J Struct Funct Genomics (2010) 11:91–100
DOI 10.1007/s10969-010-9087-6

X-ray structure determination of the glycine cleavage system protein H of *Mycobacterium tuberculosis* using an inverse Compton synchrotron X-ray source

Jan Abendroth · Michael S. McCormick · Thomas E. Edwards · Bart Staker · Roderick Loewen · Martin Gifford · Jeff Rifkin · Chad Mayer · Wenjin Guo · Yang Zhang · Peter Myler · Angela Kelley · Erwin Analau · Stephen Nakazawa Hewitt · Alberto J. Napuli · Peter Kuhn · Ronald D. Ruth · Lance J. Stewart

Received: 30 January 2010 / Accepted: 10 March 2010 / Published online: 3 April 2010

Abstract Structural genomics discovery projects require ready access to both X-ray diffraction and NMR spectroscopy which support the collection of experimental data needed to solve large numbers of novel protein structures. The most productive X-ray crystal structure determination laboratories make extensive use of tunable synchrotron X-ray light to solve novel structures by anomalous diffraction methods. This requires that frozen cryo-protected crystals be shipped to large multi acre synchrotron facilities for data collection. In this paper we report on the development and use of the first laboratory-scale synchrotron light source capable of performing many of the state-of-the-art synchrotron applications in X-ray science. This Compact Light Source is a first-in-class device that uses inverse Compton scattering to generate X-rays of sufficient flux, tunable wavelength and beam size to allow high-resolution X-ray diffraction data collection from protein crystals. We report on benchmarking tests of X-ray diffraction data collection with hen egg white lysozyme, and the successful high-resolution X-ray structure determination of the Glycine cleavage system protein H from *Mycobacterium tuberculosis* using diffraction data collected with the Compact Light Source X-ray beam.

Keywords Compact light source · X-ray crystallography · Inverse Compton

Lance J. Stewart is corresponding author for the protein structure research aspects of the publication and Ronald D. Ruth is corresponding author for the design and function of the CLS aspects of the publication.

J. Abendroth · T. E. Edwards · B. Staker · L. J. Stewart (✉)
Emerald BioStructures, 7869 NE Day Road West, Bainbridge Island, WA 98110, USA
e-mail: LStewart@embios.com

M. S. McCormick · P. Kuhn
The Scripps Research Institute, 10550 N. Torrey Pines Road, La Jolla, CA 92037, USA

R. Loewen · M. Gifford · J. Rifkin · R. D. Ruth (✉)
Lyncean Technologies, 370 Portage Ave, Palo Alto, CA 94306, USA
e-mail: Ronald_Ruth@lynceantech.com

C. Mayer · W. Guo · Y. Zhang · P. Myler
Seattle Biomedical Research Institute, 307 Westlake Ave N., Suite 500, Seattle, WA 98109, USA

A. Kelley · E. Analau · S. N. Hewitt · A. J. Napuli
University of Washington, Seattle, WA 98195, USA

R. D. Ruth
SLAC National Accelerator Laboratory, Stanford, CA 94305, USA

Introduction

Protein crystallography with the Compact Light Source

Over the past 30 to 40 years, synchrotron X-ray sources have had a rapidly growing impact on many fields of science. Their role within the biological community has been very broad and includes many novel techniques that take advantage of the coherence, intensity and brilliance of these X-ray sources, such as phase contrast imaging of soft tissue and macromolecular crystallography [7].

The Protein Structure Initiative (PSI), led by the National Institute of General Medical Sciences (NIGMS), has placed an extraordinary emphasis on the role of synchrotron light sources in collecting the vast amounts of data necessary for

determining the three dimensional structure of a wide variety of proteins and macromolecules. To further extend the influence of the special techniques developed at the large synchrotrons, we have developed a laboratory-scale synchrotron light source, called the Compact Light Source (CLS) capable of performing many of the state-of-the-art synchrotron applications in X-ray science.

The initial development of the prototype CLS was funded by the NIGMS through Small Business Innovation Research (SBIR) grants to Lyncean Technologies, Inc. (LTI). Over the last 4 years the development of the CLS for protein crystallography applications has been funded by the Accelerated Technologies Center for Gene to 3D Structure (ATCG3D), a PSI-2 Specialized Center that is supported by both the NIGMS and the National Center for Research Resources (NCRR). The ATCG3D consortium is focused on the accelerated development, integration, and deployment of emerging technologies that have high potential to improve the economics of protein structure determination by X-ray crystallographic methods.

Here we report on the first X-ray crystal structures from X-ray diffraction data collected from the NIH-funded, commercially developed, miniature synchrotron—the Compact Light Source (CLS). The CLS was first used to collect X-ray diffraction data for crystals of hen egg white lysozyme, a benchmarking protein commonly used in the development of new technologies related to protein X-ray crystallography. We then went on to use the CLS to collect X-ray diffraction data for crystals of the Glycine cleavage system protein H from *Mycobacterium tuberculosis* (MytuGCSPH), whose structure was solved in collaboration with the Seattle Structural Genomics Center for Infectious Disease. The resulting 2.0 Å resolution protein structure was published to the RCSB Protein Data Bank (PDB) (PDBID: 3IFT) together with a sister PDB entry, a second crystal of the same protein solved at 1.75 Å using X-ray diffraction data collected using a current state-of-the-art in-house (non synchrotron) X-ray source (PDB ID:3HGB). The Seattle Structural Genomics Center for Infectious Disease (SSGCID) project is funded by the National Institute of Allergy and Infectious Disease (NIAID) with the aim of solving protein structures that are relevant for potential structure-based drug design of new antimicrobial agents.

