PERSPECTIVES

A guide to the visual analysis and communication of biomolecular structural data

Graham T. Johnson and Samuel Hertig

Abstract | Biologists regularly face an increasingly difficult task — to effectively communicate bigger and more complex structural data using an ever-expanding suite of visualization tools. Whether presenting results to peers or educating an outreach audience, a scientist can achieve maximal impact with minimal production time by systematically identifying an audience's needs, planning solutions from a variety of visual communication techniques and then applying the most appropriate software tools. A guide to available resources that range from software tools to professional illustrators can help researchers to generate better figures and presentations tailored to any audience's needs, and enable artistically inclined scientists to create captivating outreach imagery.

In the past 45 years, close to 100,000 atomic-resolution structures have been deposited into the Protein Data Bank (PDB), and more than 2,000 molecular-resolution entries have been deposited into the Electron Microscopy Data Bank. Myoglobin was the first protein to be resolved at high resolution[1], which resulted in a PDB file containing data that describe 1,260 atoms. The structure of myoglobin can be explored in three dimensions directly using a web browser by typing the PDB ID code 1MBN into the search box at PDB. Small computers such as current mobile phones deliver enough power to interactively display three-dimensional (3D) representations of a small protein file such as 1MBN with ease; however, large structures that have been published more recently can quickly push the most powerful computers beyond their limits. For example, hybrid methods to explore macromolecular assemblies have recently generated atomic-resolution structures with more than 2.4 million atoms[2], and structural systems biology approaches have assembled molecular models of entire organelles[3,4], large asymmetrical viruses and portions of whole cells[5]. The generation and analysis of such massive models require the simultaneous manipulation and visualization of multiple types of data at different resolutions[6]. Large-scale molecular dynamics simulations add time as a fourth dimension to the innate complexity of structural 3D data; this can greatly increase storage and display requirements, which makes it difficult not only to view the structures but also to communicate their details.

At the same time, there is an increasing need to justify the cost of science and to educate the general public. Hence, in addition to dealing with increasingly complex data, biologists must not only communicate with the scientific community but will often need to face the challenge to engage more general audiences. Importantly, different visual solutions are usually required to communicate with these different audiences effectively. To address these diverse modelling and visualization needs, a vast array of tools for visualizing molecular data have been developed. When combined with the recent trend of using 'Hollywood-calibre' 3D animation software[7–10] or the option to hire illustrators and artists to create visual content, choosing the best tools to generate an ideal visual solution can become an overwhelming task.

In this Essay, we highlight the strengths and weaknesses of common types of software tools for biomolecular visualization in diverse contexts. We recommend techniques for solving difficult problems and outline example protocols, in addition to providing best methods, tips and tricks. The spectrum of techniques that we briefly survey represents a broad and deep continuum in which scientific and artistic minds can unite to better engage audiences ranging from peers to students and from grant reviewers to the general public.

This visualization guide, which is aimed at molecular biologists, provides a broad overview of the needs of different possible audiences and the relevant available tools. Our recommendations presented in this context will typically represent only one of many paths to achieve any particular goal. However, they intend to introduce the knowledge and confidence that a molecular biologist needs to improve a slide, tackle a journal cover, make a movie or even engage in extreme outreach activities such as scientific art or the production of educational video games.

Molecular visualization categories

Understanding the needs and interests of the audience is crucial to producing effective science communication. For the purpose of this Essay, we find it useful to group visualizations for different audiences into four categories, although the boundaries between such categories are not strict (FIG. 1).

Data analysis. Data analysis can be considered as an abstract form of science communication: the presenter and the audience are often the same person, who is using the brain's visual interpretation power to recognize patterns and to shape ideas. Analysis calls for an immediate, interactive and objective visualization of the data, in which a user can quickly switch between types of representation, take measurements, plot comparisons or carry out other evaluations to confirm experimental results or to generate new hypotheses.
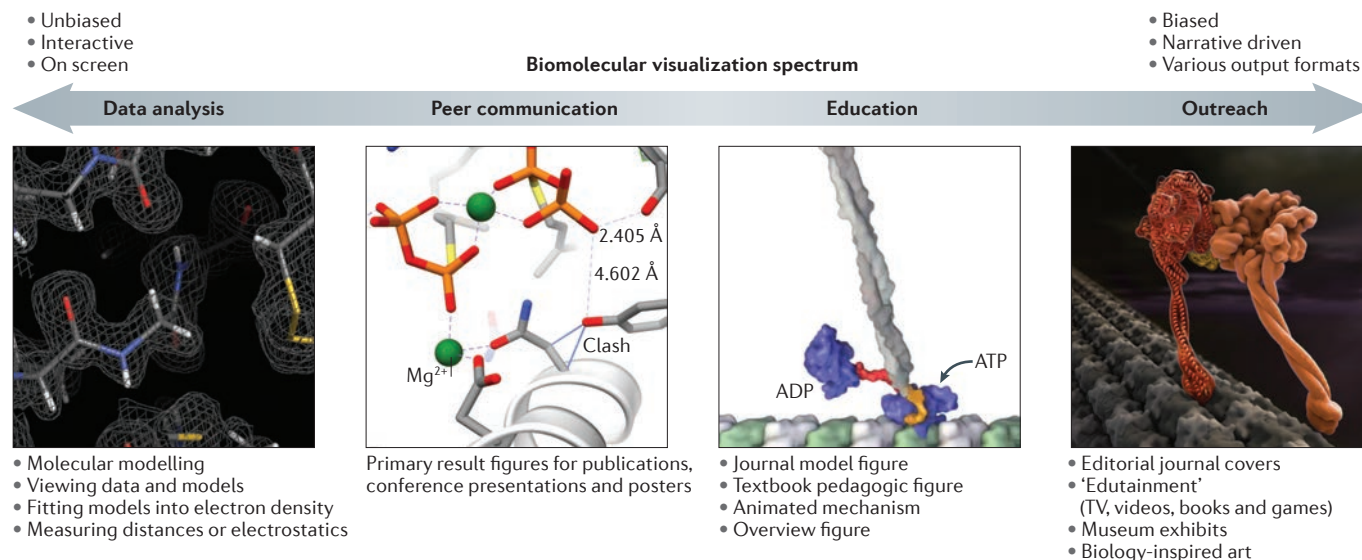
- Unbiased
- Interactive
- On screen

Biomolecular visualization spectrum

- Biased
- Narrative driven
- Various output formats

| Data analysis | Peer communication | Education | Outreach |

- Molecular modelling
- Viewing data and models
- Fitting models into electron density
- Measuring distances or electrostatics

Primary result figures for publications, conference presentations and posters

- Journal model figure
- Textbook pedagogic figure
- Animated mechanism
- Overview figure

- Editorial journal covers
- 'Edutainment'
  (TV, videos, books and games)
- Museum exhibits
- Biology-inspired art

Figure 1 | **The spectrum of biomolecular visualization.** As described by Gaël McGill[46,47], the biomolecular visualization spectrum includes a wide range of audiences. Researchers communicate most effectively when they identify where their target audience lies along this spectrum and should choose the appropriate visual techniques to meet that audience's needs. For example, an audience of colleagues at a field-specific conference will need considerably less background information than an outreach audience of first-time learners. The far left panel shows how well an atomic model represented as sticks along the bonds fits into the isovalued surface (triangular mesh) of an electron density map calculated from X-ray crystallography. The middle left panel shows the backbone of a protein as a white ribbon representation with side chains that interact with ligands represented as stick models. The middle right panel shows a frame from an animated model[48] (created by G.T.J.) to clarify the complex sequence of molecular mechanics of kinesin 'walking' along a microtubule. The far right panel shows the motor protein dynein walking along a microtubule in a dramatically lit cell interior to capture an audience's attention for the cover of a journal[49].

