Cryo-EM visualization of key factors associated with genome regulation and integrity

Abstract

The aim of these studies is to establish a structural understanding for the molecular forces driving genomic organization and regulation. The Kellogg lab uses cryo-EM as their main tool to reveal the mechanistic basis of important macromolecular complexes that regulate access to and define genome sequence. To this end we are including in this proposal two distinct core aims which require access to a high-end electron microscope. Building upon our previous work on novel transposases[1], we are interested in continuing our mechanistic studies focused on newly discovered Cascade-Tn7 systems that cooperatively utilize CRISPR and transposon genes to carry out site-directed integration. The direct outcome of these structural studies would result in a better understanding regarding how these evolutionarily unique systems work together (despite containing distinct functional modules). Our second general aim is to understand the native spectrum of transcriptional states using cryo-EM. We are developing a purification pipeline that would isolate endogenously tagged human RNA pol II. These purified complexes will represent a mixture of compositional (and transcriptional) states and will give us a unique window into the mechanistic details underlying *in vivo* transcriptional regulation.

1. Mechanistic studies of Crispr-Tn7 elements using cryo-EM

Scientific Impact Currently available genome-editing tools (such as Cas9[2] and related systems[3]), are capable of sequence-directed cleavage, but introduction of new sequences relies on DNA repair and homologous recombination[4], which are extremely difficult to control. Cascade-Tn7 systems have recently shown the potential for RNA-guided DNA insertion[5, 6], but it's molecular mechanisms are poorly understood[7, 8]. Thus, characterizing the mechanism of Cascade-Tn7 systems would not only facilitate its application to genetic engineering, but also it would give insight into the evolution of Tn7 transposons and the process of Cascade domestication.

Background and Significance

Recent bioinformatic surveys have revealed that, throughout evolution, Type I-F CRISPR genes have been co-opted and subsequently co-evolved with bacterial transposons, called Tn7 elements (referred to here as CasTn7) which mediate horizontal gene transfer in bacteria and often result in antibiotic resistance[9]. Recent biochemical reconstitution of a *Vibrio Cholerae* CasTn7 revealed that the reconstituted macromolecular complex is capable of sequence-directed integration[5].

Goals and Objectives

Motivated to explore the differences between the Cascade-like functions of the complex and the functional roles of the Tn7 adaptor protein, TniQ, we have pursued the structural characterization of the assembly pathway for a distinct lineage (previously uncharacterized) of CasTn7. Our preliminary results indicate that the role of TniQ is tightly integrated with Cascade and functions as a control 'switch' that dictates the functionality of Cascade. Our goal is to further reveal the structural and mechanistic role of TniQ using high-resolution cryo-EM.

Safety Issues/Special Technical Requirements: None

Scientific Feasibility

In collaboration with Ailong Ke's lab (Cornell), we have obtained preliminary reconstructions of three different states describing the assembly of CasTn7(Fig 1), all of which refine to between ~4-5 Å resolution. Because the reconstituted complexes contain a mixture of oligomerization states (Table I), we believe the lower resolutions stem from data limitations that can be easily circumvented using high-throughput data acquisition coupled with 3D image sorting.

Technical Feasibility

Given our preliminary data (Table I) we believe that 1-2 days of data acquisition on a Krios for each of these samples will result in 3.5 Å resolution or better. A subset of our samples demonstrates severe preferred orientation (Fig 2) which have been partially circumvented using tilted image-acquisition, resulting in a 4.5 Å cryo-EM reconstruction (Fig 1). However, to achieve near-atomic resolution we may require more sample optimization or access to the chameleon plunge-freezing device to obtain an isotropic high-resolution reconstruction.

2. Structural studies of Cellular Transcription using cryo-EM

Scientific Impact Recent high-resolution structural studies have elucidated the structural details of recombinantly expressed and reconstituted paused[10], activated[11], and elongating[12, 13] transcription complexes. These structures likely represent a small subset of physiologically relevant functional states. Our structural studies on endogenously tagged RNA Pol II will reveal the *in vivo* conformational ensemble of paused RNA pol II, resulting in a better understanding of how transcription is regulated within the cell.

