NE-CAT Communications

A Biannual Newsletter of the Northeastern Collaborative Access Team Summer 2022

Message from the Director

Steve Ealick

NE-CAT applied for a renewal to our P30 grant in October 2021. Outside of our member institutions, this grant provides all the funding for NE-CAT. In March, our renewal application was peer-reviewed by a Special Emphasis Panel and we scored very well. We received our summary statement on Friday, April 29, which has all the critiques and, most importantly to our continuation, the recommended budget from the review panel. The review panel supported our proposed budget and with our good score we are hopeful that NE-CAT will be funded for the next five years. Our P30 application has now gone to to the NIGMS advisory council where we hope all the reviewers recommendations will be upheld. By our next newsletter, we hope to have a positive conclusion to report on our funding situation. For the latest developments at NE-CAT, be sure to visit our website at: https://necat.chem.cornell.edu

Beamline Developments

APS-U

As of the writing of this newsletter, the APS is still scheduled to go dark for the APS Storage Ring Installation period on April 17, 2023. The APS has been moving forward with component delivery despite supply chain issues and the estimate remains a return to user operations in Spring 2024.

During the estimated year the APS will be down and unavailable for user operations, NE-CAT plans to offer full support from our highly trained crystallography staff for structure solution or advising users on how to collect data at other beamlines. We will also be making our computational resources, such as crystallography programs and our compute cluster, available to our users for remote and local use.

NE-CAT is planning to upgrade our beamline during the long down. As part of our upgrades during the APS-U, NE-CAT has purchased two new microdiffractometers (MD3) which have already arrived. NE-CAT will be obtaining new white beam position monitors for both beamlines, more compact NSLS-II-designed attenuators, and new cryocoolers for both monochromators. Though items have been purchased in tandem for both beamlines, 24-ID-C will be upgraded during the APS-U while 24-ID-E re-commissions with the existing beamline structure intact in order to be immediately available to users and will have its new components installed in the year following the APS-U. The second installation and commissioning will be faster as we will have already done it once on 24-ID-C.

We are also planning to build two new sample automounters (robots) to accommodate the new orientation of the goniometer on the MD3. This will require purchasing new parts and writing new software. One of the new parts in a larger capacity container which has the ability to hold 30 pucks (Fig. 1).

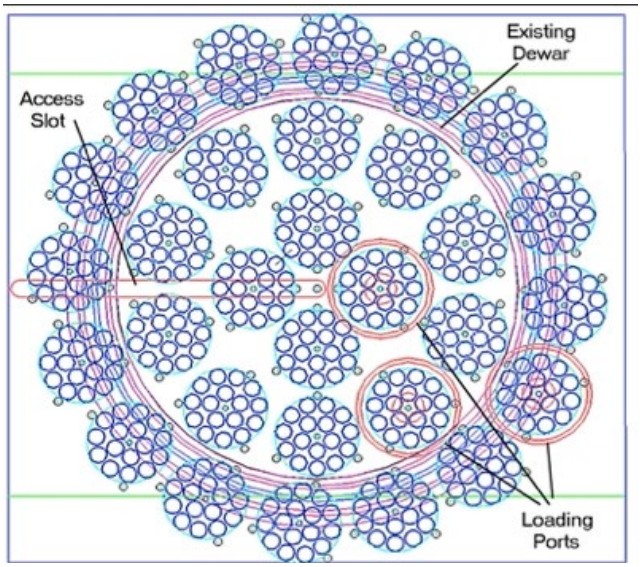

Fig. 1 CAD drawing of sample locations for the proposed 30-puck high capacity automounter.

Lustre

NE-CAT currently uses GPFS as our primary file system for data storage. GPFS is developed by IBM and we have used it since before 2008. As a result, our distribution is quite old and in order to obtain upgrades, we must pay a yearly fee to IBM. Several years ago, NE-CAT began exploring moving to Lustre, a parallel distributed file system, which is open-source, community-supported, and free, as it is available under the GNU General Public License. Currently, Lustre is the most commonly found file system on clusters. NE-CAT began transitioning to Lustre in 2021 for long term data storage. All files in long term storage will be compressed, making them one-third the size. Currently, we estimate that with 430 Terabytes of disk space NE-CAT has ~3 years of data storage capability. Users should expect availability of their data for a minimum of a year after data collection.