Communicating results to scientists. Presenting results to scientific peers requires a more streamlined type of visualization in which ancillary information can be de-emphasized or omitted to highlight or clarify a point. Even raw data must typically be arranged or condensed to more clearly support a drawn conclusion. Broad scientific audiences have stronger background knowledge on the topic concerned than educational or outreach audiences. Although there is less need to provide an overview of the subject matter, field-specific visual jargon should be clarified. For example, the fact that a ribbon diagram represents the amino acid backbone of a folded protein will rarely have to be explained, but one should consider labelling the different two-dimensional (2D) icons that are often used to represent monosaccharides on a glycosylated molecule if communicating to audiences without a background in glycobiology.

Education. Creating visuals for education usually requires further simplification or abstraction depending on the knowledge level of the target students and the learning goals. Composition, labelling, representation choice and further processing of the data all become more important when communicating to a less specialized scientific audience (that is, towards the right of the biomolecular visualization spectrum (FIG. 1)).

Outreach. Communicating to the world outside the scientific community forms the fourth and perhaps the most challenging type of visualization. The audience may have little knowledge of molecular biology, and their interest might be hard to capture. In order to deliver the message, the data might need to be embedded into a compelling narrative and presented in the context of the bigger picture, such as the *in vivo* environment of a protein under study. In this case, the goal of conveying a story or motivating an otherwise uninterested audience takes priority, which may warrant a higher degree of artistic license[11] to fill in the gaps or to 'edutain' (that is, to educate with captivating entertainment). For journal covers, visual abstraction or deep metaphors can often summarize the complex content or impact of the article (or articles) in an issue better than a diagrammatic figure. Whether literal or figurative, journal editors will usually select or commission imagery that quickly captures the attention of potential readers.

Visual communication techniques

After identifying the needs of an audience, taking the time to carefully select effective visual solutions from an array of techniques before jumping into a software package can save a lot of effort. The simplest solutions will often make the most memorable images, but simple solutions can be hard to envision if one is already mired in the dizzying array of menu and widget options of any particular software package. Therefore, one should always try to sketch ideas, verbalize thoughts and plan labels, regardless of how simple they are, using pencil and paper or some other familiar medium before opening a complex software interface[12]. BOXES 1,2 provide an overview of the useful communication techniques related to molecular visualization and a brief description of how and where to use them.

Multiple audiences: broaden but do not distract. In general, one should avoid clutter in the visuals but expand the potential audience by providing additional details in a way that would not distract. This rule should be applied in both directions along the biomolecular visualization spectrum (FIG. 1). For example, rather than having a transmembrane protein float on a white page, placing it into a lightly outlined lipid bilayer and labelling the compartments on either side of the bilayer could make the figure more useable for educational purposes that go beyond peer communication (see below). Transmembrane protein databases[13,14] can ease this process by showing the thickness of the bilayer around the

Box 1 | Molecular visualization communication techniques

Molecular biologists need to convey a broad range of information within the context of molecular and cellular structure. Different styles of structural representation, levels of detail and visualization techniques can help to clarify a broad range of data for particular audiences that range from specialist researchers to the general public.

Representation of molecular structures

An appropriate level of detail should be chosen to convey the message. One should avoid using the default representations of most molecular viewers, which commonly show every bond as a thin line regardless of the molecule's size (see the figure, part **a**). Coarse surface representations (part **b**) are generally recommended for quaternary structures; ribbons (part **c**) are ideal for secondary and tertiary structures and, along with molecular surfaces (not shown), can illustrate relationships between domains; and bond sticks or atomic spheres (part **d**) work well to illustrate chemical details.

Mixed representations

It is often necessary to use mixed representations to highlight structural features across a broad range of scales. Mixed representations can be used to describe a function, a mechanism of action or the details of chemical reactions. Using a combination of surfaces, ribbons and chemical details (part **e**) can be helpful for data analysis, as well as for presenting data to diverse audiences. An abstracted depiction, for example, that of aquaporin (part **f**), can indicate the mechanism of action of this channel more readily than a fully realistic structural representation using detailed molecular surfaces.
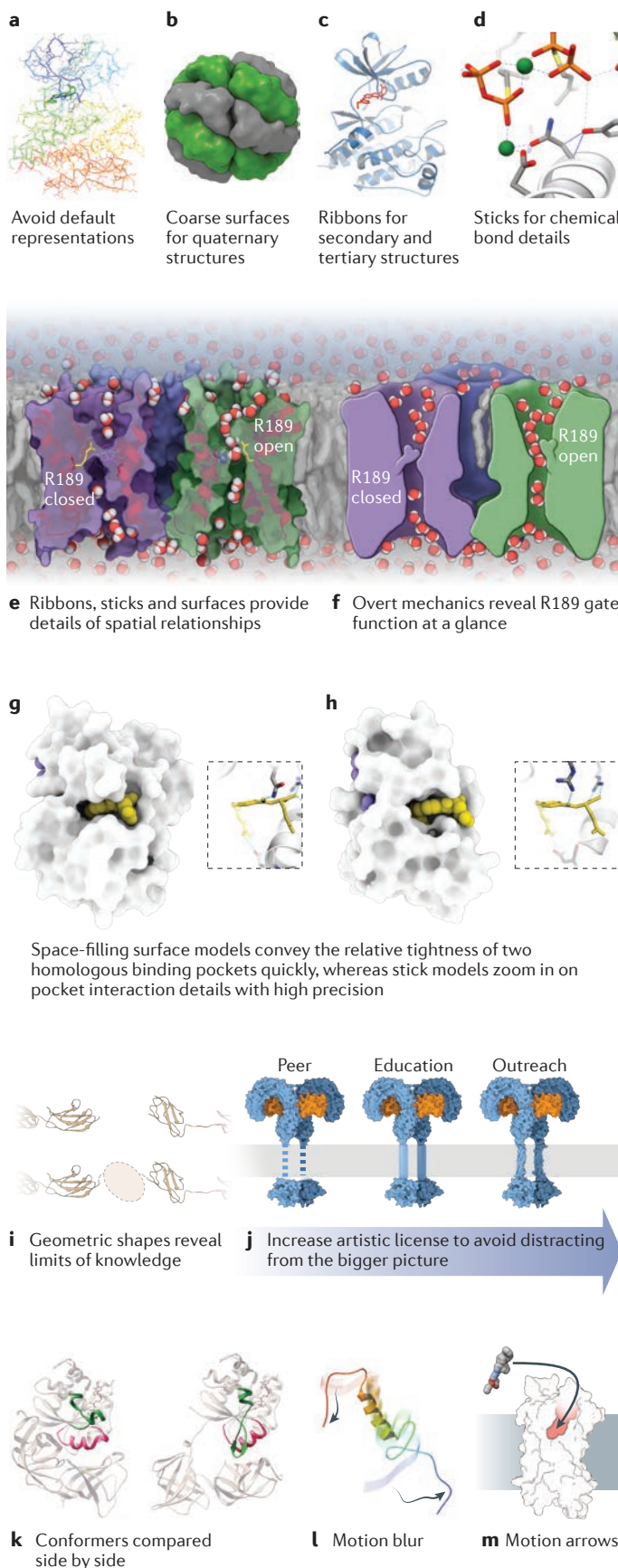
Appropriate level of detail

