Six steps closer to FRET-driven structural biology

Timothy D Craggs & Achillefs N Kapanidis

A new toolbox for structural biology that combines single-molecule fluorescence and molecular modeling is used to generate high-precision structures of protein complexes.

Atomic-resolution structures are central to understanding the function of biological molecules. Although traditional X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy have been used to generate the bulk of high-resolution structures, they are less suitable for large, multicomponent, transient or dynamic biomolecular complexes. In this issue, Kalinin *et al.*[1] present a comprehensive toolkit for determining biomolecular structure and conformations by combining single-molecule Förster resonance energy transfer (smFRET) with molecular modeling. Using this approach, the authors generated a structure of a human immunodeficiency virus reverse transcriptase–DNA complex that agrees remarkably well with an X-ray structure of the complex. This achievement marks a dramatic improvement in the accuracy of FRET-derived structures, and it confirms smFRET as a quantitative structural biology tool.

Why would conventional structural biology methods need complementary approaches such as smFRET? In addition to the difficulties of applying X-ray crystallography and NMR spectroscopy to large or dynamic structures, both methods require significant amounts of high-purity material, which is often difficult and expensive to obtain. Moreover, crystallography is limited by the need for crystal growth, and many proteins (especially membrane proteins) crystallize with difficulty, if at all. Apart from these limitations, protein conformations in crystals may not reflect their conformations in solution, particularly for flexible modules that are important for protein activity; indeed, these modules are often missing from crystal structures. Structures can also be obtained using cryo-electron microscopy, but resolution is poorer and the fact that samples must be frozen hinders the analysis of dynamics.

On the other hand, computational methods have made strides in describing conformational ensembles of biomolecules, mainly through molecular dynamics (MD) simulations. Guidance of these ensembles using experimentally determined distance restraints offers a general approach to determine three-dimensional structures of proteins and their complexes.

FRET is a well-established interaction between donor and acceptor dyes that serves as a molecular ruler for the 2- to 10-nanometer scale[2]. It can provide useful experimental distance restraints because its dynamic range matches well with the length scale of nucleic acids and proteins. Although structural determination based on ensemble FRET has been reported[3], the method has been limited by the many sources of heterogeneity found in typical samples. Through the measurement of the FRET of individual molecules[4], ensemble averaging is removed, and heterogeneity can be directly observed and either excluded from the analysis or understood in mechanistic terms. However, relating measured interdye distances to structure is complicated by the fact that the dyes populate a range of positions relative to their attachment points. Modeling these distributions is key for establishing FRET as an accurate, quantitative structural tool and has been the focus of many recent efforts.

Timothy D. Craggs and Achillefs N. Kapanidis are in the Biological Physics Research Group, Clarendon Laboratory, Department of Physics, University of Oxford, Oxford, UK.
e-mail: timothy.craggs@physics.ox.ac.uk or kapanidis@physics.ox.ac.uk

Several groups have successfully applied FRET-guided rigid-body docking using restraints derived from experiments on immobilized molecules[5–7]. A complementary probabilistic data analysis approach known as the nano-positioning system has also been developed[8,9]. Kalinin *et al.*[1] extend this work by considering the effect of averaging FRET efficiency over donor-acceptor distance distributions, using iterative docking that considers the impact of the spatial arrangement of structural units on dye distributions, performing rigorous error analysis and model validation by comparing to a known structure, and introducing FRET-guided screening of a large structural ensemble generated by MD simulations. Although each step represents an incremental improvement over existing methods, when taken together, the full toolbox provides a significant advance in the accuracy and validation of FRET for structural biology.

Kalinin *et al.*[1] generate structural models in six steps (**Fig. 1**): biomolecular modeling, dye modeling, smFRET measurement, structure docking or screening, model assessment and model precision determination. First, an initial model is generated from known structures, homology modeling or *ab initio* modeling. The model is then used to design a network of dye positions, which is achieved by considering the volumes accessible to dyes and their flexible linkers, a method similar to that of the nano-positioning system[8]. The authors also correct discrepancies between the distances spanning mean dye positions and those calculated using FRET efficiency. The differences can reach up to 30%, particularly at distances below the Förster radius (a value that defines the middle of the FRET dynamic range for a specific donor-acceptor pair).

Single-molecule FRET measurements are performed using a sophisticated confocal microscope that records donor and acceptor fluorescence as well as donor fluorescence lifetime and anisotropy; the results are then converted to interdye distances. This multiparameter fluorescence-detection approach[10], pioneered by the Seidel group, has excellent temporal resolution that can detect fast dynamics and dye rotations, and because it observes diffusing molecules, it is free from any surface effects; MFD is thus complementary to