Globus

NE-CAT currently offers data transfer through a custom sync script which requires the user to have access to Python. We maintain our own download server for the process. This data transfer has worked well since the advent of Remote but NE-CAT is seeking ways to divest ourselves of maintaining a download server. One such way is Globus.

Globus is a non-profit data transfer service managed by the University of Chicago for the research community. Globus was developed in partnership with the Argonne National Laboratory with the goal of allowing researchers to focus on research instead of IT issues.

NE-CAT initially began investigating Globus as some of our users have firewalls which make them unable to use the sync script available from Remote. We have been testing Globus data transfers with these select groups for the last couple of runs and are now ready to move to making Globus our primary data transfer mechanism. Instructions for using Globus are on our website. Once all our users have successfully transitioned to Globus then we will stop providing the data via the Remote sync script.

Remote Updates

With the arrival of the MD3 microdiffractometers, Deputy Director Malcolm Capel has been busy coding software to integrate them into the NE-CAT beamlines. The MD3 microdiffractometers come with new software capabilities which we have encoded into the current MD2 alignment routines and made them available in remote. First, some users may have noticed a green

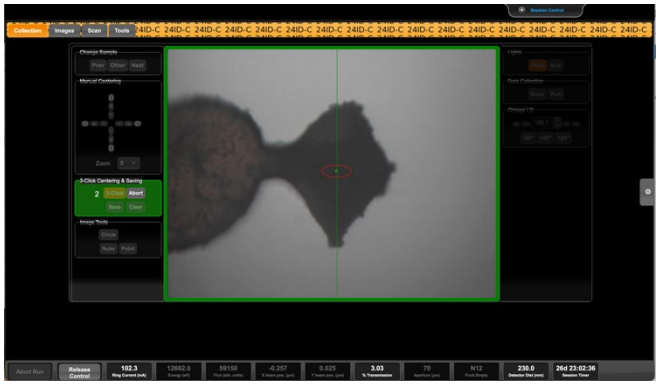

Fig. 2 Remote Data Collection GUI after first center-click to show location of new centering line.

line appearing vertically across the screen after making the first click in three-click centering (Fig. 2). The green line indicates the location, along the horizontal axis, of the first click. This should make it easier for everyone to make their second and third clicks in the same plane and improve centering accuracy. Second, we have also implemented 1-click centering in the Remote interface. To use 1-click centering, the user need only to double-click on the crystal in the centering window and the sample will move to the beam or center of the window. It only works in a 2-D plane, so in order to obtain the third direction/depth, the user will need to rotate 90 degrees to center in the orthogonal direction. 1-click centering is useful for gross centering, such as when the sample is initially mounted. However, 3-click centering is more accurate and faster for precise centering of microcrystals. Finally, if you have an idea you want implemented on the beamline, please let your support scientist or any NE-CAT staff member know.

Flux Measurements

The Flux displayed in the Remote GUI are in arbitrary units which only provide a relative reference to the presence or absence of beam during data collection. The flux density on both beamlines has not been measured with a photodiode since the replacement of the horizontal and vertical focusing mirrors on both beamlines. Due to requests from our users for accurate flux which can be used to accurately calculate radiation damage, NE-CAT has borrowed a calibrated silicon PIN diode with a 10 x 10 mm^2 active area from the APS Detector Pool. Irradiating the diode generates electron-hole pairs in the silicon which causes an electric current to flow across the junction. The PIN diode can be connected to a multimeter and the amperage can be measured to determine the flux.