One should avoid restrictive conventions and use the full spectrum of tools to best convey a particular message. For example, to visually contrast how well ligands (yellow) pack into the binding pockets of two homologous proteins, it might be worth zooming out to show the big picture. Conventional stick models are useful for showing precise details of how the proteins interact with their respective ligands. However, for the intended message, space-filling surface models more quickly convey that the ligand is deep and well protected in human dihydrofolate reductase (DHFR; Protein Data Bank (PDB) ID: 4M6K) on the left (part **g**) compared with the *Escherichia coli* DHFR protein (PDB ID: 3QL3) on the right[36] (part **h**), especially for less-technical audiences. Rendering the surface models with an objective lighting technique such as Ambient Occlusion[37] — which is available in many molecular viewers (such as Python Molecular Viewer (PMV)[24], QuteMol[37] and Visual Molecular Dynamics (VMD)[26]) and in all three-dimensional animation packages — is necessary to highlight the relative tightness and depth of the pockets at a glance.

Unknown structure

Dashed lines, circles or other shapes could be used to reveal a lack of information on a structure. The size and position of the abstractions should approximate the molecular weight of the unknown regions (part **i**). For peer and educational audiences, one should never imply knowledge of higher resolution by making these shapes resemble atomic structures. However, for an outreach audience, using artistic license to draw an informed depiction of a missing structure, such as the transmembrane domain (part **j**), can minimize distractions that would draw attention away from the bigger picture of a transmembrane signal transduction mechanism in this example.

Structural differences and dynamics

Animations are a useful tool to show protein conformational changes over time, and the differences between single structures can be highlighted using morphing[38,39]. For still images, structural differences can be conveyed using small multiples[40,41] of side-by-side comparisons (part **k**) or by superimposing structural conformers on top of one another (not shown). Dynamics can be similarly conveyed by ghosting multiple conformers using motion blur (part **l**) or by drawing simple arrows between two or three superimposed conformations (parts **l,m**).

a Avoid default representations

b Coarse surfaces for quaternary structures

c Ribbons for secondary and tertiary structures

d Sticks for chemical bond details

e Ribbons, sticks and surfaces provide details of spatial relationships

f Overt mechanics reveal R189 gate function at a glance

Space-filling surface models convey the relative tightness of two homologous binding pockets quickly, whereas stick models zoom in on pocket interaction details with high precision

i Geometric shapes reveal limits of knowledge

j Increase artistic license to avoid distracting from the bigger picture

k Conformers compared side by side

l Motion blur

m Motion arrows

structure based on hydrophobic analyses. Similarly, a model figure or textbook image of a cellular event, such as a second-messenger signalling pathway, can often become a useful resource for peer audiences if the details of cellular and molecular structure can be incorporated without distracting from the main goal of conveying the overall signalling pathway (BOX 2).

Representations that are familiar to molecular biologists are only abstract shapes to most other audiences, who often assume incorrectly that such representations have some physical reality. Thus, it is usually best to use artistic license and convey the message using a consistent representation and style. Although outreach audiences are learning that ribbon diagrams represent something nanoscopic[15,16], it is often wise to encase them in a transparent molecular surface to clarify space occupancy and shape, for example, to illustrate that water cannot pass through the sparse ribbon, which is often mistaken for a physical depiction of a protein.

Types of tools
After the goals have been identified and ideal visual solutions planned, appropriate software is needed to execute most techniques. The complete spectrum of biomolecular visualization requires more than simply software for viewing proteins[17]. Whereas almost any tool on the spectrum can be adapted to accomplish any task along this spectrum, choosing the most suitable tool can greatly simplify the work. The visualization workflow can include specialized analysis software, more general molecular viewers, 2D illustration and compositing software, pencil and paper, as well as 3D animation software (FIG. 2).

Analysis tools. Analysis of biomolecular data can be carried out using specialized software designed to handle specific data types or using more general molecular viewers (see below). As the output is typically limited to interactive visualization on computer screens, rendering enhancement options are of less concern and are rarely provided. These software packages tend to emphasize speed, interactivity and functionality. They often have limited graphical user interfaces and require some basic scripting capabilities to input or refine commands. A broad range of open-source software is available for visual analysis[18]. Examples include specialized software for building atomic models into 3D electron density maps (for example, crystallographic

Box 2 | **Guiding the focus of an audience**

Two example images — a textbook figure[28] and a journal cover (see the figure) — use principles of graphic design and various rendering techniques to capture and direct the attention of each audience. The textbook figure uses colours to visually categorize the molecules involved in a cell signalling event into their relevant chapters listed vertically on the left as a means to introduce a new section of the book on cell signalling. The journal cover image uses pills as a substrate metaphor to summarize an article about promiscuous 'multidrug' transporters that exhibit polyspecific recognition (that is, numerous chemically different compounds pass through them). Cover images are often used to attract an audience towards reading an article and ideally the rest of the journal. Some general rules of design can help to direct an audience or to focus their attention in any visualization.