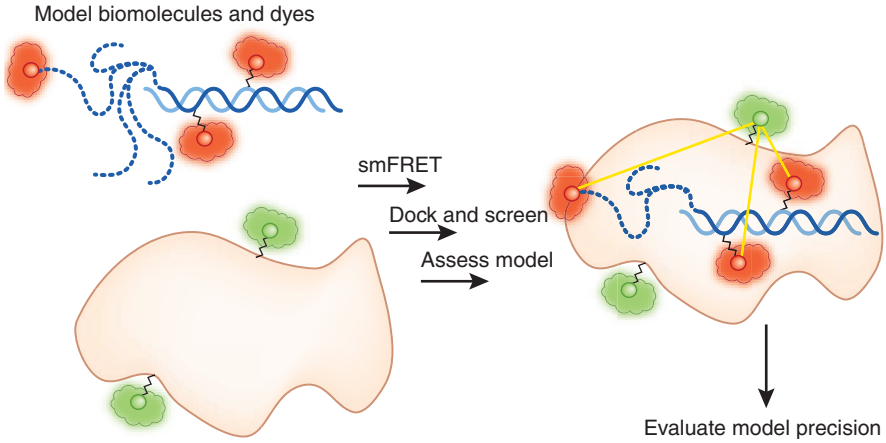

Figure 1 | A workflow combining smFRET and computer modeling can provide accurate biomolecular structures. In this example of a protein-DNA complex, double-stranded DNA (light and dark blue lines) and protein (beige) are initially modeled separately. Multiparameter fluorescence detection of complexed biomolecules conjugated to donor and acceptor fluorophores (green and red circles) then provides smFRET data that are converted to distance restraints (yellow lines). The restraints guide rigid-body docking of the modeled protein and DNA, and they allow the selection of best-fit solutions for unknown features (a flexible single-stranded overhang, dotted blue lines) from among conformations generated by molecular dynamics simulations. By considering dye spatial distributions (green and red clouds) and applying rigorous error analysis, researchers can derive high-precision models of large complexes and flexible molecules.

approaches based on total-internal-reflection fluorescence microscopy.

After generating smFRET data, structural 'candidates' are evaluated by two complementary approaches on the basis of their agreement with the FRET measurements: rigid-body docking of known substructures (wherein measured distances are modeled as springs with strengths that depend on the measurement errors) and screening of a large ensemble of putative structures generated by MD simulations. The models are ranked according to how well they fit the FRET restraints, and clustering is used to judge their uniqueness and determine confidence intervals. Finally, the precision of the best-fit model is determined.

The authors applied this workflow to a complex of human immunodeficiency virus reverse transcriptase with a primer-template DNA. Using eight FRET donor sites on transcriptase and five acceptor sites on DNA, they used a set of 20 smFRET measurements to generate a docked, FRET-restrained model that was in excellent agreement with an X-ray structure of the complex (root mean square deviation was 0.5 Å). The authors then determined the configuration of the flexible single-stranded template overhang (missing from the crystal structure) by screening a library of MD-generated structures against 16 distances between acceptors on the overhang and donors on transcriptase.

Despite the promise of this new FRET-based method, it will be challenging to turn it into a mainstream tool. One limitation is the need for many specific labeling positions on the proteins of interest. Dyes are often conjugated to proteins via surface cysteine residues, which requires that all native surface cysteines be removed by mutation. Many proteins contain cysteines that cannot be removed because they are essential for stability or function. A potential solution uses unnatural amino acids with chemical reactivities orthogonal to that of cysteine, which can be genetically encoded anywhere in the protein sequence[11].

The Kalinin *et al.*[1] method requires complex instrumentation and analysis, although commercial versions of similar microscopes are available. The confocal geometry for single-molecule detection also places an upper limit on the concentration of fluorescent species used (~500 picomolar), confining it to tight binding partners (K_d < 10 nanomolar). Vesicle encapsulation can address this issue by effectively increasing the local concentration of the interacting species.

The smFRET methods presented here and elsewhere[8] provide attractive means for studying the structures of molecules not easily accessible by standard structural biology, such as large proteins and complexes, interconverting populations in dynamic equilibrium and protein regions with increased flexibility. As FRET-based structures accumulate, we expect further standardization of the methodology and easier access to the structural models, perhaps through deposition in the Protein Data Bank. By coupling these methods with advances in smFRET detection in live cells[12], the tantalizing possibility arises of determining biomolecular structure *in vivo* and even capturing conformational states of proteins at work *in situ*.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

1. Kalinin, S. *et al. Nat. Methods* **9**, 1218–1225 (2012).
2. Stryer, L. & Haugland, R.P. *Proc. Natl. Acad. Sci. USA* **58**, 719–726 (1967).
3. Mekler, V. *et al. Cell* **108**, 599–614 (2002).
4. Ha, T. *et al. Proc. Natl. Acad. Sci. USA* **93**, 6264–6268 (1996).
5. Choi, U.B. *et al. Nat. Struct. Mol. Biol.* **17**, 318–324 (2010).
6. Brunger, A.T., Strop, P., Vrljic, M., Chu, S. & Weninger, K.R. *J. Struct. Biol.* **173**, 497–505 (2011).
7. McCann, J.J., Zheng, L., Chiantia, S. & Bowen, M.E. *Structure* **19**, 810–820 (2011).
8. Muschielok, A. *et al. Nat. Methods* **5**, 965–971 (2008).
9. Muschielok, A. & Michaelis, J. *J. Phys. Chem. B* **115**, 11927–11937 (2011).
10. Sisamakis, E., Valeri, A., Kalinin, S., Rothwell, P.J. & Seidel, C.A.M. *Methods Enzymol.* **475**, 455–514 (2010).
11. Liu, C.C. & Schultz, P.G. *Annu. Rev. Biochem.* **79**, 413–444 (2010).
12. Sakon, J.J. & Weninger, K.R. *Nat. Methods* **7**, 203–205 (2010).

GEM: crystal-clear DNA alignment

Gregory G Faust & Ira M Hall

The Genome Multitool (GEM) mapper rapidly and accurately provides all alignments of a read within a user-defined number of mismatches.

Over the past several years, the flood of data produced by second-generation DNA sequencers has motivated the development of a new generation of DNA alignment tools designed for mapping short reads to large reference genomes[1]. These aligners have allowed

Gregory G. Faust is in the Department of Computer Science, Ira M. Hall is at the Center for Public Health Genomics and both are in the Department of Biochemistry and Molecular Genetics, University of Virginia, Charlottesville, Virginia, USA.
e-mail: irahall@virginia.edu