In March, initial flux measurements were taken on 24-ID-C at 12662 eV and 6500eV using the borrowed PIN diode to determine responsivity of the diode at both energies and the linear range of the diode at different

transmission and aperture sizes (Fig. 3). As the MD2 apertures operate as masks, the total flux available is dependent upon the size of the aperture used. Initial

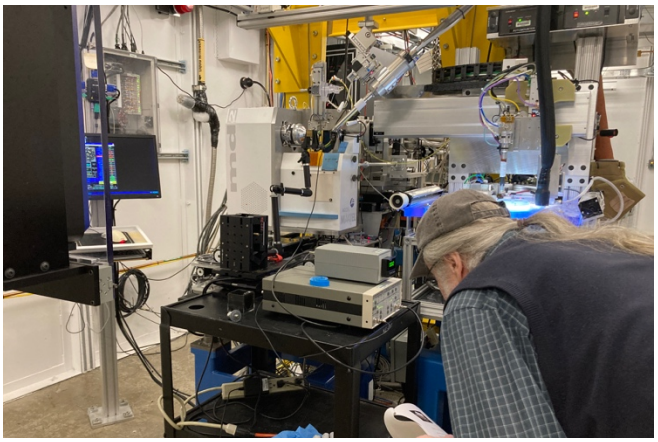

Fig. 3 Dr. Malcolm Capel checking the PIN diode setup prior to taking flux measurements on 24-ID-C.

measurements at 12662 eV indicates a flux of 3.0 x 10^{12} photons/sec at full beam using the 70 micron aperture. Final flux measurements will be taken on 24-ID-C and 24-ID-E at the beginning of run 2022-2.

Crystallography Class

As X-ray crystallography becomes more automated and the plethora of programs available increases, the new user to the field can be overwhelmed by the sheer number of choices available for any task from data processing, data analysis, structure solution and refinement. In order to make crystallography more accessible, NE-CAT has begun offering a virtual course in the practical aspects of crystallography, guiding users at all levels: undergraduate student, graduate student, postdoc or new investigator. The course covers the programs (as used at NE-CAT) needed to take a crystal from an initial dataset to a final publication-ready structure. This course is currently being offered to a limited number of our users as we test our lessons, pacing, and add/remove programs from the curriculum.

If you are interested in participating in the course or have a student in your lab who could benefit from the course, NE-CAT is offering the course twice a year, once during run 1 and once during run 3. Under-represented minorities are given first preference for participation. Below is a sample course syllabus:

1. Data collection using the NE-CAT Remote GUI, including collecting data suitable for SAD structure solution.
2. Data integration, processing and scaling using XDS
3. Data processing and scaling using CCP4 programs: POINTLESS, AIMLESS, CTRUNCATE, FREERFLAG
4. Data analysis: Matthews Coefficient, Pseudotranslation and Twinning analysis using XTRIAGE, Self-Rotation Functions using MOLREP
5. Homology modelling using HHPRED
6. Model preparation (for Molecular Replacement) using SCULPTOR
7. Structure solution using PHASER, SHELXC/D/E
8. Data columns in MTZ files
9. Model building in COOT
10. Refinement using PHENIX.REFINE
11. Figure preparation using PYMOL
12. Data Table preparation using PHENIX.TABLE_ONE

The class is currently free. It is offered via Zoom and lasts between five and ten weeks. To signup for a spot, use our Google form: https://forms.gle/DASKYeaoboCvqq4R6

