4)	6.9 (6.2)	5.9 (5.6)	5.9 (5.8)

<sup>a</sup> Numbers in parentheses refer to the highest resolution shells

<sup>b</sup>  $R_{\text{merge}} = \sum |I_{\text{obs}} - I_{\text{avg}}| / \sum I_{\text{avg}}$

observed depending on ligand concentration), and incubated on ice for 30 min prior to set-up crystallization. For X-ray data collection, a minimum of five protein crystals grown were selected based on size and morphology, harvested and flash-cooled in liquid N<sub>2</sub>, with the addition of 10–20% glycerol in the buffer as cryoprotectant.

Monochromatic datasets were collected at the beam lines 5.0.1 and 5.0.2 at the Advanced Light Source (ALS). The wavelengths used for each data collection and the estimated standard deviations (esd's) of the wavelengths are listed in Table 2. All datasets were processed with the HKL2000 program suite [35]. Detailed data collection statistics are listed in Table 2.

## Results and discussion

### Protein preparation and nucleotide ligand analysis by dye-ligand affinity chromatography

To determine whether the ligand(s) identified by F3GA dye-ligand affinity chromatography method can enhance crystallization, we selected 26 proteins from *Mtb* cell extracts for which the approach identified potential ligand interactions [32]. All 26 target genes were successfully cloned, and expression of each was confirmed. Solubility

tests indicated that 10 of these were insoluble when expressed in *E. coli*, and these were not considered further. Crystal structures of three of the remaining 16 soluble proteins were already present in the Protein Data Bank (PDB, [36]; PDB codes listed in Supplementary Table 1) and these proteins were also not considered further. We purified the remaining 13 proteins.

The first step in our dye-ligand affinity chromatography experiment is normally evaluation of protein adsorption onto the Cibacron Blue F3GA dye resin. The buffer used in these experiments contains phosphate (50 mM), which we note might possibly compete weakly with dye binding and decrease weak or non-specific interaction of protein with dye resin. All proteins in this report had previously shown interactions with the F3GA resin, and binding to the resin was confirmed for all proteins (data not shown). For three of the 13 proteins, the ligand-specific elution analysis described below did not show interactions with the ligands we tested. These three (3-hydroxybutyryl-CoA dehydrogenase, probable fatty acid oxidation protein, and tryptophanyl-tRNA synthetase) were not considered further. Crystallization trials without and with ligand(s) identified in this analysis were set up for the remaining 10 proteins (Supplementary Table 1). Eight of these ten proteins (all but NAD(P) transhydrogenase and short-chain-type oxidoreductase) had previously been targeted by members of

the Tuberculosis structural genomics consortium (TB SGC) for crystallization and structure determination, but various attempts to solve structures had failed (<http://www.webtb.org>).

