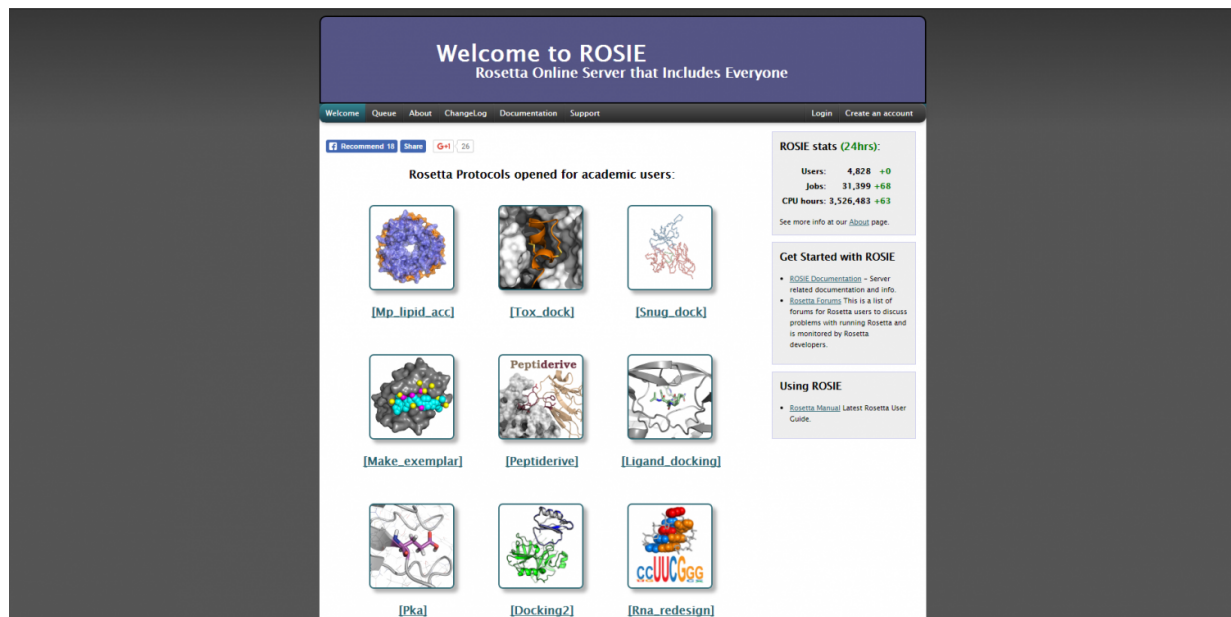


# Rosetta online server that includes everyone

April 25 2017



ROSIE is an easy-to-use web interface (also known as a 'gateway') that provides access to the Rosetta software suite and encapsulates the body of rapidly evolving tools for the 3-D structure prediction and high-resolution design of proteins, nucleic acids, and a growing number of non-natural polymers that were created by members of the RosettaCommons. Credit: Gray Lab

Our bodies are made of biomolecules like proteins, nucleic acids, fats and sugars. These biomolecules are folded into specific 3D structures—predetermined by the DNA and RNA sequences that build them—which allows them to do everything they need to do in our bodies.

Biomolecules are frequently long and can bend in lots of different ways, creating an immense number of possible forms. For scientists trying to understand how a protein works, or how to design a biomolecule that accomplishes a specific action, the task of determining what it might look like in 3D is daunting.

To deal with this problem, scientists have developed computer algorithms that are clever enough to map out biomolecules' 3D forms, or create entirely new ones, based on their DNA or RNA sequence. However, doing so requires powerful supercomputers and specialized software that can take advantage of them.

One of the most widely used such programs is Rosetta. Originally developed as a structure prediction tool more than 17 years ago in the laboratory of David Baker at the University of Washington, Rosetta has been adapted to solve a wide range of common computational macromolecular problems. It has enabled notable scientific advances in computational biology, including protein design, enzyme design, ligand docking, and structure predictions for biological macromolecules and macromolecular complexes.

"The structure prediction problem is to take a sequence and ask, 'What does it look like?'" said Jeffrey Gray, a professor of Chemical and Biomolecular Engineering at Johns Hopkins University and a collaborator on the project.

"The design problem asks 'What sequence would fold into this structure?' That's at the heart of Rosetta, but Rosetta does a lot of other things," Gray said.

Over the years, Rosetta evolved from a single tool, to a collection of tools, to a large collaboration called RosettaCommons, which includes more than 50 government laboratories, institutes, and research centers

(only nonprofits).

## The ROSIE Science Gateway

Most recently, with support from the National Science Foundation (NSF), it has morphed once again into ROSIE: the Rosetta Online Server that Includes Everyone. ROSIE is an easy-to-use web interface (also known as a 'gateway') that provides access to the Rosetta software suite and encapsulates the body of rapidly evolving tools for the 3D [structure prediction](#) and high-resolution design of proteins, [nucleic acids](#), and a growing number of non-natural polymers that were created by members of the [RosettaCommons](#).