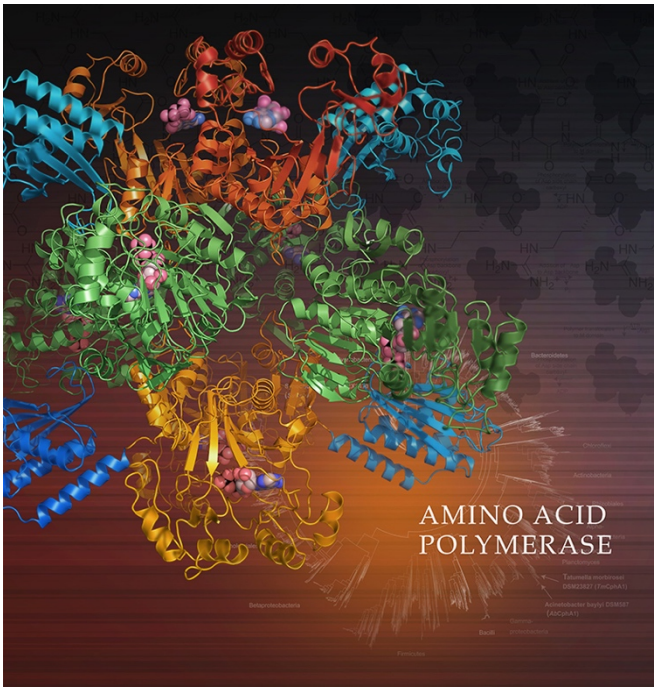

Fig. 4 Cyanophycin synthetase is a beautiful, multi-functional peptide making nanomachine. Image credit: Martin Schmeing, Itai Sharon and Larissa Ulisko

Research Highlights

T. Martin Schmeing, James McGill Professor, Department of Biochemistry & Directeur, Centre de Recherche en Biologie Structurale, McGill University, Montreal, Québec

In the Schmeing lab, we love peptides and the macromolecular machines that make them. Those topics usually bring the ribosome to mind, and although we've done some ribosome research, we are more focused on cellular enzymes that make interesting peptides which don't end up as proteins.

NRPSs: The most common such enzymes are the modular nonribosomal peptide synthetases[1] (NRPSs). NRPSs are megaenzymes found in many bacteria and fungi that produce a vast variety of small peptide molecules with important and diverse biological activity. For example, NRPSs synthesize antifungals, antibacterials, antivirals, antitumours, siderophores, immunosuppressants, industrial agents and green chemicals. The impact of NRPS products on human health has been revolutionary: Penicillin, cyclosporin, cephalosporin and daptomycin have saved tens of millions of lives worldwide. NRPSs are true macromolecular machines, huge proteins with multiple active sites, moving parts and a coordinated synthetic process. They consist of a series of multidomain modules of ~1100 residues, with each module responsible for adding one specific amino acid to the growing peptide molecule, in assembly line fashion.

The research into NRPSs that we have accomplished with contribution from NE-CAT facilities and staff have shed substantial light into the structures and functions of NRPS. Our structures of part of the NRPS that makes the topical antibiotic linear gramicidin (LgrA) provided the most complete view of an NRPS in action[2]. They show massive domain movements and exquisite intra- and inter-domain interactions that the NRPS requires to accomplish biosynthesis. Other studies address the 3D arrangement of full NRPSs, which has been debated for decades: Our structure of a cross-module NRPS construct showed no interaction between core domains of adjacent modules, and a near continuum of module-module conformations[3]. Our structures of dimodular LgrA NRPS reveal massive conformation changes (>200 Å) and prove NRPSs are incredibly dynamic, with no regular, higher-order structure[4].

We are also fascinated with how NRPSs accomplish co-synthetic tailoring by use of embedded tailoring domains and accessory tailoring proteins. Tailoring domains in NRPSs allow products to access far more chemical space and are commonly found in these synthetases. For example, cyclosporin synthetase contains methyltransferase domains; daptomycin (Cubicin) synthetase, epimerization domains; bactitracin (BACiiM) synthetase, a heterocyclization domain. These domains enable key functionalities of the product by providing protease resistance, enabling novel interactions, improving affinity or allowing product to assume its active conformation. Our studies with NRPSs with include formylation domains and reductase domains visualized how a horizontal gene transfer produced these domains, and how they interact productively with the core NRPS domains[2,4-7]. Our studies of a NRPS:oxidase complex allows us to understand how in trans, co-synthetic modification occurs[8].

Biosynthesis and degradation of a non-NRPS peptide: Our love of peptides extends past nonribosomal peptides, and includes an unusual molecule called cyanophycin, a "non-nonribosomal peptide". Cyanophycin is a green biopolymer composed produced widely in the bacterial kingdom. It was discovered over 130 years ago as large, insoluble granules within cyanobacterial cells, and was later understood to be very long chains of poly-L-Asp residues with L-Arg residues attached to each Asp side chains through isopeptide bonds ((β-Asp-Arg)$_{\sim 80\text{-}400}$). Cyanophycin serves as an important nitrogen reservoir

Fig. 5 Members of the Schmeing lab.