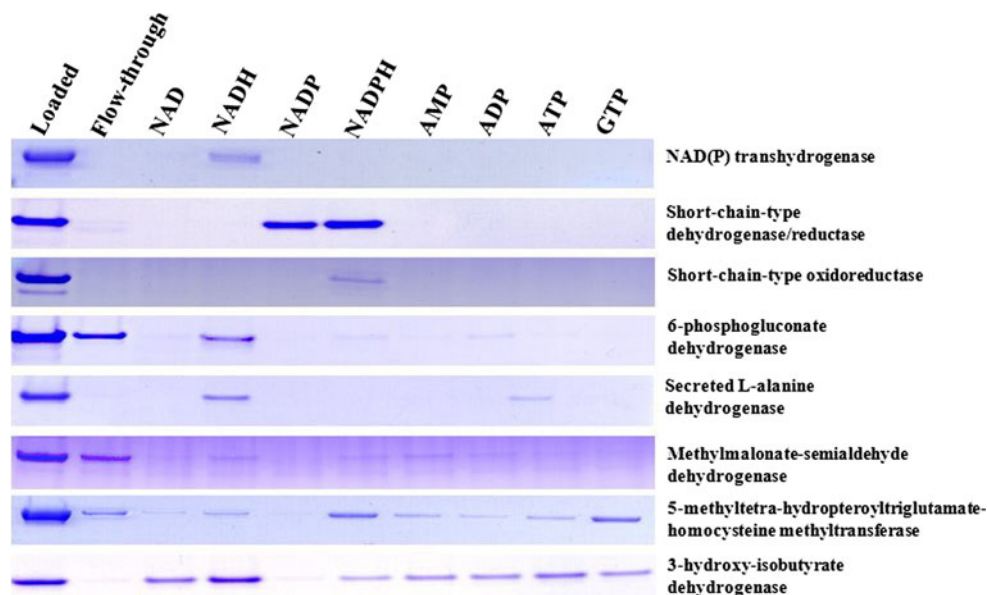
The enhancement of crystallization for one of these proteins (aldehyde dehydrogenase) with nucleoside ligands has been reported previously [17], and the experimental data for the remaining nine proteins are presented here. For all but one of these (the conserved hypothetical protein) addition of ligands improved crystallization (Supplementary Table 1). The eight proteins for which ligand addition improved crystallization are examined in more detail below. Figure 1 shows the application of the dye-ligand affinity chromatography approach to these eight *Mtb* proteins, testing interactions with eight common nucleotide ligands. The basis of this experiment is that the binding of ligand affects the binding of the protein to the resin. Initial binding of each protein to the resin was confirmed by the small relative amount in the flow-through fraction (lane 2) compared with the amount loaded (lane 1). For five of eight proteins, essentially all of the applied protein was bound to the resin. For the three others (6-phosphogluconate dehydrogenase, methylmalonate-semialdehyde dehydrogenase, and 5-methyltetra-hydropteroyl triglutamate-homocysteine methyltransferase), unbound protein accounted for 14–37% of the loaded protein amount, respectively (Table 1). Ligands interacting with each of the eight proteins were identified by selective elution of the proteins. A relative measure of the interaction of each ligand with each protein could be obtained from the fraction of resin-bound protein that is eluted by that ligand (see “Materials and

methods” and Supplementary Table 2), and is presented in Table 1.

Of eight proteins shown in Fig. 1, five revealed very obvious interactions with ligands. The proteins annotated as NAD(P) transhydrogenase, short-chain-type oxidoreductase, 6-phosphogluconate dehydrogenase, and secreted L-alanine dehydrogenase showed elution dominated by one ligand (NADH or NADPH). The short-chain-type dehydrogenase/reductase was eluted with two ligands (NADP and NADPH). In the case of methylmalonate-semialdehyde dehydrogenase, three ligands showed interactions. In contrast, the proteins annotated as 5-methyltetra-hydropteroyl triglutamate-homocysteine methyltransferase and 3-hydroxy-isobutyrate dehydrogenase interacted with most of the ligands tested. Additionally, the conserved hypothetical protein showed interaction with FAD (Supplementary Tables 1 and 2).

Based on existing annotations, the eight proteins shown in Fig. 1 are likely all to be dehydrogenases or reductases, and they would be expected to interact with either NAD/NADH or NADP/NADPH. The ligand interactions indicated by our dye-elution analysis give clues as to the likely cofactors for several of these proteins. As NAD(P) transhydrogenase, 6-phosphogluconate dehydrogenase, and secreted L-alanine dehydrogenase are eluted most readily with NADH, we suggest that they may be NAD/NADH-dependent dehydrogenases. In contrast, short-chain-type dehydrogenase/reductase and short-chain-type oxidoreductase are eluted specifically by NADP and/or NADPH, and therefore may be NADP/NADPH-dependent enzymes. The proteins NAD(P) transhydrogenase, 6-phosphogluconate

**Fig. 1** Nucleotide ligand interaction assay using *Mtb* recombinant proteins by dye-ligand affinity chromatography. The dye-ligand affinity chromatography results from SDS-PAGE show binding of eight target proteins to Cibacron Blue F3GA and subsequent selective elution in response to exposure to a sequence of eight different nucleotide ligands (NAD, NADH, NADP, NADPH, AMP, ADP, ATP, and GTP)



\* Annotation of target proteins is obtained from *Mycobacterium tuberculosis* Structural Genomics Consortium (TB SGC) web site (<http://www.webtb.org>).