Background

While the total flux from a typical home laboratory (homelab, rotating anode) X-ray source may be quite large, the necessary filtering and focusing results in limited monochromatic flux at prescribed, non-tunable energies. As a consequence, long exposure times are required to record a crystallographic data set in comparison to the speed with which a high resolution data set can be collected using intense X-rays generated at large national synchrotron facilities. In addition, the lack of tunability limits homelab sources from fully exploiting techniques such as single wavelength anomalous dispersion (SAD) and multiple wavelength anomalous dispersion (MAD). Homelab X-ray sources have therefore been used heavily for screening crystals for diffraction quality prior to study at a synchrotron where cryopreserved crystals are sent by shipment in dry-shipper containers held at liquid nitrogen temperatures for data collection.

However, the difficulties of conducting research at a remote location continue to pose challenges for the structural biologist, many of whom would still prefer to conduct research in their own laboratories. In fact a common sentiment voiced among synchrotron users is the wish for a local, tunable source, with suitable properties to perform their favorite synchrotron-developed applications. This sentiment has driven synchrotron beamlines to invest heavily in remote, web-enabled data collection services using software such as Web-ICE.

The idea for the CLS is a by-product of high-energy physics research at the Stanford Linear Accelerator Center (SLAC) aimed at producing low-emittance electron beams [8]. Later, further research at SLAC explored the viability of using a miniature electron storage ring and a high-finesse optical cavity optimized for creating high average flux in the hard X-ray regime [13]. These early studies led to the formation of Lyncean Technologies, Inc., which developed the CLS based on the prior concept but completely re-designed and engineered for commercialization. By combining a "low-energy" (MeV scale) electron storage ring and a high-power optical cavity, X-rays are generated when the electrons and photons collide through an interaction known as inverse Compton scattering. While this phenomenon had been originally studied by colliding high-power laser pulses with electrons from a linear accelerator, the average flux was always too low to be relevant for most applications. However, using an electron storage-ring, together with the optical cavity, can provide several orders of magnitude higher X-ray flux; enough to yield a practical X-ray source [8].

Existing synchrotron light sources employ multi-GeV electron beams stored in large rings of magnets. Special insertion devices, such as undulator magnets, cause the electron beam to wiggle and produce the characteristic radiation of synchrotron light. For example, to produce 1 Å wavelength radiation with a 2 cm-period undulator magnet requires an electron beam with an energy of several GeV stored in a ring that is about 300 m in diameter. However, a laser beam colliding with a electron beam has the same effect as an undulator magnet. The electric and magnetic fields of the laser pulse cause the electrons to wiggle, inducing a radiation spectrum similar to a long undulator magnet. To produce 1 Å radiation with a laser wavelength

of 1 micron requires an electron beam energy of only 25 MeV, reducing the scale of the storage ring by a factor of about 200; a 25 MeV electron storage ring is small enough to fit within the footprint of a large desk.

The CLS was designed by LTI to operate over a wide range of electron energies (20–45 MeV) and can be tuned to produce X-rays in the range between 10 keV (1.23 Å) and 35 keV (0.35Å). The X-rays are directed in a narrow cone in the direction of the electron beam. The spot size of the X-ray beam is largely determined by the focused spot of the laser beam, which has a radius of 50 μm rms, while the divergence of the X-ray beam, a few mrad, is largely determined by the divergence of the electron beam at the interaction point. The electron beam divergence and energy spread determine the natural energy bandwidth of the X-rays of about 3–4%.

For crystallography experiments, a set of 1:1 multilayer X-ray mirrors are placed near the X-ray output of the CLS to focus the emitted cone beam back to its original source size at a focal point located approximately 2 m from the optics. This also results in a bandwidth filter which decreases the X-ray beam energy spread to the range of ~0.1–1%.

The Compact Light Source: technology

A CAD engineering model of the CLS system, including an experimental area for crystallography, is shown to scale in Fig. 1. The CLS itself consists of an electron beam injector, a storage ring, and an optical cavity system. The injector is shown with two commercial high-power microwave sources that power the accelerator sections. The electron storage ring and optical cavity circulate beams which synchronously interact at one particular location to produce X-rays. The entire CLS is contained within a shielded enclosure, shown partially transparent in Fig. 1. The narrow cone of

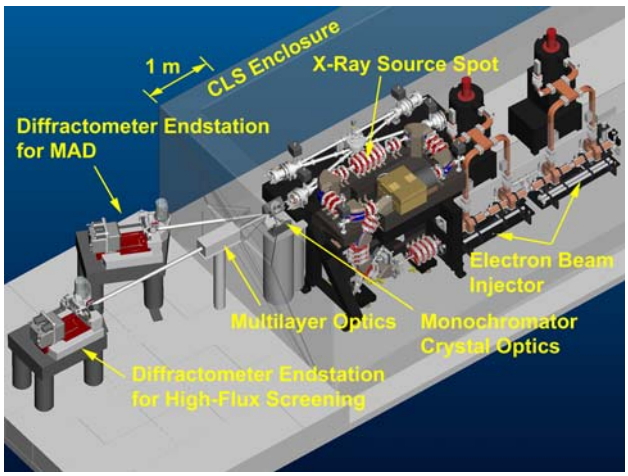

Fig. 1 CAD drawing of the Compact Light Source together with endstation. The experimental area is separated by a shielding wall (shown mostly transparent)

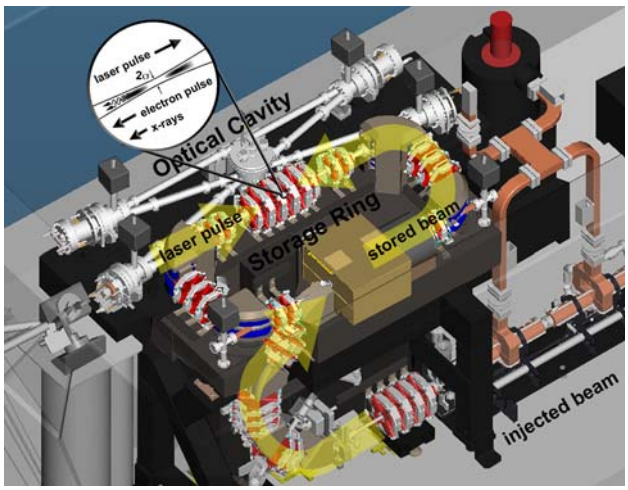

Fig. 2 CAD drawing of the Compact Light Source illustrating the laser-electron pulse interaction. Major components are the injector (electron gun not shown), the electron storage ring, and the optical cavity. Electron-photon scattering at the interaction point produces naturally collimated, narrow bandwidth X-rays that exit a window at left of figure. The storage ring is a *rectangle* of approximately 1 m by 2 m

X-rays exits through a window towards sets of X-ray optics which can monochromatize and focus the beam. The X-ray beam is transported through evacuated chambers to end stations (shown with *MarDTB* and *Rayonix* MX-225 CCD) where it can be used for data collection.

