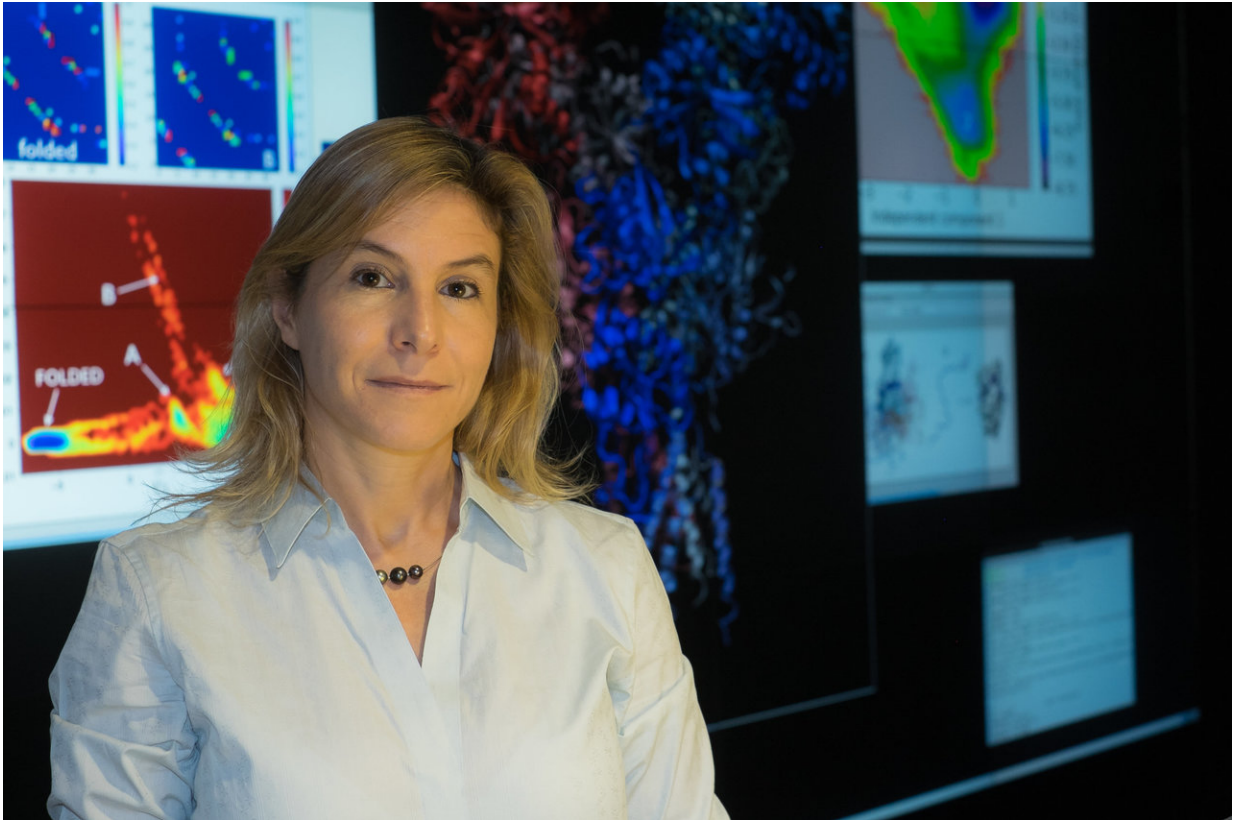


Simpler model gets to the point with proteins

June 6 2018



Cecilia Clementi. Credit: Jeff Fitlow/Rice University

Computational models have come a long way in their ability to simulate the most basic biological processes, such as how proteins fold. A new technique created by Rice University researchers should enable scientists to model larger molecules with greater accuracy than ever.

The Rice lab of computational chemist Cecilia Clementi has developed a molecular modeling framework that can more accurately reproduce experimental results with simple coarse-grained models used to simulate [protein dynamics](#).

The framework, Observable-driven Design of Effective Molecular Models (ODEM), incorporates available experimental data in the definition of a coarse-grained [simulation model](#). For a given coarse-grained model, repeating the simulation with incremental changes in the model parameters improves the algorithm's ability to predict, for instance, how a protein will find its functional form.

The work led by Clementi and Rice graduate student and lead author Justin Chen appears in the American Chemical Society's *Journal of Chemical Theory and Computation*.

"Understanding proteins, especially their dynamics, is essential to understanding life," Clementi said. "There are two complementary ways to do this: either through simulation or experimentation. In an experiment, you measure something that's real, but you're very limited in the quantities you can measure directly. It's like putting together a puzzle with only a very few pieces."

She said simulations allow researchers to look at every aspect of protein dynamics, but models that incorporate the properties of every atom can take supercomputers months or years to compute, even if the proteins themselves fold in seconds in vivo. For faster results, scientists often use coarse-grained models, simplified simulations in which a few effective "beads" represent groups of atoms in a protein.

"In very simple models you have to make strong approximations, and as a consequence, the results may differ from reality," Clementi said. "We combine these two approaches and use the power of simulation in a way

that reproduces the experiments. That way, we get the best of both worlds."

Acquiring initial data is not an issue, Chen said. "There is a wealth of experimental data about proteins already, so it's not hard to find," he said. "It's just a matter of finding a way to model that data in a simulation."

Clementi said the data can come from any one or a combination of sources like Förster resonance energy transfer (FRET), mutagenesis or nuclear magnetic resonance. The computational framework takes advantage of Markov models to combine multiple short protein simulations to obtain the equilibrium distribution of protein configurations that is used in ODEM. "Markov models let us combine and explore different parts of the configurational space of a protein," she said. "It's a clever way to divide and conquer."

The key, according to the researchers, is to include only as much physical detail as necessary to model the process accurately.

"There are models that are very accurate, but they are computationally too expensive," Clementi said. "There's too much information in those models, so you don't know what are the most important physical ingredients.

"In our simplified models, we include only the physical factors we think are important," she said. "If by using ODEM the simulations improve their agreement with experiments, it means that the hypothesis was correct. If they do not, then we know there are ingredients missing."

The researchers found their technique can reveal unanticipated molecular properties. In the process of testing their algorithm, the researchers discovered a new detail about the folding mechanism of

FiP35, a common WW domain protein that is a piece of larger signaling and structural proteins. FiP35, with only 35 amino acids, is well-understood and often used in folding studies.

The ODEM model of FiP35, based on [experimental data](#) from simulated FRET results, revealed several regions where localized frustration forced changes in the folding process. Their analysis showed the interactions are important to the process and likely evolutionarily conserved, but they said the data leading to that conclusion would never have appeared if the simulated FRET data were not used in the coarse-grained model.

"Now we're scaling it up to larger systems, like 400-residue proteins, about 10 times larger than our test [protein](#)," Chen said. "You cannot do full-atom simulations of these large motions and long time scales, but if you do 10 or 11 iterations of a coarse-grained model with ODEM, they take only a few hours. That's a huge reduction of the time it would take a person to see reasonable results."

More information: Justin Chen et al, Learning effective molecular models from experimental observables, *Journal of Chemical Theory and Computation* (2018). [DOI: 10.1021/acs.jctc.8b00187](https://doi.org/10.1021/acs.jctc.8b00187)

Provided by Rice University

Citation: Simpler model gets to the point with proteins (2018, June 6) retrieved 2 May 2024 from <https://phys.org/news/2018-06-simpler-proteins.html>

<p>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.</p>
--