- High contrast and saturated colours, as well as detailed and focused surfaces, draw a viewer's eye. Colours should be used sparingly to save saturated colours for the most important parts of a story[42–44].
- Low contrast, desaturated colours and blurred surface details command less focus or even disappear into the background.
- 'Noisy' or even-toned images fail to guide focus.
- Composition (that is, relative position of objects in the image), as well as carefully placed leader lines and labels, can direct attention to specific locations in the image or guide an audience through a story[45] (see the figure, left panel).
- Outlines can emphasize edge boundaries, but overuse can clutter an image. For example, the beaded ribbons available in PyMOL[25] visually outline the helices to clarify overlaps when images are printed at small size. Most molecular viewers provide ink-like outline options that have similar effects, as shown by the line-drawing of a protein rendered from Chimera[22] (FIG. 3a).
- Depth cueing (where distant objects fade into the background) or a narrow depth-of-field (that is, when objects out of a focal plane become increasingly blurred) can clarify z-depth on an otherwise busy three-dimensional (3D) image. The two background glycoprotein complexes in the figure (right panel) are blurred by the depth-of-field effect applied to the 3D camera when the model was rendered.
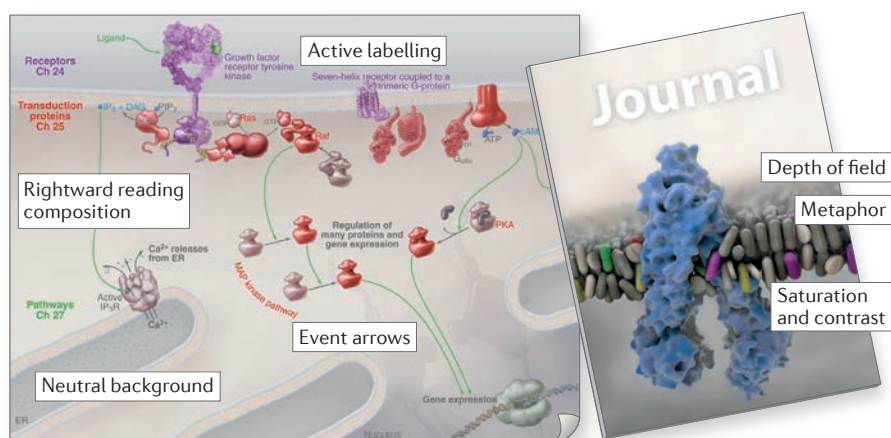
The figure (left panel) was published in *Cell Biology*, second edition, Pollard, T. D., Earnshaw, W. & Lippincott-Schwartz, J, page 425, © Elsevier (2007). The image (right panel) is created by Graham Johnson of www.grahamj.com and TSRI (2010), and was used as the cover image in *Clin. Pharmacol. Ther.*, Vol. 87, Issue 1.

object-oriented toolkit (Coot)), molecular modelling[19] and drug design (for example, SYBYL), and the analysis and comparison of NMR conformers (for example, MOLMOL). A large selection of software packages, such as IMOD[20], Amira, ImageJ and OsiriX, offer tools to visualize and analyse image data spanning from organelles to organisms[21]. Results from these larger-scale tools provide useful distance relationships and other statistical constraints for positioning molecular detail

into cellular contexts more correctly. Visualization experts working anywhere on the biomolecular visualization spectrum will often find it useful to be able to interactively explore published data by loading it into these analysis tools to answer the highly detailed questions that are sometimes asked when producing visualizations elsewhere on the spectrum, for example, when labelling the distance measured between two atoms of an animated macromolecular complex.
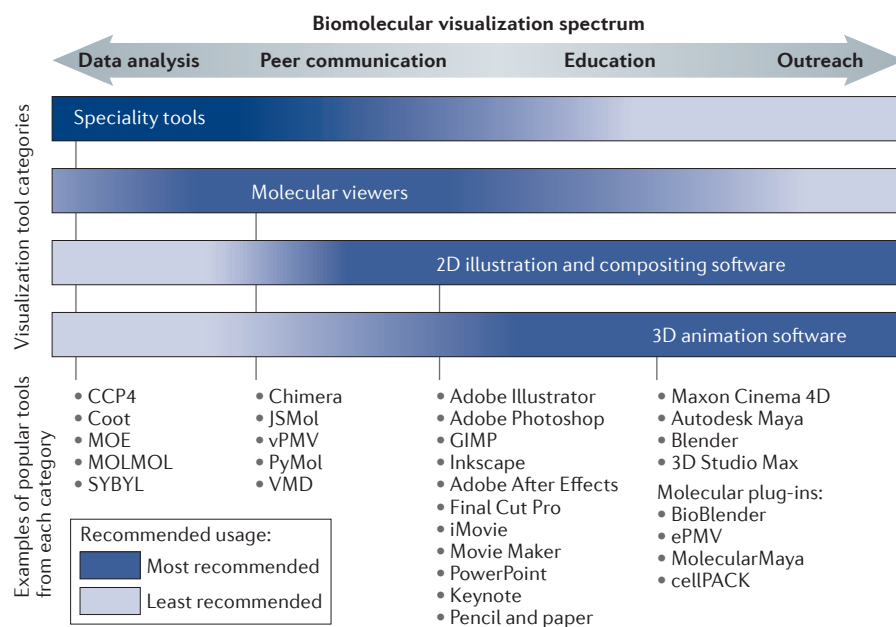
Biomolecular visualization spectrum

Figure 2 | **Visualization tools for specific audience needs.** Certain visualization tools are more ideally suited than others to meet the needs of particular audiences and communication goals listed along the biomolecular visualization spectrum (top arrow). Choosing the right tools can increase the impact of the end product and improve production efficiency. One should work smart by taking the time to plan and select the best techniques, before selecting the most suitable tools with which to execute those techniques. For each of the four tool categories listed below the spectrum, a colour gradient indicates where in the spectrum they are most applicable. Examples of tools that fit into each tool category are listed at the bottom. 2D, two-dimensional; 3D, three-dimensional; Coot, crystallographic object-oriented toolkit; ePMV, embedded Python Molecular Viewer; MOE, Molecular Operating Environment; VMD, Visual Molecular Dynamics.

Molecular viewers. Molecular graphics software carry out a wide variety of tasks and can handle multiple file types. Widely used examples include UCSF (University of California, San Francisco) Chimera[22], Jmol[23], Python Molecular Viewer (PMV)[24], PyMOL[25] and Visual Molecular Dynamics (VMD)[26]. These tools can be used for fast and precise molecular visualization tasks[51], such as emphasizing the protein side chains that interact with a ligand by representing the side chain bonds as sticks branching out from a ribbon backbone (FIG. 1, middle left panel), or for producing a video of a molecular dynamics trajectory. Many molecular viewers can also visualize and analyse volumetric data, such as 3D electron tomography maps or super-resolution microscopy data, with limited capabilities, which makes them useful for bridging the scale from atoms to cells. These viewers typically feature advanced rendering capabilities, which can produce high-quality journal figures, slides and supplementary movies for scientific publications. However, most scientific software has no or limited functionality to undo mistakes. By saving sessions frequently or working on duplicated models, one can usually return to a reasonable point in the workflow after making an irreversible mistake.