A more detailed view of the CLS electron storage ring and optical cavity is shown in Fig. 2. The electron source produces a single electron bunch using an RF gun with a laser photocathode. A short linear accelerator then accelerates the electron bunch to the full energy desired in the ring—20 to 45 MeV. This electron bunch is injected and stored in the miniature storage ring for about one million turns, and the injector periodically refreshes the electron bunch in the storage ring to maintain high beam quality and intensity. A typical re-injection rate of 30 Hz is sufficient to provide a stable, continuous, circulating electron bunch, with little noticeable X-ray intensity variation between injection cycles. (The rapid reinjection is required for a low-energy storage ring due to natural gas scattering even under ultra-high vacuum conditions.)

On one side of the ring is a straight section in which the electron beam is transversely focused to a small spot. This straight section also serves as one leg of the optical gain-enhancement cavity for the laser pulse. The electron bunch and the laser pulse collide each turn at the interaction point producing a burst of X-rays. The high flux of the CLS results from the high circulation rate of the storage ring and cavity.

The native CLS X-ray output is summarized in Table 1. Present flux has not yet reached design values, although the intensity has been already sufficient to demonstrate

Table 1 CLS native X-ray beam output parameters

Parameter	Present	Planned	Notes
Total flux	$\sim 10^{11}$ ph/s	$\sim 10^{13}$ ph/s	Full bandwidth
Total flux (output BW)	$\sim 10^{9}$ ph/s	$\sim 10^{11}$ ph/s	3–4% bandwidth
Flux on crystal	$\sim 10^{7}$ ph/s	$\sim 10^{9-10}$ ph/s	0.1–1% bandwidth
Source spot size	50 μm rms	30 μm rms	Also image size for 1:1 optics
Source divergence	~2.5 mrad	~2.0 mrad	
X-ray energy range	10–20 keV	7–35 keV	Tunable

published results for Differential Phase Contrast Imaging [2, 3] in addition to crystallography. Ongoing development and funded hardware upgrades of the CLS are targeting the future performance levels listed in Table 1.

External X-ray optics provide further manipulation of the CLS X-ray output. The cone beam can be intercepted just past the output window of the device and collimated or focused, with varying energy bandwidths depending on the type of optics employed. The reduction in flux should be proportional to the reduction in bandwidth of the output beam, limited only by the efficiency of the X-ray optics. Although the prototype optics that were used for the experiments described here were very inefficient, new X-ray optics are being manufactured that will provide close to optimal performance for bandwidths ranging from 0.5 to 1.5%.

Images of selected components of the CLS and the native X-ray spectrum are shown in Fig. 3.

The glycine cleavage system

The Glycine cleavage system [11] is the major route for glycine catabolism. The four-enzyme complex catalyzes in a reversible manner the conversion of glycine, tetrahydrofolate and NAD^+ to 5,10-methylene-tetrahydro-folate, carbon dioxide, ammonia, and NADH. The four proteins of the complex include three enzymes and one carrier protein: The P-protein (200 kDa, multimer) is a glycine dehydrogenase which contains pyridoxal phosphate; the T-protein (40 kDa) is an aminomethyl-transferase and carries out the tetrahydrofolate-dependent reaction; the L-protein is a dihydrolipamide dehydrogenase; the H-protein (14 kDa) shuttles reaction intermediates and reduction equivalents between the different components of the system.

The first crystal structure of a Glycine cleavage system protein H was determined at 2.6 Å resolution by Pares et al. [20], PDBID code 1HPC. The protein was extracted

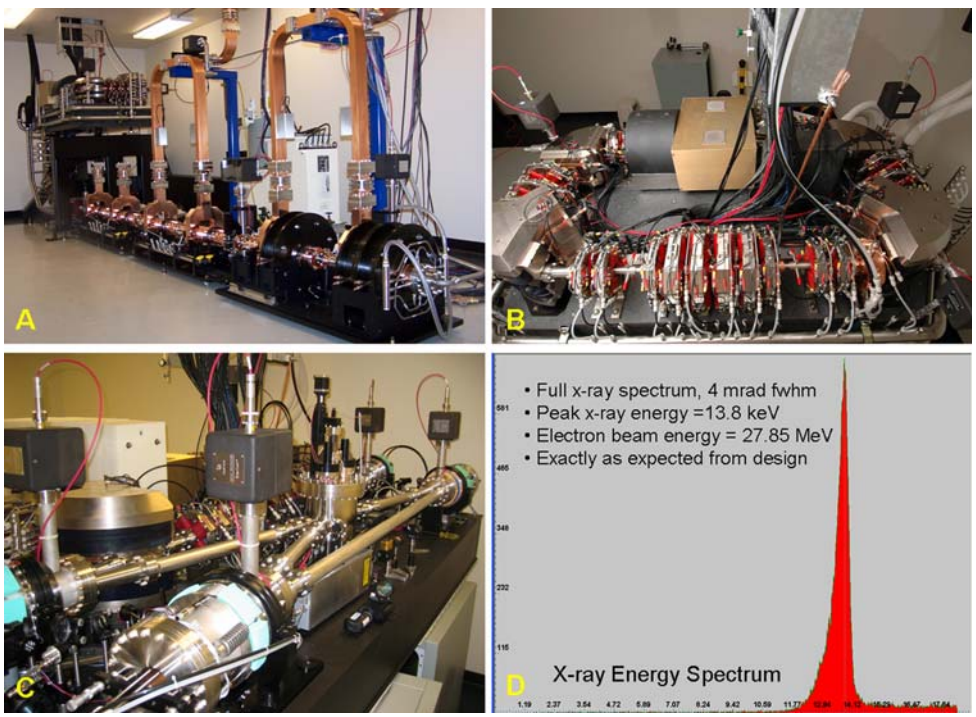

Fig. 3 **a** CLS injector, **b** electron storage ring, **c** CLS optical cavity, **d** energy spectrum of X-rays from CLS output