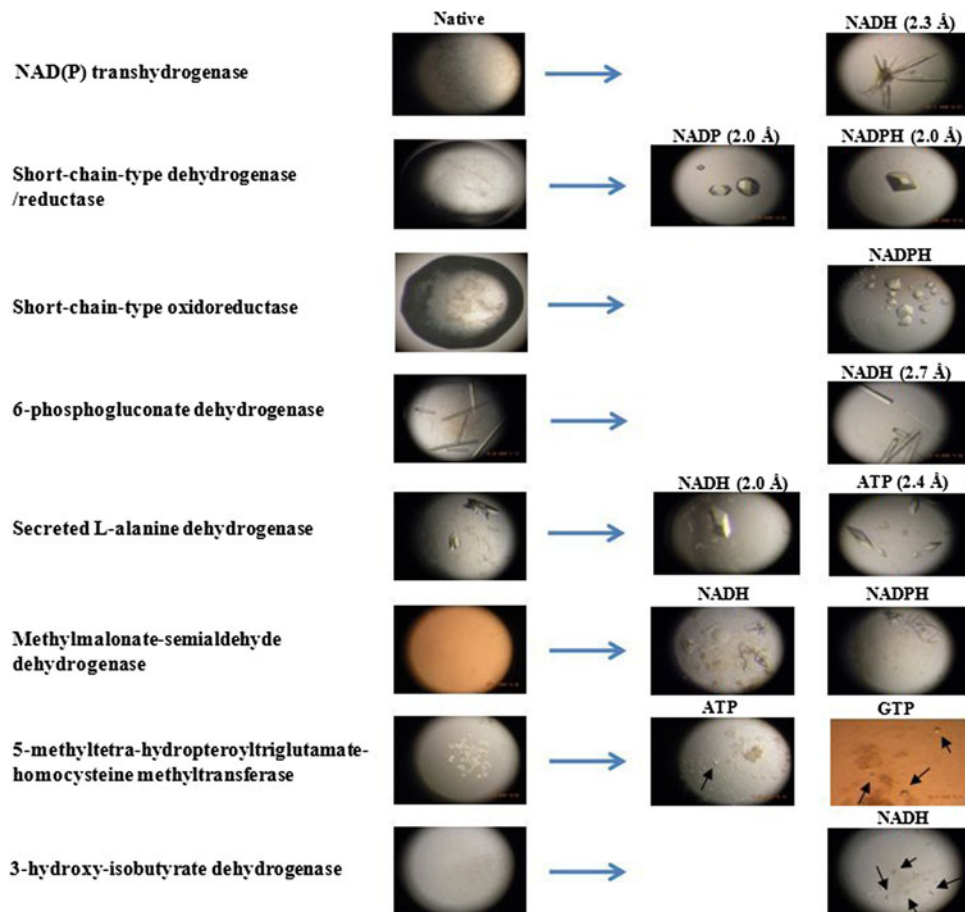
dehydrogenase, secreted L-alanine dehydrogenase, and short-chain-type oxidoreductase showed preferred interactions with the reduced form of cofactors (NADH or NADPH), suggesting that these enzymes may be involved in a reduction process.

During our work, structural and interaction data for secreted *Mtb* L-alanine dehydrogenase with NADH and other ligands was reported by Agren et al [37]. In agreement with our ligand interaction data, a strong interaction of this protein with NADH was confirmed ( $K_d = 8.2 \mu\text{M}$  and  $K_m = 10.7 \mu\text{M}$ ) showing that this enzyme is likely to be involved more in reductive amination converting pyruvate and ammonia to L-alanine rather than reverse oxidative deamination. These observations were also supported by kinetic parameters of this enzyme:  $K_{cat} = 694 \pm 33/\text{s}$ ,  $K_{m,\text{pyruvate}} = 0.76 \pm 0.05 \text{ mM}$  for reductive amination versus  $K_{cat} = 126 \pm 4/\text{s}$ ,  $K_{m,\text{L-alanine}} = 15.64 \pm 1.09 \text{ mM}$  for oxidative deamination. This agreement with our observations suggests that protein–ligand interaction profiles obtained by dye–ligand affinity chromatography can produce not only information about ligand binding to a protein, but also useful functional information.