3D animation software with embedded molecular viewers. Although molecular viewers are great for handling and displaying DNA, RNA, proteins and some volumetric data, they are less suitable for creating and displaying other 3D content, for example, graphical shapes that represent missing regions of proteins or biological structures such as bilayers and organelles that span scales to lend context to molecular data. Therefore, scientific illustrators and some biologists have turned to 3D animation software to create biomolecular visualizations. These tools offer almost unlimited possibilities for arranging molecules with other data into a common 3D space, for animating or simulating objects and for creating geometric shapes to depict the environment of a molecule over a large range of scales. Examples of such 3D animations include a molecular dynamics simulation of two fibronectin domains in the context of an entire fibronectin fibre (see Supplementary information S1 (movie)), a simulation of

the motor protein kinesin walking along a microtubule[27] (see Supplementary information S2 (movie)), the mechanics of the molecular machine ATP synthase[27] (FIG. 3; see Supplementary information S3 (movie)) and the detection of odours by the olfactory system in an animation that spans from tissues to molecules, which was composited from a collection of 2D images[28,50] (see Supplementary information S4 (movie)). Compared to the other tools, it is relatively easy to unify all of these visualization tasks into one project using 3D animation software, for example, to animate a series of complex cellular events that involve the movement and interaction of both organelles and the molecules within the organelles. These packages have fairly steep learning curves and can be overwhelming for researchers who seek to produce a quick slide, but artistic-minded researchers have excelled in their use. For example, many researchers contribute images to The Art and Science of Cell Biology art show[29], and the team of researcher Dan Gurnon and illustrator Arik Thuesen won second place in the autoPACK Visualization Challenge[30]. Popular examples of 3D animation software include Blender, Maxon Cinema 4D and Autodesk Maya. To directly import atomic-resolution structure files from the PDB or electron microscopy density maps, several plug-ins have been developed for 3D animation software to facilitate the workflow and enable continuous and direct interaction with the data. These molecular plug-ins include BioBlender, the embedded PMV (ePMV)[9] and MolecularMaya. Biologists taking our molecular animation classes tend to prefer to learn with Maxon Cinema 4D, which is well known for its ease of use, but all professional animation packages supported by molecular plug-ins have similar power and capability. Most 3D animation packages are freely available with limited licenses for academic trials. Blender is free and open-source, although learning to use it can be relatively difficult. MolecularMovies is a portal that provides access to a wide range of high-quality life science animations and tutorials.

2D illustration and compositing software. We strongly recommend using 2D compositing software to assemble or complete all types of visual science communication. Using such programs to assemble and label the images from analysis tools, molecular viewers and 3D software saves time and keeps visuals editable and legible. Although molecular viewers and 3D animation

packages have the ability to render text, such as residue labels and distance measurements, the text gets rendered into these images as pixels with limited resolution and is not editable outside the software. Text rendered into images is more difficult to read in journal figures or on slides than the mathematically defined 'vector graphics' text that can be added on top of the image as an additional step in any 2D compositing software. To maintain a non-destructive workflow (that is, one in which undoing back to a previous state in the process is easy), individual molecule renderings, labels and other graphics should be assembled on discrete layers using some type of compositing software so that they can be edited or repurposed easily later without the need to revisit the data. The molecular animator Drew Berry[31] builds up much of the sophisticated intricacies of his work[32] by compositing dozens of transparent layers, which are exported from Autodesk Maya, together using Adobe After Effects. For images, there are some open-source alternatives to the powerful but expensive Adobe Illustrator and Photoshop packages, for example, GIMP and Inkscape, but users often find these interfaces harder to use than the commercial packages. Microsoft PowerPoint or Keynote presentation software is often the best choice for simple labelling or adding context when preparing slides. Although options are limited, the ability to repurpose a slide simply by editing a label may be worth forgoing visual appeal (FIG. 3a). The availability of open-source software for movie editing is relatively limited and, if budgets restrict access to professional software (for example, Adobe After Effects, Apple's Final Cut Pro, Autodesk's Smoke and The Foundry's Nuke), users can choose to use Blender (open source), iMovie (for Mac; inexpensive) or Movie Maker (for Windows; free) for simple movie editing and labelling tasks.

Compositing approaches are best used for quickly adding context. For example, instead of struggling to add a bilayer to a molecular viewer, one could draw a rectangle in PowerPoint in less than ten seconds to get the same message across, often more clearly. Micrographs or review articles should be used for reference when drawing or modelling background context (typically, a larger scale of information such as the environment of a protein). It is usually worth mimicking organic shapes for outreach audiences, but rectangles and squares are often good enough to make a quick slide that is intended only for peers.

Programming. Complex animations are often easier to simulate or program rather than to animate using more traditional 'keyframe' interpolation approaches. All molecular viewers and 3D packages mentioned above have Python programming interfaces to unify this process, and Python is therefore a useful programming language to learn[33] for biologists who are interested in producing more involved molecular visualizations. The increasing prevalence of online molecular viewers makes the JavaScript language similarly useful. The increasingly popular computational platform MATLAB offers a molviewer application with Jmol scripting access as part of its bioinformatics toolbox, as well as universal and other biology-specific analysis, modelling, simulation and visualization capabilities.

Software protocol examples
After goals have been identified and ideal visual solutions planned, the appropriate software could be quickly selected to execute the techniques. FIG. 3 outlines the creation of a sample educational textbook-type figure, an editorial game image and a teaching animation. It highlights the efficient application of compositing tools and protocols to assemble visual stories from carefully chosen, but quickly generated, molecular representations. Regardless of the complexity of a project, consistently identifying an audience's needs, planning visual solutions and then applying appropriate tools can markedly improve impact while minimizing production time. As an example, to highlight the relationship between the bound ligand of a transmembrane G protein-coupled receptor (GPCR) and the hydrophobic region of a lipid bilayer, we choose to de-emphasize the protein and the bilayer by fading them into the colour of the background (FIG. 3a). This can be achieved by carefully selecting colouring and lighting options in the molecular visualization software or, more efficiently, by post-processing the default output images during compositing, in which results of subtle parameter changes can be immediately perceived in the final context of the labelled figure. In this example, we render grey outlines of the GPCR using the 'silhouette' feature in Chimera, then paste them together and apply background graphics (that is, rectangles for the leaflets of the bilayer and colour gradients for the extracellular matrix and cytoplasm) using Apple's Keynote presentation software to maintain a non-destructive workflow

that quickly results in an editable slide. Reducing the opacity of the GPCR binding pocket and fading off the jagged top edge help to clarify that this pocket is found inside the protein. Photo-editing software provides more options to make a slightly cleaner slide, but often this might not be worth the effort for peer communication or for other well-prepared educational audiences. Investing more time into compositing can put the focus on the binding pocket of the receptor (see tutorials for both approaches at Mesoscope). An experienced user can also use molecular-enabled 3D software (for example, using the ePMV plug-in) to create a powerful editorial image or moody game character[34] in less than an hour (FIG. 3b). The same basic process can make the production of an animation manageable. Short animations benefit from a simple sketch to decide how the components can be fitted into the scene with some brief action descriptors, for example, arrows of intended molecular movement and labels. The same general workflow can also be applied to complex projects, such as a three-minute animation to teach the mechanism of ATP synthase[27] (FIG. 3c).