from pea leaves and has a lipoate moiety bound to Lys 63, called the lipoyl-lysine arm. Since, the structure of the protein from pea leaves has been determined in various forms: oxidized [4] (PDBID:1HPC), reduced [6] (PDBID:1DXM), apo [14] and bound to the reaction intermediate methylamine [4] (PDBID:1HTP). Crystal structures from *Themus thermophilus* [17] (PDBID:1ONL) and *Thermotoga maritima* (PDBID:1ZKO) have been determined in the recent years.

The Seattle Structural Genomics Center for Infectious Disease (SSGCID) is a collaborative center funded by the National Institute of Allergy and Infectious Disease (NIAID) to solve numerous protein structures of infectious disease targets from NIAID category A-C threat agents. The SSGCID targets are all selected with the long term goal that they can represent a blueprint for future structure-guided drug design. The Glycine cleavage system in bacteria has been selected by SSGCID as a pathway that could potentially be disrupted by a small molecule anti-bacterial agent that targets one or another of the components of the Glycine cleavage system, with the goal that such a compound would compromise cell viability. In order to provide a variety of comparative structural information on selected targets, the SSGCID has drafted numerous homologous targets from a variety of our selected pathogen organisms, which includes *Mycobacterium tuberculosis*, the causative agent of tuberculosis (TB). The SSGCID strives to avoid duplication of effort with other structural genomics centers by coordinating target selection through the PSI funded TargetDB, a database of targets that allow centers to communicate progress towards structure determination of selected targets (http://tagetdb.PDB.org). At the time of initiating our SSGCID project, the Glycine cleavage protein H of *Mycobacterium tuberculosis* was not being worked on by other structural genomics centers, and therefore it was selected for structure determination by SSGCID.

Materials and methods

Protein expression and purification

The target gene was PCR amplified from purified *Mycobacterium tuberculosis* strain H37Rv genomic DNA (gift from David Sherman, SBRI) using the following primers: MytuD.01046.a.A1_FWD (5′GGGTCCTGGTTCGATGG TGAGCGATATCCCGTCCG) and MytuD.01046.a.A1_ REV (5′CTTGTTCGTGCTGTTTATTACTCGGTCAGTG TGCCGCGGT). The purified amplicon was subsequently cloned into the bacterial expression vector AVA0421 using a ligation-independent cloning (LIC) methodology [1] and transformed into the amplification host NovaBlue (Novagen). The expression vector AVA0421 is derived from pET14b, regulated by the T7 promoter, and contains the *bla* gene conferring ampicillin resistance. AVA0421 yields protein constructs with an N-terminal His_6-tag and a 3C protease cleavage site: MAHHHHHHMGTLEAQTQ GPGSM-ORF. Cleavage of the His_6-tag by 3C protease yields proteins with an N-terminal sequence: GPGSM-ORF. The resultant plasmid was transformed into the Rosetta Oxford bacterial host (gift from Ray Hue, Structural Genomics Consortium), grown in auto-induction medium [22], the cells lysed, the supernatant passed over Ni^{2+} beads, and soluble protein quantified by SDS–PAGE to evaluate expression. Glycerol stocks were made at this stage and DNA prepared for sequencing to confirm that the correct target has been cloned and does not contain frameshifts or premature stop codons.

The frozen cells were thawed and resuspended in 200 ml of Lysis Buffer (20 mM HEPES, pH 7.2–7.4, 300 mM NaCl, 5% glycerol, 30 mM Imidazole, 0.5% CHAPS, 10 mM $MgCl_2$, 3 mM β-mercaptoethanol, 25 units/ml of Benzonase® nuclease, and 0.05 mg/ml lysozyme). The resuspended cell pellet was then disrupted on ice for 15 min with a Branson Digital Sonifier 450D (settings at 70% amplitude, with alternating cycles of 5 s of pulse-on and 10 s of pulse-off). The cell debris was clarified by centrifugation on a Sorvall RC5 at 6,000 RPM for 60 min at 4°C. The hexa-histidine tagged protein of interest was purified from the clarified cell lysate by immobilized metal affinity chromatography binding on Ni Sepharose High Performance resin (GE Biosciences, Piscataway, NJ) equilibrated with Binding Buffer (20 mM HEPES, pH 7.2–7.4, 300 mM NaCl, 5% glycerol, 30 mM Imidazole). The recombinant protein was eluted in 500 mM imidazole plus 5 mM β-mercaptoethanol and was further resolved by size-exclusion gel chromatography (SEC, Superdex 75 26/60; GE Biosciences, Piscataway, NJ). Pure fractions collected in SEC Buffer (20 mM HEPES pH 7.0, 300 µM NaCl, 2 mM DTT, and 5% glycerol) as a single peak were analyzed by sodium dodecyl sulfate–polyacrylamide (SDS) gel electrophoresis and Invitrogen Corp. SimplyBlue Safestain. The protein was then pooled, concentrated, flash frozen and stored at −80°C in SEC Buffer.