Back row, L to R: Mike Tarry, Angelos Pistofidis, Elias Kalthoff, Martin Schmeing

Front row, L to R: Max Eivaskhani, Sarah Alvey, Camille Fortinez, Itai Sharon

for cells and has promising green industrial applications. However, efficient production remains an obstacle in realizing the polymer's commercial potential. Cyanophycin is made by CphA1, which catalyzes the ATP-dependent polymerization of Asp and Arg, or by CphA2, which polymerizes β-Asp-Arg dipeptides. Cyanophycin is degraded in two steps: CphB hydrolyzes cyanophycin into dipeptides, and isoaspartyl dipeptidases cleave dipeptides into Asp and Arg.

With important contributions from NE-CAT, we have characterized the cyanophycin complete metabolism. We determined the first structures of CphA1, and show how nature combined three unrelated domains into a Swiss Army knife enzyme[9]: One domain adds Asp, another adds Arg, and the third both holds on to the growing cyanophycin chain to allow processive synthesis (https://tinyurl.com/CyanoSynth). Our crystal structure of CphA2, show how it evolved to contain only the single active site required for its activity and how that site is altered to allow its new dipeptide substrate to bind[10]. We have also captured the cyanophycinase CphB mid-degradation with the use of the unnatural amino acid diaminopropoanic acid, and solved structures of several isoaspartyl dipeptidases, providing structural insight in to cyanophycin catabolism. These studies have together shed much light on cyanophycin metabolism, and we have used insight gained to increase yield of the polymer from a heterologous host, a key goal in the field.

NRPSs have the potential to produce new bio-active molecules, make them an almost limitless source of novel therapeutics. Cyanophycin derivatives could be used as biodegradable plastic alternatives. Our studies further our fundamental understanding of these complex enzymes that make these peptides, and potentiate their engineering to produce new, useful molecules. But mainly we've just figured out some cool stuff about peptide synthesis. We love peptides.

References

1. Reimer, J.M., Haque, A.S., Tarry, M.J. & Schmeing, T.M. Piecing together nonribosomal peptide synthesis. Curr Opin Struct Biol 49, 104-113 (2018).
2. Reimer, J.M., Aloise, M.N., Harrison, P.M. & Schmeing, T.M. Synthetic cycle of the initiation module of a formylating nonribosomal peptide synthetase. Nature 529, 239-42 (2016).
3. Tarry, M.J., Haque, A.S., Bui, K.H. & Schmeing, T.M. X-ray Crystallography and Electron Microscopy of Cross- and Multi-Module Nonribosomal Peptide Synthetase Proteins Reveal a Flexible Architecture. Structure 25, 783-793 e4 (2017).
4. Reimer, J.M. et al. Structures of a dimodular nonribosomal peptide synthetase reveal conformational flexibility. Science 366(2019).
5. Reimer, J.M. et al. Structural Insight into a Novel Formyltransferase and Evolution to a Nonribosomal Peptide Synthetase Tailoring Domain. ACS Chem Biol 13, 3161-3172 (2018).
6. Alonzo, D.A., Chiche-Lapierre, C., Tarry, M.J., Wang, J. & Schmeing, T.M. Structural basis of keto acid utilization in nonribosomal depsipeptide synthesis. Nat Chem Biol 16, 493-496 (2020).
7. Alonzo, D.A. & Schmeing, T.M. The ribosome makes sweeping arrests. Nat Chem Biol (2016).
8. Fortinez, C.M. et al. Structures and function of a tailoring oxidase in complex with a nonribosomal peptide synthetase module. Nat Commun 13, 548 (2022).
9. Sharon, I. et al. Structures and function of the amino acid polymerase cyanophycin synthetase. Nat Chem Biol 17, 1101–1110 (2021).
10. Sharon, I., Grogg, M., Hilvert, D. & Schmeing, T.M. Structure and function of the β-Asp-Arg polymerase cyanophycin synthetase 2. ACS Chem Biol 17, 670-679 (2022).