Outlook
Although the suite of diverse software tools can be daunting to navigate, they now provide complete access to the biomolecular visualization spectrum. We find from teaching and experience that most biologists prefer to use academic freeware for molecular analysis and data visualization, but tend to learn faster and prefer using commercial packages for compositing and presentation, as they are easier to use. New trends towards renting cloud-managed software from commercial companies may make these options more attractive to many research groups who use them less frequently. Using the best tool for any particular job should reduce frustration and increase efficiency, which may be worth any associated costs.

Researchers acquire specialized skills throughout their careers. After assessing an audience's needs, a scientist may reason that visualization techniques falling far outside his or her range of skills on the spectrum could provide the best solution. Therefore, one should decide early in the process whether to invest the time and cost to learn these techniques or whether it would be wiser to collaborate. Many laboratories might have artistic talents and would be able and eager to solve these

a Three steps to produce efficient and effective visuals

Step 1: Identify goals

Show the ligand binding to emphasise depth of GPCR pocket relative to bilayer for overview slide at Monday's lab meeting

**Step 2: Plan solution and select ideal tools**
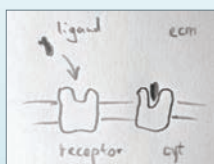
Sketch a visual solution

1. Render GPCR as surface
2. Render isolated pocket to make it fade as if inside
3. Render ligand as atoms
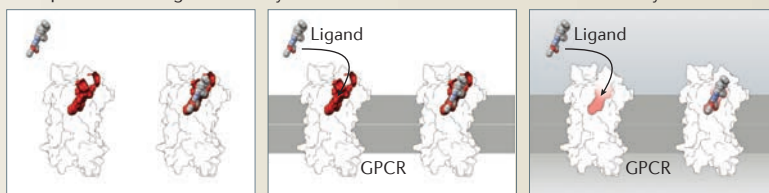4. Make rectangle for bilayer using Keynote

Write a protocol

Step 3: Apply the visualization tools as planned

Export image components on transparent backgrounds from a molecular viewer

Surface

Binding pocket

Ligand
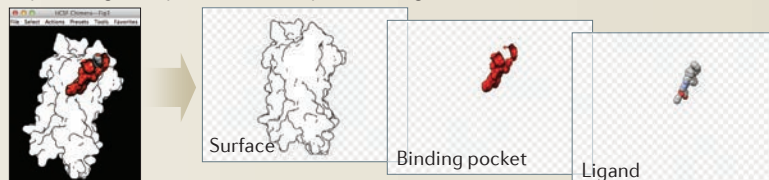
Composite the images into a layered slide to save time and retain editability

Arrange layers

Add graphics and text

Adjust emphasis and focus

b Dramatically different styles suit different audiences
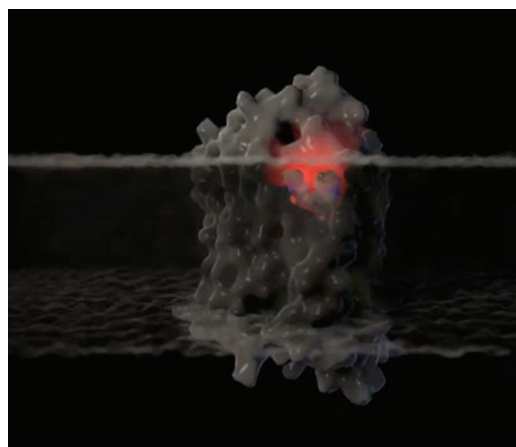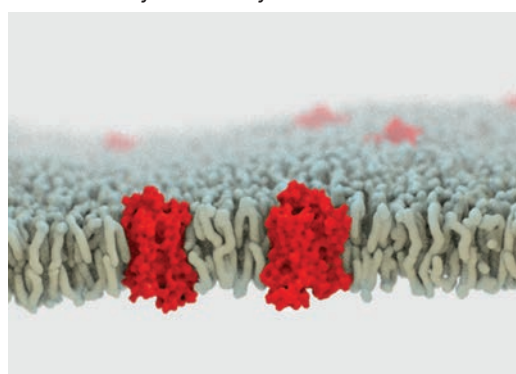
c The same process applies to create more complex visuals such as animations

Step 1: Identify goals

Clarify ATP synthase function with structural anatomy and turbine mechanics for freshmen biology students.
Include:
proton gradient energy to mechanical energy to ATP chemical bond energy
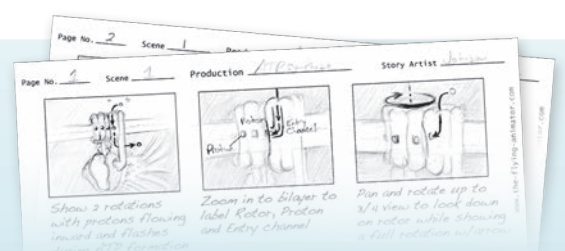
Step 2: Plan solution and select ideal tools

Sketch a rough concept

1. Model and assemble ATP synthase PDB fragments as ribbons and coarse surfaces
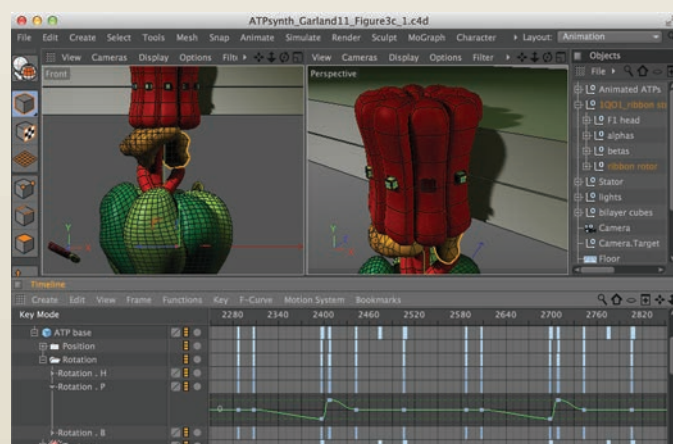2. Rig conformation changes from 120° superimposed rotations
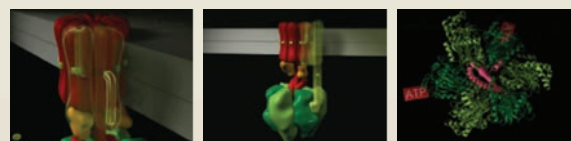
Write a protocol

