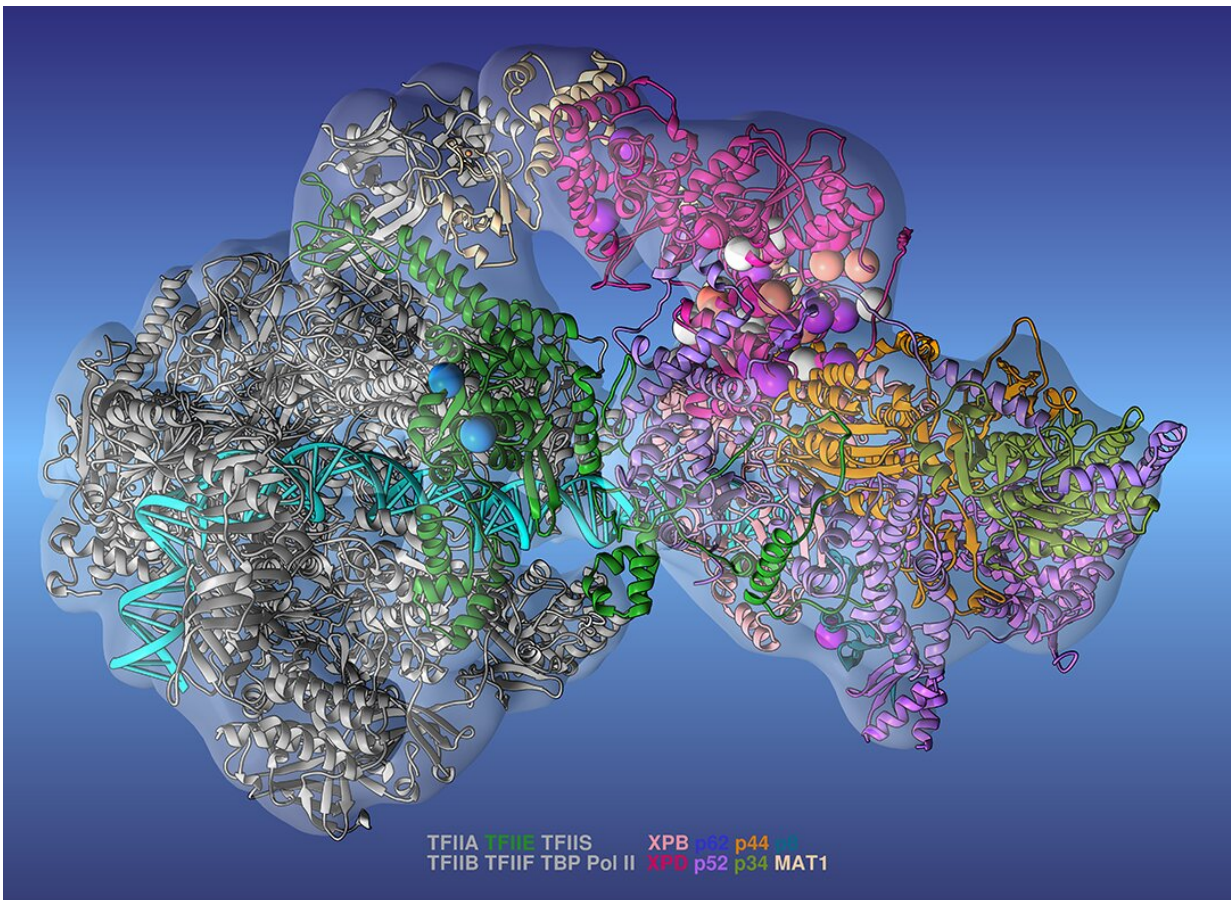


Researchers create most complete model of complex protein machinery

May 21 2019, by Elizabeth C Rosenthal



The researchers used the new model to accurately identify clusters of gene mutations (spheres), which helped them study the emergence of various genetic diseases. Credit: Ivaylo Ivanov, Georgia State University

Environmental conditions, lifestyle choices, chemical exposure, and foodborne and airborne pathogens are among the external factors that can cause disease. In contrast, internal genetic factors can be responsible for the onset and progression of diseases ranging from degenerative neurological disorders to some cancers.

A team led by Ivaylo Ivanov of Georgia State University used the 200-petaflop IBM AC922 Summit system, the world's smartest and most powerful supercomputer, to develop an integrative model of the transcription preinitiation complex (PIC), a complex of proteins vital to [gene expression](#). Results from this work are published in *Nature Structural & Molecular Biology*.