Ligand(s) of a protein can be predicted from sequence-based annotation and from homologous protein's structure solved with ligand. The challenges of ligand prediction by sequence-based annotation were mentioned above. Since all eight proteins used in this experiment are annotated in the category of dehydrogenases or reductases, we can predict all four nicotinamide nucleotides (NAD, NADH, NADP and NADPH) and can further suggest many nucleotides and nucleosides that are related to these as potential ligands, but it is hard to tell which one actually interacts with protein or improves crystallization until we screen each protein's crystallization conditions with all of them. Similarly, when all the ligands bound to structures of proteins in the PDB with 30% or higher amino acid sequence identity for each of first five proteins in Fig. 1 are listed, many potential ligands are predicted for each protein. In contrast, ligands that showed clear interaction with protein by our method often enhanced crystallization as shown in Fig. 2, allowing us to make progress up to X-ray data collection as shown in Table 2.

Inhibition studies of several enzymes such as phosphoglycerate kinase, phospho-glycerate mutase, cyclic 3',5'-monophosphate dependent protein kinase, myosin

**Fig. 2** Ligand-enhanced crystallization of *Mtb* recombinant proteins. Proteins were crystallized in the absence (Native, left) or presence (right) of various nucleotide ligands, as indicated. Ligand selection was based on the results of the ligand interaction assay (Fig. 1 and Table 1). Enhanced crystallization of protein–ligand co-crystals, relative to native protein alone, was observed for all eight proteins, and allowed for structural resolution of four proteins (resolution of the collected diffraction data is indicated in parentheses)



subfragment 1, chloroplast coupling factor 1 using Cibacron Blue F3GA dye indicate that the dye inhibits enzyme activities by pre-occupation of the ligand binding site [38–41]. In the dye-ligand affinity chromatography approach, with the assumption that protein elution occurs by replacement of dye with nucleotide ligand at nucleotide binding sites, we may be able to design and test inactive enzyme variants by mutation of residues that are critical for ligand interaction once the enzyme-ligand interaction mode is elucidated by solving the enzyme structure bound to the ligand.

#### Enhancement of crystallization by the ligands identified with dye-ligand elution

To examine the effects on crystallization of ligand(s) identified with our methods, nine *Mtb* proteins were crystallized without and with corresponding ligand(s) (Supplementary Table 1). Crystallization of six native proteins (without ligand) revealed no crystals in any condition and two (6-phosphogluconate dehydrogenase and secreted L-alanine dehydrogenase) generated relatively small crystals compared to proteins set up with ligands. In contrast, the co-crystallization of eight proteins (all but the conserved hypothetical protein) with ligands enhanced crystallization in various ways that we have never achieved with the native proteins (Fig. 2). Given the negative or poor crystallization results of native proteins relative to the results of protein with ligand(s) under identical crystallization conditions, we did not further pursue crystallization with native proteins by further screening of crystallization conditions or by protein modification. We categorize the enhancements by ligands in three groups: immediate formation of large new crystals (group 1), increases in crystal size (group 2), and the initiation of new crystals of small size (group 3). We found that of the proteins in group 1, NAD(P) transhydrogenase and short-chain-type dehydrogenase/reductase yielded new crystals diffracting to resolutions of 2.0–2.3 Å, whereas the crystals of short-chain-type oxidoreductase with NADPH diffracted but did not generate useful diffraction data and might need intensive screening of cryo-conditions. Of the proteins in group 2, larger crystals were obtained from 6-phosphogluconate dehydrogenase in the presence of NADH, yielding diffraction data to 2.7 Å resolution, and the crystals from secreted L-alanine dehydrogenase with NADH and ATP gave diffraction data to 2.0 Å and 2.4 Å resolution, respectively. When more than one ligand is identified, it may be useful to try tight-binding ligands first because they may yield better diffracting crystals [17]. The group 3 proteins methylmalonate-semialdehyde dehydrogenase, 5-methyltetra-hydropteroyl triglutamate-homocysteine methyltransferase, and 3-hydroxyisobutyrate dehydrogenase yielded protein crystals with the

ligands (NADH and NADPH, ATP and GTP, and NADH, respectively) that were confirmed by polarization and staining with dye, although those were too fragile or too small (5–10 µm in the largest dimensions) to harvest (the native proteins gave mere precipitation and no trace of crystals).