Refined and detailed storyboards save animation time

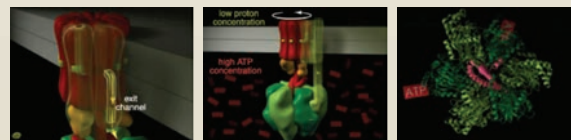**Step 3: Apply the visualization tools as planned**

Molecular models are generated and animated to match the storyboard

Storyboard frames are typically turned directly into a narrated animatic movie, and gradually replaced with polished sequences of animation

Three frames from the raw animated movie

Three frames from the composited movie

◄ Figure 3 | **Three general steps to produce effective visuals in any medium.** Despite having different audiences, goals and presentation media, the two sample project protocols follow the same workflow pattern to accomplish visual communication efficiently by identifying goals (step 1), planning solutions (step 2) and then applying visualization tools (step 3). **a** | In this example, a researcher wants to emphasize the depth of the binding pocket of a G protein-coupled receptor (GPCR) relative to the surrounding bilayer in a slide that he or she will present to an audience of general molecular biologists. By planning ahead (steps 1 and 2), simple images rendered from a molecular viewer can be composited efficiently with additional graphics directly in slide presentation or other compositing software (step 3) to create an information-rich slide that allows easy editing or future repurposing. **b** | Different molecular representations, as well as different content, composition, colour and mood choices, create different image styles that suit different audiences. Although the fast composition method (outlined in part **a**) works well for peer communication and education, broader audiences tend to respond to more artistic representations. By contrast, animation software offers more options for generating more artistic molecular representations for cover art or other outreach goals. For example, the top panel shows a homologous GPCR in a detailed lipid bilayer to make an engaging web banner image, whereas the lower panel shows the same GPCR rendered using the three-dimensional (3D) animation software Maxon Cinema 4D through embedded Python Molecular Viewer (ePMV) for a game prototype to engage teenagers. **c** | The same steps outlined in part **a** can be applied to more complex visuals such as animations, although longer animations require considerably more planning. After identifying the goals in step 1, step 2 involves writing a script that includes a description of the actions and the narration while optionally exploring mood, style and composition by creating concept art (that is, static images of critical 'scenes' planned in the final animation). Next, a storyboard provides a visual depiction of the script. Storyboards should be created with the visual tools that are most comfortable to the researcher, for example, by sketching with pencil and paper as shown, or by arranging simple shapes in slide presentation or compositing software. To transition efficiently into step 3, the storyboard images are then cut into individual frames, which are played sequentially to provide a slideshow version of the movie timed to the narration to create an 'animatic'. The animatic imagery is gradually replaced with rendered frames from the polished movie. The output of this protocol is shown in Supplementary information S3 (movie). 2D, two-dimensional; ECM, extracellular matrix. Part **c** © 2002 from *Molecular Biology of the Cell*, 4th Edition by Alberts *et al.* Reproduced by permission of Garland Science/Taylor & Francis LLC.

problems; alternatively, one could select from hundreds of professional or academic scientific visualization experts from organizations such as the Association of Medical Illustrators and the Guild of Natural Science Illustrators. Medical illustrators are rigorously trained to convey complex information to a range of audiences clearly, accurately and aesthetically. For extreme artistic license from Hollywood-style talent, particularly for grabbing the attention of an outreach audience, one could consider working with members of the Computer Graphics Society and many others.

Several imminent projects hold promise to bring powerful visualization tools and community collaboration options directly to web browsers and mobile devices[17]. Molecular Flipbook offers user-friendly molecular modelling and animation options, and Autodesk's Project Cyborg provides powerful scripting, simulation and visualization modules in a browser-to-cloud interface. The cellPACK project[35] is building a browser-based interface to a community-curated database of whole cells modelled in molecular detail and has already been used to engage outreach audiences by hosting a contest to create molecular movies about HIV (see autoPack

Visualization Challenge). These projects mark the beginning of a revolution in computing convenience and collaboration that will make the entire spectrum of biomolecular visualization more accessible to everyone.

Graham T. Johnson and Samuel Hertig are at the California Institute for Quantitative Biosciences (QB3), University of California, San Francisco; and the Resource for Biocomputing, Visualization, and Informatics, Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94158, USA.

Graham T. Johnson is also at the the Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, California 94158, USA.