Protein crystallization

Lysozyme

Hen egg white lysozyme (Sigma) was crystallized using the hanging drop vapor diffusion method at room temperature. Protein solution contained 100 mg/ml protein in 100 mM sodium acetate trihydrate (Fluka) at pH 4.8. Precipitant solution contained 30% (w/v) polyethylene glycol monomethyl ether 5,000 (Fluka), 1.0 M sodium chloride (Fisher Scientific), and 50 mM sodium acetate

trihydrate at pH 4.8. Hanging drops comprised 2 µl protein solution added to 2 µl precipitant solution, suspended over 1.0 ml precipitant solution. Crystals formed within minutes and were harvested after overnight growth. The 0.6 × 0.6 × 0.6 mm crystalline sample used for data collection was transferred to a cryogenic solution containing the precipitant solution components with added 20% (v/v) glycerol (Fisher Scientific), flash frozen in liquid nitrogen, and immediately mounted on the goniometer head in cryo stream at 100 K prior to diffraction data collection.

MytuGCSPH

The purified MytuGCSPH protein was concentrated to 27 mg/ml, and various commercial crystallization screens were set up using 0.4 µl + 0.4 µl drops and Emerald BioSystems Compact Junior 96 well crystallization plates. Initial crystal hits were obtained in the JCSG+ screen (Emerald BioSystems) [18], condition B11: 1.6 M Na-citrate. The crystals for the data set collected in-house were obtained directly from the screen and vitrified without further cryo-protection. Optimized crystals were obtained with the following reservoir solution: 1.5 M Na-citrate, 50 mM Bis–Tris pH 6.3. The crystals grew to a size of up to 250 µm in each direction, see Fig. 4. The optimized crystals were cryopreserved by freezing the crystals in liquid nitrogen following their transfer into a cryopreservation solution comprised of three parts of the reservoir crystallization cocktail mixed with one part of ethylene glycol as a cryoprotectant such that it had a final concentration of 25% v/v ethylene glycol. The cryopreserved crystals, mounted in nylon loops were used for data collection with the CLS.

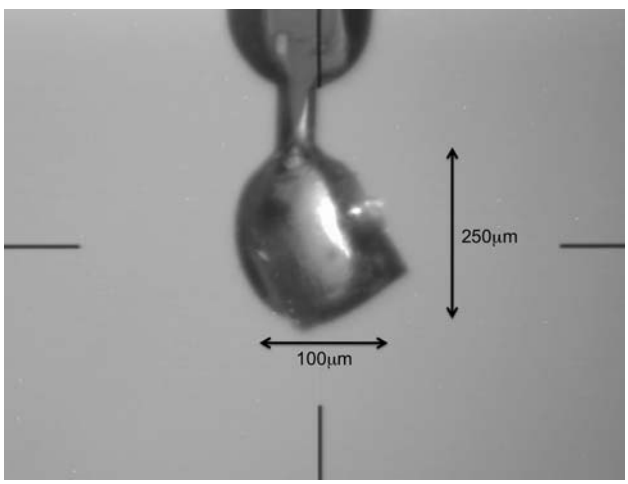

Fig. 4 Crystal of MytuGCSPH cryo-protected and mounted to the goniometer head of the compact light source. The size of the crystal is about 250 × 250 × 100 µm

Data collection

Experimental set up

For crystallography experiments, a set of 1:1 crystal or multilayer X-ray mirrors can be placed near the X-ray output of the CLS which focuses the emitted cone beam back to its original source size at a focal point located approximately 2 m from the optics. Because the X-ray radiation from the CLS is produced in only a narrow bandwidth, there are no thermal or other shielding issues associated with the optics. The focused beam can then be directly aligned onto a commercial goniostat providing extremely low background radiation on the detector. In the following experiments, the X-ray beam path was through air with no vacuum or Helium enclosure. At 15 keV, this resulted in an intensity attenuation of about 30%.

Lysozyme data collection-CLS

For the lysozyme crystal experiment, the CLS X-ray beam was focused by two bent silicon crystals (Si 111). These were arranged in a K-B configuration following a design by Prof. Jens Als-Nielsen, and manufactured by JJ X-ray (Lyngby, Denmark). The end station was a *MarDTB* with a *Rayonix* SX-165 CCD detector (Rayonix LLC, Evanston IL, USA). The measured beam spot sizes on the detector were approximately 200 µm FWHM, dominated by the point spread function. The X-ray beam size at the crystal was about 120 µm FWHM.

Due to figure errors in the optics, the flux was decreased from 10^9 down to 3×10^5, a reduction of about 3,000. This loss is far more than expected from bandwidth reduction. These X-ray optics will not be used in future experiments.

X-ray diffraction data collection occurred over a 2-day period while the lysozyme crystal was held in the cryostream of nitrogen gas at 100 K. Ninety images were collected using a 1.0° oscillation range per frame and an exposure time of 8 min. Although the CLS runs stably for long periods, the data collection was performed over 2 days during daytime business hours to comply with Lyncean's lease terms. Additional data collection information and statistics are reported in Table 2.

MytuGCSPH (MytuD.01046.a) data collection: CLS

For the MytuGCSPH crystal experiment the CLS X-ray beam was focused by two bent multilayers on loan from the Niels Bohr Institute, University of Copenhagen, Denmark. These were developed by Anette Jensen and Jens Als-Nielsen [9]. For this experiment, the endstation location was shifted to the angle appropriate for the optics, but otherwise the set up was identical to that for the lysozyme experiment.