#### Conclusion

We have demonstrated here the applicability of a novel method, dye-ligand affinity chromatography, to ligand screening of nine recombinant *Mtb* proteins.

A unique feature of this dye-ligand affinity chromatography procedure is the generation of information about ligand interactions, depending on each ligand's ability to replace dye at the binding site of protein (Fig. 1). Depending on the character of the Cibacron Blue F3GA dye-protein interaction, the ligand-interaction profile obtained for a given protein can give clues to the protein's function.

The advantages of this method over other ligand analyses, such as the thermal shift assay with ligands [13] and the ligand chips [42] are that (a) no protein modification is required, (b) no special instruments and skill are required, and (c) the potential for direct monitoring of ligand interaction by displacement of the dye from the binding site of the protein. These advantages may allow the development of a high-throughput ligand analysis system that can be operated in a cost- and time-effective manner.

In this report, the ligand(s) identified by this method demonstrated a 50% success rate (5 of 10 tested proteins, including aldehyde dehydrogenase, reported previously by Kim et al. [17]) for obtaining diffraction data by crystallization improvement. The structure of each protein for which we have collected data has been solved and is in the process of refinement (these structures will be reported elsewhere).

According to earlier reports about the proteins interacting with Cibacron Blue F3GA dye [25–31], most nucleoside/nucleotide-interacting proteins, or about half of all enzymes [43], can be candidates for ligand analysis using the dye-ligand affinity chromatography method. Based on our data, the method seems especially efficient for identifying ligand(s) if the target protein is annotated as an NAD/NADH-binding (and NADP/NADPH-binding) protein.

We suggest that a useful structural genomics approach will be to use our methods to identify ligands for target proteins early in the process and then to co-crystallize target proteins with the identified ligand(s), potentially yielding functional insights and improving crystal formation and diffraction data.



**Acknowledgments** The authors are grateful to the Midwest Center for Structural Genomics for its gift of the pMCSG7 expression vector. We also would like to thank the staff at the BL 5.0.1 and BL 5.0.2—managed by the Berkeley Center for Structural Biology (BCSB) at the ALS—for technical support. The BCSB is supported in part by the National Institutes of Health, National Institute of General Medical Sciences. The ALS is supported by the director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. This work was in part supported by the LANL-UCR CARE program (STB-UC:06-29) and the NIGMS Protein Structure Initiative program (NIH U54 GM 074946).