"The idea was to take this collaboration of 50 labs and institutions and make a single gateway," Gray said. "Rather than duplicating the work that everyone else was doing we agreed to work together. We decided to use NSF resources for the back end to provide the computational power. Now, it's easy to maintain 18 different web servers."

First described in [PLOS One](#) in May 2013, it continues to add new elements. In January 2017, a team of researchers, including Gray, reported in Nature Protocols on the latest additions to the gateway: antibody modeling and docking tools called RosettaAntibody and SnugDock that can run fully automated via the ROSIE web server or manually, with user control, on a personal computer or cluster.

Currently, the ROSIE gateway serves approximately 5,000 users and has run more than 30,000 jobs.

Some of the calculations enabled by ROSIE require 10 minutes of compute time; others require 200 computer processing hours. With several thousand users, the computing needs quickly add up.

"XSEDE [the Extreme Science and Engineering Discovery Environment] was a natural fit for a shared national resource that allows many different scientists to do science using large compute facilities," Gray said.

Initially funded by a five-year, \$110-million grant from NSF, XSEDE is the most advanced, powerful, and robust collection of integrated advanced digital resources and services in the world. It is a single virtual system that scientists can use to interactively share computing resources, data, and expertise.

The Stampede supercomputer at the Texas Advanced Computing Center (TACC), one of the resources allocated through XSEDE, provides the lion's share of the computing power. Gray had used TACC resources as a graduate student in Texas in the late 1990s, so he knew about TACC and some of the other NSF supercomputing facilities.

"We've been using Stampede and applied for it through XSEDE," Gray said. "We have a Stampede allocation for my lab and we have a separate allocation for ROSIE."

Stampede serves as the back-end computing system for the thousands of researchers who use ROSIE. It has provided roughly two million compute hours for the project since 2013. Though scientists may not be aware that they are using a supercomputer, the project could not be as successful without a massive, on-demand supercomputer humming away in the background.

In Gray's own lab, he is exploring the structure and interactions of membrane proteins, which behave differently than many other types of proteins because they are in a bilayer of fatty lipids. How proteins interact and fold inside the cell membrane is an open question that his lab is trying to solve.

"The other big new thrust in the lab is glycoproteins," Gray said.

"Most of the proteins in your body have sugars attached to them, which makes them glycoproteins. Traditionally, people ignored the glycans, but they are very important to cancer, heart disease, diabetes, aging, and infectious diseases. We're adding carbohydrates into the structure, and modeling their effects on protein folding and binding interactions using the Rosetta software and the Stampede supercomputer."

## **Getting Help from XSEDE's Experts**

Beyond providing raw computing power to the nation's researchers, XSEDE also runs an Extended Collaborative Support Service (ECSS) program, which pairs researchers with cyberinfrastructure experts who have a variety of expertise. ECSS experts, many with advanced degrees in domain areas, are available for collaborations lasting months to a year to help researchers fundamentally advance their use of XSEDE resources.

"There were a couple of places that we needed ECSS's help," Gray said. "One was setting up the ROSIE science gateway. To run a gateway there are many security concerns—you have people logging in from different locations, and the computer cluster is a hacking target. To assuage this concern, the software engineer that developed ROSIE worked with TACC staff to make sure the gateway worked properly. That was very successful."

In addition, Gray and other researchers needed the ability to write their own code in Rosetta beyond simply running canned software. Thus, Gray also worked with ECSS to install the Rosetta Python modules, called PyRosetta, which was created in Gray's lab.

"It's a Python interface to all of the Rosetta tools," Gray said. "It allows

people to make their own customized scripts for tailored modeling."

PyRosetta is installed on Stampede as a module so that a scientist who is more of an expert can log into Stampede, load the module, and have access to all of the Rosetta code and functionality, allowing them to tailor their own scripts for their own particular molecules or designs that they are trying to calculate.

"I'm hugely grateful for NSF, XSEDE and TACC for making these resources available," Gray said. "We spent so many years and had so many students put all of their research effort into making great tools to model and design biomolecules and you want other people to be able to use it. However, biomolecular prediction and design requires tremendous computing time, so having XSEDE there makes it possible for us to share our tools broadly and allow them to have impact across the scientific community."

As ROSIE and the community it supports continues to grow, so do its computing needs.

"There's a huge life sciences community out there that wants to perform structural predictions on their biomolecules, but we can't handle it all with the current demand on Stampede."

For that reason, Gray is eagerly awaiting [Stampede2](#), TACC's newest supercomputer which is due to come online later in 2017, "so we have the capacity to handle the great demand for computing time."

**More information:** Sergey Lyskov et al. Serverification of Molecular Modeling Applications: The Rosetta Online Server That Includes Everyone (ROSIE), *PLoS ONE* (2013). [DOI: 10.1371/journal.pone.0063906](#)

Provided by University of Texas at Austin

Citation: Rosetta online server that includes everyone (2017, April 25) retrieved 27 April 2024 from <https://phys.org/news/2017-04-rosetta-online-server.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.