Table 2 Diffraction data and refinement statistics

Data collection	Lysozyme	MytuGCSPH	MytuGCSPH
Beamline	CLS	CLS	Rotating anode
Wavelength [Å]	0.92939	0.81836	1.5418
Space group	$P4_32_12$	$C2$	$C2$
a [Å]	78.61	86.45	86.86
b [Å]	78.61	51.01	51.47
c [Å]	37.30	32.57	32.53
beta [°]	90	95.10	94.64
Resolution [Å]	50–2.80 (2.87–2.80)	20–2.00 (2.05–2.00)	20–1.75 (1.80–1.75) [2.02–1.96]
Reflections (unique)	3,131 (156)	9,647 (705)	14,091 (1,067) [943]
Redundancy	4.0 (3.9)	3.1 (3.1)	2.3 (1.4) [1.7]
Completeness [%]	98.9 (99.6)	98.8 (99.3)	96.9 (97.7) [97.4]
I/sigma	12.7 (4.2)	10.2 (3.3)	23.8 (6.2) [10.7]
Rsym	0.105 (0.268)	0.095 (0.385)	0.027 (0.088) [0.053]
Refinement			
R_{work}	*0.209*	0.181	0.181
R_{free}	*0.278*	0.253	0.220
RMSD bonds [Å]	*0.012*	0.016	0.008
RMSD angles [°]	*1.37*	1.59	1.19
RMSD chirals [Å^3]	*0.104*	0.076	0.073
Ramachandran plot			
Preferred	*119 (94.4%)*	129 (97.0%)	121 (96.8%)
Allowed	*7 (5.6%)*	3 (2.3%)	4 (3.2%)
Outliers	*0 (0%)*	1 (0.7%)	0 (0%)
PDB ID code	NA	3IFT	3HGB

The benchmark data set of lysozyme was not fully refined and the structure not deposited. Hence, the values in the table are italicized

Numbers in parenthesis represent highest resolution shell of data, numbers in brackets for the high resolution data set for MytuGCSPH represent the resolution shell that roughly corresponds to the highest resolution shell for the CLS data set

$R_{merge} = (|\Sigma I_{hkl} - \langle I \rangle|/(\Sigma I_{hkl})$, where the average intensity $\langle I \rangle$ is taken over all symmetry equivalent measurements and I_{hkl} is the measured intensity for any given reflection

$I/\sigma I$ is the mean reflection intensity divided by the estimated error

$R_{work} = ||F_o| - |F_c||/|F_o|$, where F_o and F_c are the observed and calculated structure factor amplitudes, respectively

R_{free} is equivalent to R_{cryst} but calculated for 5% of the reflections chosen at random and omitted from the refinement process

The measured spot size on the detector was again around 200 μm FWHM. The X-ray beam on the crystal was about 120 μm FWHM. The multilayer X-ray optics resulted in a factor of 200 reduction in flux (from 10^9 down to 5×10^6 ph/s), once again far more than expected from simple bandwidth reduction. This is understandable because they were developed for a rotating anode source. In order to reduce ice rings, the cryopreserved mounted crystal was annealed twice by allowing the crystal to thaw at room temperature followed by re-application of the cryo-stream nitrogen gas to rapidly freeze and maintain the crystal at 100 K.

MytuGCSPH (MytuD.01046.a) data collection: rotating anode

A data set for a MytuGCSPH crystal was initially collected using a conventional rotating anode (Rigaku FR-E+ Superbright) equipped with a Rigaku Saturn 944+ CCD detector. A 300 μm collimator was used. In order to include high angle spots, the detector was swung in 2θ by 10°. 180 images were collected in 1° wedges with 10 s exposure per image.

Structure solution and refinement

Lysozyme

Diffraction images for lysozyme were analyzed and processed in HKL2000 [19]. The best 60 images from the 90-image data collection were selected to yield a data set up to 2.8 Å resolution.

The Molecular Replacement was done with PHASER [15], using published a 1.65 Å resolution structure of tetragonal lysozyme (PDB code 2HU1) with all non-

protein atoms removed as the search model. Rotation and translation function values for the molecular replacement were 12.2 and 29.3, respectively. The resulting initial structure was fit using COOT [5] and refined using REFMAC5 [16], to yield the refinement statistics reported in Table 2.

MytuGCSPH

The diffraction data sets for the MytuGCSPH crystals were reduced with the XDS [10] package and scaled with XSCALE [10] to 1.75Å (rotating anode) and 2.0Å (CLS), respectively. The diffraction statistics are summarized in Table 2.

The Molecular Replacement for both data sets was done with PHASER [15] using the PDB-entry 1ONL (in-house data, GCSPH from *Thermus thermophilus*, 51% sequence identity) or 3HGB (in-house structure for the CLS data set) as the search model. The models were iteratively built with COOT [5] and refined with REFMAC5.5.0088 [16]. The structures were refined to Rwork = 0.181 and Rfree = 0.220 (in-house data) and R Rwork = 0.181 and Rfree = 0.253 (CLS data) with good stereochemistry (see Table 2). Coordinates and structure factors were deposited in the PDB with the PDB-code 3HGB (rotating anode data) and 3IFT (CLS data), respectively.

Results and discussion

As of mid-2009, the Compact Light Source prototype had reached a number of milestones in both protein crystallography and biological imaging. The CLS prototype development began with initial funding in 2002 and after 1 year of design work and 2 years of construction, the prototype electron storage ring first achieved a stored electron beam in June of 2005. The optical cavity system required further development, but it was installed 6 months later, and on February 23, 2006 the CLS prototype achieved its first X-ray beam. The first test of prototype crystal optics followed in June of 2006 when the CLS achieved its first focused X-ray beam. The next year focused on studies of the X-ray flux and spectrum. In June of 2007, the CLS prototype was used in its first science experiment to study Differential Phase Contrast Imaging (DPCI) [21]. One year later in June of 2008, the first crystal diffraction was achieved with a small molecule. Several experiments followed in 2008 including the second DPCI experiment and the first protein diffraction and partial data set in November and December of 2008. The next year, 2009, began with the first scientific publication from the CLS which reported on the imaging results from the DPCI experiment [2, 3]. In March 2009 a trial lysozyme data set was collected and later in July, data for the first novel protein was collected and reported herein.