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

## References

- Goh C-S, Lan N, Douglas SM, Wu B, Echols N, Smith A, Milburn D, Montelione GT, Zhao H, Gerstein M (2004) Mining the structural genomics pipeline: identification of protein properties that affect high-throughput experimental analysis. *J Mol Biol* 336:115–130
- Terwilliger TC, Stuart D, Yokoyama S (2009) Lessons from structural genomics. *Annu Rev Biophys* 38:371–383
- Newman J, Egan D, Walter TS, Meged R, Berry I, Jelloul MB, Sussman JL, Stuart DI, Perrakis A (2005) Expanding screening space through the use of alternative reservoirs in vapor-diffusion experiments. *Acta Cryst D* 61:1426–1431
- Bergfors T (2003) Seeds to crystals. *J Struct Biol* 142:66–76
- Deniaud A, Liguori L, Blesneac I, Lenormand JL, Pebay-Peyroula E (2010) Crystallization of the membrane protein hVDAC1 produced in cell-free system. *Biochim Biophys Acta* 1798:1540–1546
- Bursey EH, Kim CY, Yu M, Terwilliger TC, Hung LW (2006) An automated high-throughput screening method for the identification of high-yield, soluble protein variants using cell-free expression and systematic truncation. *J Struct Funct Genomics* 7:139–147
- Dong A, Xu X, Edwards AM, Chang C, Chruszcz M, Cuff M, Cymborowski M, Di Leo R, Egorova O, Evdokimova E, Filippova E, Gu J, Guthrie J, Ignatchenko A, Joachimiak A, Klostermann N, Kim Y, Korniyenko Y, Minor W, Que Q, Savchenko A, Skarina T, Tan K, Yakunin A, Yee A, Yim V, Zhang R, Zheng H, Akutsu M, Arrowsmith C, Avvakumov GV, Bochkarev A, Dahlgren LG, Dhe-Paganon S, Dimov S, Dombrowski L, Finerty P Jr, Flodin S, Flores A, Graslund S, Hammerstrom M, Herman MD, Hong BS, Hui R, Johansson I, Liu Y, Nilsson M, Nedyalkova L, Nordlund P, Nyman T, Min J, Ouyang H, Park HW, Qi C, Rabeh W, Shen L, Shen Y, Sukumard D, Tempel W, Tong Y, Tresagues L, Vedadi M, Walker JR, Weigelt J, Welin M, Wu H, Xiao T, Zeng H, Zhu H (2007) In situ proteolysis for protein crystallization and structure determination. *Nat Methods* 4:1019–1021
- Gao X, Bain K, Bonanno JB, Buchanan M, Henderson D, Lorimer D, Marsh C, Reynes JA, Sauder JM, Schwinn K, Thai C, Burley SK (2005) High-throughput limited proteolysis/mass spectrometry for protein domain elucidation. *J Struct Funct Genomics* 6:129–134
- Hu G, Lin G, Wang M, Dick L, Xu RM, Nathan C, Li H (2006) Structure of the *Mycobacterium tuberculosis* proteasome and mechanism of inhibition by a peptidyl boronate. *Mol Microbiol* 59:1417–1428
- Strong M, Sawaya MR, Wang S, Phillips M, Cascio D, Eisenberg D (2006) Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from *Mycobacterium tuberculosis*. *Proc Natl Acad Sci USA* 103:8060–8065
- Derewenda ZS (2004) Rational protein crystallization by mutational surface engineering. *Structure* 12:529–535
- Tereshko V, Uysal S, Koide A, Margalef K, Koide S, Kossiakoff AA (2008) Toward chaperone-assisted crystallography: protein engineering enhancement of crystal packing and X-ray phasing capabilities of a camelid single-domain antibody (VHH) scaffold. *Protein Sci* 17:1175–1187
- Vedadi M, Niesen FH, Allali-Hassani A, Fedorov OY, Finerty PJ Jr, Wasney GA, Yeung R, Arrowsmith C, Ball LJ, Berglund H, Hui R, Marsden BD, Nordlund P, Sundstrom M, Weigelt J, Edwards AM (2006) Chemical screening methods to identify ligands that promote protein stability, protein crystallization, and structure determination. *Proc Natl Acad Sci USA* 103:15835–15840
- Chayen NE, Saridakis E (2008) Protein crystallization: from purified protein to diffraction-quality crystal. *Nat Methods* 5:147–153
- Cooper GM, Brown CD (2008) Qualifying the relationship between sequence conservation and molecular function. *Genome Res* 18:201–205
- Stein L (2001) Genome annotation: from sequence to biology. *Nat Rev Genetics* 2:493–503
- Kim C-Y, Webster C, Roberts JKM, Moon JH, Lyon EZA, Kim H, Yu M, Hung L-W, Terwilliger TC (2009) Analysis of nucleoside-binding proteins by ligand-specific elution from dye resin: application to *Mycobacterium tuberculosis* aldehyde dehydrogenases. *J Struct Funct Genomics* 10:291–301
- McPherson A, Cudney B (2006) Searching for silver bullets: an alternative strategy for crystallizing macromolecules. *J Struct Biol* 156:387–406
- Jones CE, Brown AL, Baumann U (2007) Estimating the annotation error rate of curated GO database sequence annotations. *BMC Bioinformatics* 8:170–178
- Schnoes AM, Brown SD, Dodevski I, Babbitt PC (2009) Annotation error in public databases: misannotation of molecular function in enzyme superfamilies. *PLoS Comput Biol* 5:e1000605
- Chen L-L, Ma B-G, Gao N (2008) Reannotation of hypothetical ORFs in plant pathogen *Erwinia carotovora* subsp. *atroseptica* SCRI1043. *FEBS J* 275:198–206
- Vongsangnak W, Olsen P, Hansen K, Krogsgaard S, Nielsen J (2008) Improved annotation through genome-scale metabolic modeling of *Aspergillus oryzae*. *BMC Genomics* 9:245–258
- Yu J-F, Sun X (2010) Reannotation of protein-coding genes based on an improved graphical representation of DNA sequence. *J Comput Chem* 31:2126–2135
- Subramanian S (1984) Dye-ligand affinity chromatography: the interaction of Cibacron Blue F3GA with proteins and enzymes. *CRC Crit Rev Biochem* 16:169–205
- Arica MY, Testereci HN, Denizli A (1998) Three-dimensional structures of apo- and holo-L-alanine dehydrogenase from *Mycobacterium tuberculosis* reveal conformational changes upon coenzyme binding. *J Chromatogr A* 799:83–91
- Dean PDG, Watson DH (1979) Protein purification using immobilised triazine dyes. *J Chromatogr* 165:301–319
- Denizli A, Pişkin E (2001) Dye-ligand affinity systems. *J Biochem Biophys Methods* 49:391–416
- Denizli A, Salih B, Piskin E (1997) Cibacron Blue F3GA and Zn(II) containing poly(ethylene glycol-dimethacrylate-hydroxyethylmethacrylate) microbeads for albumin adsorption. *J Macromol Sci Pure Appl Chem A34:1353–1368*