Staff Activities

Talks

Schuermann, J. "Data Flow at NE-CAT," 2022 Workshop on High Data Rate Macromolecular Crystallography (HDRMX), Brookhaven National Laboratory, Virtual, April 27-28, 2022.

Publications

Chen, B., Liu, Z., **Perry, K.**, and Jin, R. (2022) Structure of the glucosyltransferase domain of TcdA in complex with RhoA provides insights into substrate recognition, *Sci Rep 12*, 9028. PMID: 35637242.

Arrigoni, C., Lolicato, M., Shaya, D., Rohaim, A., Findeisen, F., Fong, L. K., Colleran, C. M., Dominik, P., Kim, S. S., **Schuermann, J. P.**, DeGrado, W. F., Grabe, M., Kossiakoff, A. A., and Minor, D. L., Jr. (2022) Quaternary structure independent folding of voltage-gated ion channel pore domain subunits, *Nat Struct Mol Biol 29*, 537-548. PMID: 35655098.

Martyn, G. D., Veggiani, G., Kusebauch, U., Morrone, S. R., Yates, B. P., Singer, A. U., Tong, J., Manczyk, N., Gish, G., Sun, Z., **Kurinov, I.**, Sicheri, F., Moran, M. F., Moritz, R. L., and Sidhu, S. S. (2022) Engineered SH2 Domains for Targeted Phosphoproteomics, *ACS Chem Biol 17*, 1472-1484. PMID: 35613471.

Veggiani, G., Yates, B. P., Martyn, G. D., Manczyk, N., Singer, A. U., **Kurinov, I.**, Sicheri, F., and Sidhu, S. S. (2022) Panel of Engineered Ubiquitin Variants Targeting the Family of Human Ubiquitin Interacting Motifs, *ACS Chem Biol 17*, 941-956. PMID: 35385646.

Chen, B., Basak, S., Chen, P., Zhang, C., **Perry, K.,** Tian, S., Yu, C., Dong, M., Huang, L., Bowen, M. E., and Jin, R. (2022) Structure and conformational dynamics of Clostridioides difficile toxin A, *Life Sci Alliance 5*. PMID: 35292538.

Moeller, N. H., Shi, K., Demir, O., Belica, C., **Banerjee, S.**, Yin, L., Durfee, C., Amaro, R. E., and Aihara, H. (2022) Structure and dynamics of SARS-CoV-2 proofreading exoribonuclease ExoN, *Proc Natl Acad Sci U S A 119*. PMID: 35165203.

Lam, K. H., Tremblay, J. M., **Perry, K.,** Ichtchenko, K., Shoemaker, C. B., and Jin, R. (2022) Probing the structure and function of the protease domain of botulinum neurotoxins using single-domain antibodies, *PLoS Pathog 18*, e1010169. PMID: 34990480. **PMC8769338**.

Ubah, O. C., Lake, E. W., Gunaratne, G. S., Gallant, J. P., Fernie, M., Robertson, A. J., Marchant, J. S., Bold, T. D., Langlois, R. A., Matchett, W. E., Thiede, J. M., Shi, K., Yin, L., Moeller, N. H., **Banerjee, S.,** Ferguson, L., Kovaleva, M., Porter, A. J., Aihara, H., LeBeau, A. M., and Barelle, C. J. (2021) Mechanisms of SARS-CoV-2 neutralization by shark variable new antigen receptors elucidated through X-ray crystallography, *Nat Commun 12*, 7325. PMID: 34916516. **PMC8677774**.

Acknowledgements

NE-CAT is supported by a grant from the National Institute of General Medical Sciences (P30 GM124165) and contributions from the following NE-CAT institutional members: Columbia University, Cornell University, Harvard University, Massachusetts Institute of Technology, Memorial Sloan-Kettering Cancer Center, Rockefeller University, and Yale University.

When publishing work resulting from data collected at NE-CAT, we ask our users to acknowledge us, by mentioning our grant number, in your funding or acknowledgements section. For suggested text and a complete list of grants, see our Acknowledgement Request on our website:
https://necat.chem.cornell.edu/acknowledgement