Background and Significance

RNA polymerase II (Pol II) is the machine that transcribes all mRNA genes and many other transcription units upon which transcription regulators act[14]. Pol II also is coupled to transcript processing and linked to control of alternative splicing. Knowing when and where Pol II loses and gains association with other factors is critical in assessing the function of different Pol II associated factors at key regulatory steps in transcription. This is particularly critical in understanding promoter-proximal pausing and release to productive elongation, where major changes in the entourage of Pol II-associated factors takes place under control of specific transcription factors[15].

Goals and Objectives

We (in collaboration with Lis lab, Cornell) are developing a pipeline to purify endogenous Pol II and its associated factors using a highly specific aptamer[16]. Using this, we will isolate human Pol II in selected states of the transcriptional cycle; in addition, we are working to selectively enrich our target on functionalized EM grids (fabricated in-house). Finally, we plan to exploit our expertise as well as recent advances in image-classification to further characterize the variability within our datasets. With the additional help of transcriptional profiling (Lis lab) and mass spectrometry proteomics (Yu lab, Cornell) experiments from our collaborators, this study will result in an integrative understanding of complex transcriptional regulation in human cells.

Safety Issues/Special Technical Requirements: None

Scientific Feasibility (fit as a cryoEM project)

Raw cryo-EM images of *S. pombe* RNA poll II purification using our purification pipeline contains approximately 400 particles/image (Fig 3), suitably dense for high-throughput data collection.

From a one-day imaging session on Cornell's Arctica/K3, we were able to achieve 5 Å resolution (Fig 3). In order to facilitate identification of auxillary subunits in the cryo-EM density maps, we will utilize mass spectrometry (in collaboration with Haiyuan Yu's lab, Cornell) to identify the associated subunits contained in our imaged samples.

Technical Feasibility (ability to be completed within a defined amount of resources/time)

Due to the anticipated increased heterogeneity (potentially both conformational and compositional in nature) we anticipate that regular high-throughput imaging of the most promising purifications combined with biochemical and mass spectrometry analysis will be a key factor in the success of this aim. Our lab is capable of producing at least one large-scale human purification a week, therefore we anticipate a monthly Krios imaging session would be sufficient for our purposes. Given our preliminary results, we anticipate sample optimization and structural analysis can be completed within the timeframe proposed (2 years).

Justification for Projects 1&2

NCCAT Resources Requested

Due to the number of projects we have outlined in this proposal (2 separate projects), we accordingly request a total of 20 Krios days over the duration of the project (2 years). We anticipate these sessions will be evenly spaced (monthly or every other month) and correspond to 1-2 day sessions. One of our samples demonstrates a severe preferred orientation problem (See Aim 1,Fig 2), we therefore request access to the Chameleon plunge freezing device in order to investigate the possibility of circumventing the preferred orientation problem with this particular sample using this state-of-the-art freezing device.

Geographic/Demographics

The Kellogg lab is located in Ithaca, New York. As upstate New York is relatively isolated, we do not have any similar resources closer to us than NCCAT. We will take the Cornell bus (C2C), which is a 5 hour bus ride to NYC in order to carry out our imaging experiments. At Cornell we have an Arctica/K3 cryo-electron microscope which we use for all our cryo-screening and preliminary data collection sessions. As an alternative, we also have a dry shipping container that we would be able to use for shipping samples to and from NCCAT.

BAG justification

Due to the limited microscopes available for biological single-particle cryo-EM at Cornell, we carry out most of our screening and preliminary cryo-EM work exclusively on Cornell's Arctica/K3. As evidenced by the preliminary reconstructions included in this proposal, we are submitting for high-resolution structure determination samples that are likely to refine to high-resolution (3.5 Å or better). As demonstrated by Table I, these samples are likely to be data-limited, therefore would be the ideal specimens for a high-throughput center such as NCCAT.

In order for these projects to be successful, we require regular access to a high-end instrument. In particular, for project 2 integration of cryo-EM data with mass spectrometry and biochemical analysis is necessary in order to fully characterize the purified samples and cannot be fully determined prior to large-scale imaging.