Before the detailed discussion of the last two experiments, it is important to note that in parallel to the CLS prototype experimental development, the ATCG3D 'Beta' CLS was under construction starting in 2005. This Beta CLS took advantage of the lessons learned from the prototype, and the prototype took advantage of the Beta development. Several ATCG3D subsystems were installed on the CLS prototype and used for the experiments described below.

After the successful completion of a trial lysozyme data set in March 2009, the CLS prototype at Lyncean Technologies was prepared to collect data on a novel, unsolved protein. The ATCG3D/SSGCID collaboration proposed a suitable target (MytuGCSPH, target ID MytuD.01046.a), samples of which were sent to Lyncean Technologies in July of 2009. After screening a dozen crystals, a candidate crystal was selected and mounted to a commercial goniostat, the MarDTB, outfitted with a Rayonix SX-165 detector. The crystal size ($\sim$250 $\times$ 250 $\times$ 100 microns) was well matched to the focused X-ray beam ($\sim$200 micron FWHM).

The CLS was tuned-up for X-rays at 15.1 keV (0.818 A) with a bandwidth of 1.4% (due to the multilayer optics) for this data collection, comprising 152 images. Each image was taken with 6 min exposure times, with times-two dezingering. Diffraction spots extended to 2.0 Å resolution (Fig. 5). The entire data set was taken during normal daytime working hours over a 3-day span. The structure was solved using Molecular Replacement with 3HGB; the

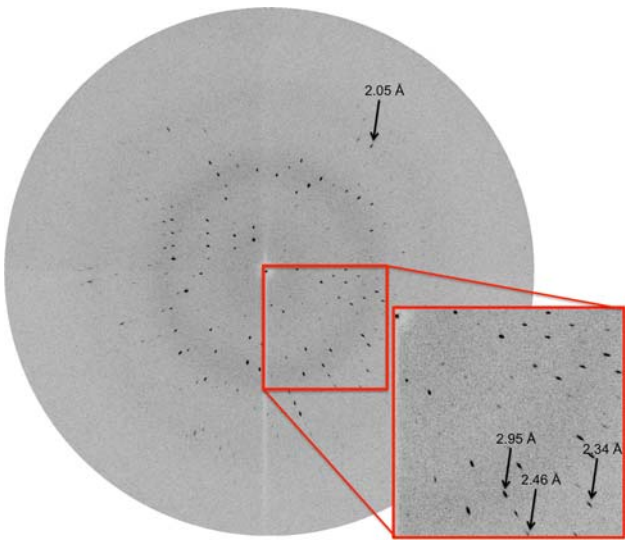

Fig. 5 Diffraction image of MytuGCSPH from the compact light source. The lower right part is magnified. The enhanced radial portion of the mosaic spread is caused by the 1.4% spread of wavelengths used for data collection

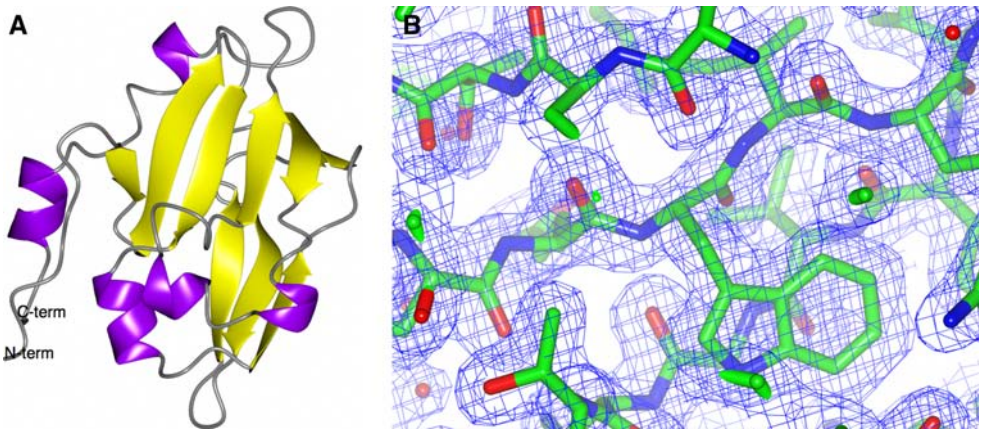

Fig. 6 Structure and electron density of MytuGCSPH. **a** Overview of the structure for MytuGCSPH in ribbon representation. **b** Electron density from the MytuGCSPH CLS data set at 2.0 Å resolution centered around Trp 14

sister structure was solved using data collected from a homelab source. The electron density for both data sets is very clear: an example for the CLS data set is shown in Fig. 6b. Residues from the N-term including two residues from the His_6 purification tag could be modeled without any gaps throughout the C-terminus of the construct. MytuGCSPH has a very compact fold in which two central anti-parallel β-sheets are flanked by a few α-helices (Fig. 6a). The fold is member of the biotinyl-lipoyl-domain superfamily. Only few protein structures with significant sequence homology are known, all of them are members of the Glycine cleavage system. A search for structural homologues using SSM [12] yields the same set of structures. The two structures determined with data from the CLS and a conventional rotating anode are virtually identical: all Ca atoms superimpose with a RMSD of 0.23 Å. The lysine arm (Lys65) which is lipoate-bound in the protein extracted from pea leaves, is ordered however, unmodified in both structures of MutuGCSPH. In fact, a modification of this residue might disturb the crystallographic packing of this crystal form.

Although the focused X-ray flux for this experiment was significantly lower than the design specifications of the CLS, its stable operation and inherently low radiation background provided a high signal-to-noise ratio for the diffraction data. Likewise, the 1.4% energy bandwidth of the X-ray beam from the multilayer optics did not substantially affect the accuracy of the data reduction.

Summary and future plans

Herein, we have described the first protein crystal structure determination using X-ray diffraction data collected using inverse Compton X-ray light from the CLS. In the spring of 2010, we anticipate to commission the new Beta version of the CLS for ATCG3D, which will include several improvements to performance. Ongoing near term technical development of the CLS is targeted to achieve another factor of 100 in native intensity. The X-ray optics used for the crystallography experiments described herein were extremely inefficient and will be replaced with new optics currently under development to improve the flux on sample by another factor of ten. This will increase the flux on the crystal by a factor of 1,000 in the near term, and permit screening with several second exposures and MAD experiments with exposures about 10 times longer.