Gene expression involves the conversion of genetic information that originates in DNA to produce functional molecules such as proteins—the building blocks of all living organisms—through steps known as transcription and translation. Because gene mutations can interfere with gene expression and cause disease, biomedical scientists are particularly interested in making sense of the connection between a patient's unique genetic makeup, or genotype, and the external manifestation of a disease, or phenotype.

A better understanding of the complex relationship between a genotype and a phenotype could reveal how mutations cause genetic diseases and thus inform the development of more effective treatments. Researchers do not yet fully comprehend how inherited mutations affect protein function.

"Like a broken gear in a machine, mutational changes break down the function of the defective protein, a process that involves alterations in both structure and dynamics," Ivanov said. "This confluence of factors presents a challenge for conventional structural biology methods."

Throughout the complex and highly regulated process of gene transcription, enzymes called Pol I, Pol II, and Pol III—referred to collectively as RNA polymerases—play a significant role. Pol II helps mediate protein synthesis, the process of transforming genetic information into proteins.

During initiation—the first step of transcription—Pol II and a host of general transcription factors (GTFs) assemble in a region of DNA called a promoter to form the PIC. Opening the promoter depends on transcription factor II human (TFIIH), a GTF consisting of multiple protein chains, which has the ability to unwind the double helix strands of DNA to initiate transcription. TFIIH also contributes to DNA repair.

Because the biochemical pathways responsible for gene expression and repair are intertwined, understanding the [molecular mechanism](#) behind this process is crucial to advancing biomedical applications. For example, the presence of mutations in three subunits of TFIIH directly leads to severe genetic diseases, including autoimmune and neurological disorders.

Previous attempts to characterize the PIC have been limited by incomplete models. The most complete model of the PIC to date, the team's new version provides superior insights into the structural organizations of these proteins, which transcribe genes and repair DNA.

To develop their PIC model, the researchers combined data from cryo-electron microscopy (CryoEM)—a structural biology method that uses an electron beam to study cryogenically frozen protein samples—and large-scale molecular dynamics simulations on Summit using the Nanoscale Molecular Dynamics (NAMD) code. Summit is located at the Oak Ridge Leadership Computing Facility (OLCF), a US Department of Energy (DOE) Office of Science User Facility at DOE's Oak Ridge National Laboratory (ORNL).

"The new model gives us the most complete view of the structure of TFIID, which helps us understand the dynamics of these proteins and lets us map the origins of patient-derived mutations, potentially enabling future biochemical experiments focused on understanding the structural mechanisms of TFIID," Ivanov said.

The simulations revealed the hierarchical organization of the PIC and explained how its numerous structural components function to modify DNA. By mapping 36 different patient-derived mutations to the PIC model, the team determined that the mutations tend to cluster at crucial areas of TFIID, including a subunit known as XPD, which prevents the GTF from working properly and leads to disease.

"Computation is absolutely instrumental in providing a link between the structure, which comes from CryoEM data, and the disease phenotype, which is a high-level concept difficult to explain with answers purely based on traditional biochemistry and structural biology," Ivanov said.

From these findings, the team gained detailed insights into three distinct genetic disorders associated with cancer, aging, and developmental defects by unveiling their distinguishing molecular mechanisms.

"If you have a handle on which regions of a [protein](#) are affected, then you can potentially develop therapies for genetic diseases, but without a fundamental understanding of the underlying mechanism, all bets are off," Ivanov said.

This accomplishment provides a foundation for future experimental and computational efforts that could precisely pinpoint the mutations that cause genetic disorders, expound on TFIID's distinct contributions to transcription and DNA repair, and delve deeper into the mechanisms of gene expression.

Although this research would have been feasible on other high-performance computing platforms, access to Summit significantly sped up the team's simulations.

"Running on Summit accelerates our research," Ivanov said. "Instead of spending several months running on another system, we were able to complete our calculations in a matter of days, which saved us a lot of time and effort."

This year, the team will run related calculations on Summit through a 2019 Innovative and Novel Computational Impact on Theory and Experiment (INCITE) program allocation. The researchers have mainly studied Pol II, but they plan to expand their project to investigate the functional dynamics of Pol I and Pol III as well, which could lead to more groundbreaking insights.

"We look forward to moving beyond merely describing the mechanisms of transcription to elucidate their connection to genetic diseases," Ivanov said.

More information: Chunli Yan et al, Transcription preinitiation complex structure and dynamics provide insight into genetic diseases, *Nature Structural & Molecular Biology* (2019). [DOI: 10.1038/s41594-019-0220-3](https://doi.org/10.1038/s41594-019-0220-3)

Provided by Oak Ridge National Laboratory

Citation: Researchers create most complete model of complex protein machinery (2019, May 21) retrieved 13 March 2024 from <https://phys.org/news/2019-05-complex-protein-machinery.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private

study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.