BAG spokesperson and members

Spokesperson: Elizabeth Kellogg

References

1. Ghanim, G. E.; Kellogg, E. H.; Nogales, E.; Rio, D. C., Structure of a P element transposase-DNA complex reveals unusual DNA structures and GTP-DNA contacts. *Nat Struct Mol Biol* **2019**, 26, 1013-1022.
2. Jinek, M.; Chylinski, K.; Fonfara, I.; Hauer, M.; Doudna, J. A.; Charpentier, E., A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **2012**, 337, 816-21.
3. Knott, G. J.; Doudna, J. A., CRISPR-Cas guides the future of genetic engineering. *Science* **2018**, 361, 866-869.
4. Richardson, C. D.; Ray, G. J.; DeWitt, M. A.; Curie, G. L.; Corn, J. E., Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat Biotechnol* **2016**, 34, 339-44.
5. Klompe, S. E.; Vo, P. L. H.; Halpin-Healy, T. S.; Sternberg, S. H., Transposon-encoded CRISPR-Cas systems direct RNA-guided DNA integration. *Nature* **2019**, 571, 219-225.
6. Peters, J. E.; Makarova, K. S.; Shmakov, S.; Koonin, E. V., Recruitment of CRISPR-Cas systems by Tn7-like transposons. *Proc Natl Acad Sci U S A* **2017**, 114, E7358-E7366.
7. Halpin-Healy, T. S.; Klompe, S. E.; Sternberg, S. H.; Fernandez, I. S., Structural basis of DNA targeting by a transposon-encoded CRISPR-Cas system. *Nature* **2020**, 577, 271-274.
8. Jia, N.; Xie, W.; de la Cruz, M. J.; Eng, E. T.; Patel, D. J., Structure-function insights into the initial step of DNA integration by a CRISPR-Cas-Transposon complex. *Cell Res* **2020**, 30, 182-184.
9. Peters, J. E., Targeted transposition with Tn7 elements: safe sites, mobile plasmids, CRISPR/Cas and beyond. *Mol Microbiol* **2019**, 112, 1635-1644.
10. Vos, S. M.; Farnung, L.; Urlaub, H.; Cramer, P., Structure of paused transcription complex Pol II-DSIF-NELF. *Nature* **2018**, 560, 601-606.
11. Vos, S. M.; Farnung, L.; Boehning, M.; Wigge, C.; Linden, A.; Urlaub, H.; Cramer, P., Structure of activated transcription complex Pol II-DSIF-PAF-SPT6. *Nature* **2018**, 560, 607-612.
12. Farnung, L.; Vos, S. M.; Cramer, P., Structure of transcribing RNA polymerase II-nucleosome complex. *Nat Commun* **2018**, 9, 5432.
13. Baluapuri, A.; Hofstetter, J.; Dudvarski Stankovic, N.; Endres, T.; Bhandare, P.; Vos, S. M.; Adhikari, B.; Schwarz, J. D.; Narain, A.; Vogt, M.; Wang, S. Y.; Duster, R.; Jung, L. A.; Vanselow, J. T.; Wiegering, A.; Geyer, M.; Maric, H. M.; Gallant, P.; Walz, S.; Schlosser, A.; Cramer, P.; Eilers, M.; Wolf, E., MYC Recruits SPT5 to RNA Polymerase II to Promote Processive Transcription Elongation. *Mol Cell* **2019**, 74, 674-687 e11.
14. Chen, F. X.; Smith, E. R.; Shilatifard, A., Born to run: control of transcription elongation by RNA polymerase II. *Nat Rev Mol Cell Biol* **2018**, 19, 464-478.
15. Jonkers, I.; Kwak, H.; Lis, J. T., Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. *Elife* **2014**, 3, e02407.
16. Shui, B.; Ozer, A.; Zipfel, W.; Sahu, N.; Singh, A.; Lis, J. T.; Shi, H.; Kotlikoff, M. I., RNA aptamers that functionally interact with green fluorescent protein and its derivatives. *Nucleic Acids Res* **2012**, 40, e39.