29. Denizli A, Yavuz H, Arica Y (2000) Monosize and non-porous p(HEMA-co-MMA) microparticles designed as dye- and metal-chelate affinity sorbents. *Colloids Surf A* 174:307–317
30. Denizli A, Yavuz H, Garipcan B, Arica MY (2000) Non-porous monosize polymeric sorbents: dye and metal chelate affinity separation of lysozyme. *J Appl Polym Sci* 76:115–124
31. Lowe CR, Small DAP, Arkinson A (1981) Some preparative and analytical applications of triazine dyes. *Int J Biochem* 13:33–40
32. Roberts JKM, Webster C, Terwilliger TC, Kim C-Y (2008) High-throughput analysis of nucleoside- and nucleotide-binding by proteins. In: Hicks MG, Kettner C (eds), *Proceedings of systems chemistry symposium*, pp. 207–219, Beilstein Institute, Bozen. <http://www.beilstein-institut.de/index.php?id=222&L=3>
33. Stols N, Gu M, Dieckman L, Raffin R, Collart FR, Donnelly MI (2002) A new vector for high-throughput, ligation-independent cloning encoding a tobacco etch virus protease cleavage site. *Protein Expr Purif* 25:8–15
34. Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227:680–685
35. Otwinowski Z, Minor X (1997) *Methods in enzymology*, 276. In: Carter CW Jr, Sweet RM (eds) *Macromolecular crystallography*. Academic Press, New York, pp 307–326
36. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The protein data bank. *Nucleic Acids Res* 28:235–242
37. Agren D, Stehr M, Berthold CL, Kapoor S, Oehlmann W, Singh M, Schneider G (2008) Three-dimensional structures of apo- and holo-L-alanine dehydrogenase from *Mycobacterium tuberculosis* reveal conformational changes upon coenzyme binding. *J Mol Biol* 377:1161–1173
38. Kawai K, Eguchi Y (1980) Effect of Cibacron Blue F3GA on phosphoglycerate kinase of *Lactobacillus plantarum* and phosphoglycerate mutase of *Leuconostoc dextranicum*. *J Biochem* 88:1227–1231
39. Loyola-Guzman G, Vallejos RH (1983) Binding of Cibacron blue F3GA to chloroplast coupling factor. Competitive inhibition of ATP synthesis and hydrolysis. *Biochim Biophys Acta* 724: 427–432
40. Reisler E, Lie J (1981) Interaction of myosin subfragment 1 with Cibacron Blue F3GA. *Biochemistry* 20:6745–6749
41. Witt JJ, Roskoski RJ (1980) Adenosine cyclic 3',5'-monophosphate dependent protein kinase: active site directed inhibition by Cibacron Blue F3GA. *Biochemistry* 19:143–148
42. Goldberg MD, Lo RC, Abele S, Macka M, Gomez FA (2009) Development of microfluidic chips for heterogeneous receptor-ligand interaction studies. *Anal Chem* 81:5095–5098
43. Scopes RK (1994) *Protein purification: principles and practice*, 3rd edn. Springer, NY