Together, these flux enhancements will make the CLS a useful tool for protein crystallography and other synchrotron applications. In particular, the crystallography community has shown interest in the CLS for rapid turnaround of results from crystallization trials, to X-ray screening and data collection. The synchrotron imaging community has been hoping for a CLS-like source for more than 20 years. Ultimately, we believe that the CLS can fill the role of a local synchrotron X-ray source, expanding the breadth and impact of X-ray science on biology and human health by opening avenues for exploration that might not be considered or possible at a large facility.

Acknowledgments We gratefully acknowledge NIH funding for the developments included in this paper. The prototype CLS was funded by the NIGMS under grant R44-GM66511. The X-ray station and optics development was funded by the NIGMS and NCRR under grant R44-GM074437. The development of the Beta CLS was funded by the NIGMS and NCRR which have co-sponsored the Accelerated Technologies Center for Gene to 3D Structure (www.atcg3d.org), a PSI-2 Specialized Center, under Grant U54 GM074961. The production of MytuGCSPH protein and crystals was funded by NIAID under Federal Contract No. HHSN272200700057C which supports the Seattle Structural Genomics Center for Infectious Disease (www.SSGCID.org). Special thanks to Shellie Dieterich, Becky Poplawski, and Jeff Christensen at Emerald BioStructures for their support in molecular biology and protein crystallization. The authors also wish to thank the organizing committee of the PSI organized Enabling Technologies Meetings held each spring on the NIH campus. LTI collaborators wish to thank the staff at LTI, Michael Blum, Jens Als-Nielsen, Anette Jensen and Bill Weis.

References

1. Aslanidis C, De Jong PJ (1990) Ligation-independent cloning of PCR products (LIC-PCR). Nucleic Acid Res 18:6069–6074
2. Bech M (2009) X-ray imaging with a grating interferometer. Ph.D. Dissertation. University of Copenhagen, Denmark
3. Bech M, Bunk O, David C, Ruth R, Rifkin J, Loewen R, Feidenhans'l R, Pfeiffer F (2009) Hard X-ray phase-contrast imaging with the compact light source based on inverse compton X-rays. J Synchrotron Rad 16:43–47
4. Cohen-Addad C, Faure M, Neuberger M, Ober R, Sieker L, Bourguignon J, Macherel D, Douce R (1997) Structural studies of the glycine decarboxylation complex from pea leaf mitrochondria. Biochimie 79:637–644
5. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. Acta Crystallogr Sect D Biol Crystallogr 60:2126–2132
6. Faure M, Bourguignon J, Neuburger M, Macherel D, Sieker L, Ober R, Kahn R, Cohen-Addad C, Douce R (2000) Interaction between the lipoamide-containing H-protein and the lipoamide dehydrogenase (L-protein) of the glycine decarboxylase multienzyme system. Eur J Biochem 267:2890–2898
7. Hendrickson WA, Smith JL, Sheriff S (1985) Direct phase determination based on anomalous scattering. Methods Enzymol 115:41–55
8. Huang Z, Ruth RD (1998) Laser-electron storage ring. Phys Rev Lett 80:976–979
9. Jensen A (2007) Optical elements for hard X-ray radiation. Ph.D. Dissertation. University of Copenhagen, Denmark
10. Kabsch W (1988) Automatic indexing of rotation diffraction patterns. J Appl Cryst 21:67–72
11. Kikuchi G, Motokawa Y, Toshida T, Hiraga K (2008) Glycine cleavage system: reaction mechanism, pysiological significance, and hyperglycinemia. Proc Jpn Acad Ser B 84:246–262
12. Krissinel E, Henrick K (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr Sect D Biol Crystallogr 60:2256–2658
13. Loewen R (2003) A compact light source: design and technical feasibility study of a laser electron storage ring x-ray source. Ph.D. Dissertation. SLAC/Stanford University, SLAC-R-632
14. Macherel DBJ, Forest E, Fuare M, Cohen-Addad C, Douce R (1996) Expression, lypoylation and structure determination of recombinant pea H-protein in *Eschericia coli*. Eur J Biochem 236:27–33
15. Mccoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ (2007) Phaser crystallographic software. J Appl Cryst 40:658–674
16. Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecuar structures by the maximum-likelihood method. Acta Crystallogr Sect D Biol Crystallogr 53:240–255
17. Nakai T, Ishijima J, Masui R, Kuramitsu S, Kamiya N (2003) Structure of thermus thermophilus HB8 H-protein of the glycine-cleavage system, resolved by a six-dimensional molecular-replacement method. Acta Crystallogr Sect D Biol Crystallogr 59:1610–1618
18. Newman J, Egan D, Walter TS, Meged R, Berry I, Ben Jelloul M, Stuart DI, Perrakis A (2005) Towards rationalization of crystallization screening for small- to medium-sized academic larboartories: the PACT/JCSG + strategy. Acta Crystallogr Sect D Biol Crystallogr 61:1426–1431
19. Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol 276: 307–326
20. Pares S, Cohen-Addad C, Sieker L, Neuburger M, Douce R (1994) X-ray structure determination at 2.6AA resolution of a lipoate-containing protein: The H-protein of the glycine decarboxylase complex from pea leaves. Proc Natl Acad Sci USA 91:4850–4853
21. Pfeiffer F, Bech M, Bunk O, Kraft P, Eikenberry EF, Brönnimann C, Grünzweig C, David C (2008) Hard-X-ray dark-field imaging using a grating interferometer. Nat Mater 7:134–137
22. Studier FW (2005) Protein production by auto-induction in high density shaking cultures. Protein Expr Purif 41:207–234