Correspondence to G.T.J.
e-mail: graham.johnson@ucsf.edu

1. Kendrew, J. C. *et al.* A three-dimensional model of the myoglobin molecule obtained by X-ray analysis. *Nature* **181**, 662–666 (1958).
2. Zhao, G. *et al.* Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature* **497**, 643–646 (2013).
3. Takamori, S. *et al.* Molecular anatomy of a trafficking organelle. *Cell* **127**, 831–846 (2006).
4. Wilhelm, B. G. *et al.* Composition of isolated synaptic boutons reveals the amounts of vesicle trafficking proteins. *Science* **344**, 1023–1028 (2014).
5. Johnson, G. *et al.* 3D molecular models of whole HIV-1 virions generated with cellPACK. *Faraday Discuss.* (in the press).
6. Ward, A. B., Sali, A. & Wilson, I. A. Biochemistry. Integrative structural biology. *Science* **339**, 913–915 (2013).
7. Iwasa, J. H. Animating the model figure. *Trends Cell Biol.* **20**, 699–704 (2010).
8. McGill, G. Molecular movies… coming to a lecture near you. *Cell* **133**, 1127–1132 (2008).
9. Johnson, G. T., Autin, L., Goodsell, D. S., Sanner, M. F. & Olson, A. J. ePMV embeds molecular modeling into professional animation software environments. *Structure* **19**, 293–303 (2011).
10. Autin, L., Johnson, G., Hake, J., Olson, A. & Sanner, M. uPy: a ubiquitous CG Python API with biological-modeling applications. *IEEE Computer Graph. Appl.* **32**, 50–61 (2012).
11. Goodsell, D. S. & Johnson, G. T. Filling in the gaps: artistic license in education and outreach. *PLoS Biol.* **5**, e308 (2007).
12. Wong, B. & Kjaegaard, R. S. Pencil and paper. *Nature Methods* **9**, 1037 (2012).
13. Kozma, D., Simon, I. & Tusnády, G. E. PDBTM: Protein Data Bank of transmembrane proteins after 8 years. *Nucleic Acids Res.* **41**, D524–D529 (2013).
14. Lomize, M. A., Lomize, A. L., Pogozheva, I. D. & Mosberg, H. I. OPM: orientations of proteins in membranes database. *Bioinformatics* **22**, 623–625 (2006).
15. Rundgren, C. J. & Tibell, L. A. E. Critical features of visualizations of transport through the cell membrane — an empirical study of upper secondary and tertiary students' meaning-making of a still image and an animation. *Int. J. Sci. Math. Educ.* **8**, 223–246 (2010).
16. Kramer, I. M., Dahmani, H. R., Delouche, P., Bidabe, M. & Schneeberger, P. Education catching up with science: preparing students for three-dimensional literacy in cell biology. *CBE Life Sci. Educ.* **11**, 437–447 (2012).
17. O'Donoghue, S. I. *et al.* Visualizing biological data — now and in the future. *Nature Methods* **7**, S2–S4 (2010).
18. O'Donoghue, S. I. *et al.* Visualization of macromolecular structures. *Nature Methods* **7**, S42–S55 (2010).
19. Hornus, S., Levy, B., Lariviere, D. & Fourmentin, E. Easy DNA modeling and more with GraphiteLifeExplorer. *PLoS ONE* **8**, e53609 (2013).
20. Kremer, J. R., Mastronarde, D. N. & McIntosh, J. R. Computer visualization of three-dimensional image data using IMOD. *J. Struct. Biol.* **116**, 71–76 (1996).
21. Walter, T. *et al.* Visualization of image data from cells to organisms. *Nature Methods* **7**, S26–S41 (2010).
22. Pettersen, E. F. *et al.* UCSF chimera — a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
23. Hanson, R. M. Jmol — a paradigm shift in crystallographic visualization. *J. Appl. Crystallogr.* **43**, 1250–1260 (2010).
24. Sanner, M. F. Python: a programming language for software integration and development. *J. Mol. Graph. Model.* **17**, 57–61 (1999).
25. DeLano, W. L. The PyMOL molecular graphics system. *DeLano Scientific* [online], http://www.pymol.org (2002).
26. Humphrey, W., Dalke, a. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38, 27–28 (1996).
27. Alberts, B. *et al. Molecular Biology of the Cell* 4th edn (Garland Science, 2002).
28. Pollard, T. D., Earnshaw, W. & Lippincott-Schwartz, J. *Cell Biology* 2nd edn Ch. 27 (Elsevier, 2007).
29. Iwasa, J. H. & Johnson, G. The art and science of cell biology (ASCB²). *ASCB Newsletter* **35**, 3–5 (2012).
30. Johnson, G. T. Announcing the Winners of the autoPACK visualization Challenge2012: present HIV in blood plasma. *AMI Newsletter* (AMI, 2013).
31. Johnson, G. T. Interview iwth Drew Berry, 2010 MacArther Fellow. *AMI Newsletter* (Association of Medical Illustrators, 2011).
32. Berry, D. Molecular animation of cell death mediated by the fas pathway. *Sci. STKE* **2007**, tr1 (2007).
33. Downey, A. *Think Python* (O'Reilly Media, 2012).
34. Lv, Z. et al. Game on, science — how video game technology may help biologists tackle visualization challenges. *PLoS ONE* **8**, e57990 (2013).
35. Johnson, G. T. *et al.* cellPACK: a virtual mesoscope to model and visualize structural systems biology. *Nature Methods* (in the press).

36. Bhabha, G. *et al.* Divergent evolution of protein conformational dynamics in dihydrofolate reductase. *Nature Struct. Mol. Biol.* **20**, 1243–1249 (2013).
37. Tarini, M., Cignoni, P. & Montani, C. Ambient occlusion and edge cueing to enhance real time molecular visualization. *IEEE Computer Graph. Appl.* **11**, 1237–1244 (2006).
38. Krebs, W. G. & Gerstein, M. The morph server: a standardized system for analyzing and visualizing macromolecular motions in a database framework. *Nucleic Acids Res.* **28**, 1665–1675 (2000).
39. Farrell, D. W., Lei, M. & Thorpe, M. F. Comparison of pathways from the geometric targeting method and targeted molecular dynamics in nitrogen regulatory protein C. *Phys. Biol.* **8**, 026017 (2011).
40. Tufte, E. R. Envisioning information. *Optom. Vision Sci.* **68**, 322–324 (1991).
41. Tufte, E. R. & Graves-Morris, P. *The Visual Display of Quantitative Information* (Graphics Press Cheshire, 1983).
42. Wong, B. Points of view: color coding. *Nature Methods* **7**, 573–573 (2010).
43. Wong, B. Points of view: color blindness. *Nature Methods* **8**, 441–441 (2011).
44. Wong, B. Points of view: avoiding color. *Nature Methods* **8**, 525–525 (2011).
45. McCloud, S. *Understanding Comics: the Invisible Art* (Tundra Pub., 1993).
46. McGill, G. Protein dynamics. *EMBL Conference on Visualizing Biological Data* [online], http://vizbi.org/Videos/26199717 (2011).
47. Miko, I. Episode 19: HMS's Gael McGill on molecular visualization. *Nature Education* [online], http://www.nature.com/scitable/blog/natureedcast/episode_19_hmss_gael_mcgill (2011).
48. Vale, R. D. & Milligan, R. A. The way things move: looking under the hood of molecular motor proteins. *Science* **288**, 88–95 (2000).
49. Carter, A. P., Cho, C., Jin, L. & Vale, R. D. Crystal structure of the dynein motor domain. *Science* **331**, 1159–1165 (2011).
50. Pollard, T.D., Earnshaw, W. C. & Johnson, G. T. *Electronic Image Collection for Cell Biology* (W. B. Saunders, 2002).
51. Mura, C., McCrimmon, C. M., Vertrees, J. & Sawaya, M. R. An introduction to biomolecular graphics. *PLoS Comput. Biol.* **6**, e1000918 (2010).

FURTHER INFORMATION
Association of Medical Illustrators: http://ami.org
AutoPack Visualization Challenge: http://autopack.cgsociety.org
Bioblender: http://www.bioblender.eu
cellPACK project: http://www.cellpack.org
Computer Graphics Society: http://www.cgsociety.org
ePMV: http://epmv.scripps.edu
GraphiteLifeExplorer: http://www.lifeexplorer.eu/gallery
Guild of Natural Science Illustrators: http://gnsi.org
Mesoscope: http://www.mesoscope.org/publications/nrmcb
Molecular Flipbook: http://www.molecularflipbook.org
MolecularMaya: http://molecularmovies.com/toolkit
MolecularMovies: http://www.molecularmovies.com
PDB: http://www.pdb.org
Project Cyborg: http://cyborg.autodesk.com
The Art and Science of Cell Biology: http://www.ascb2.org

SUPPLEMENTARY INFORMATION
See online article: S1 (movie) | S2 (movie) | S3 (movie) | S4 (movie)

ALL LINKS ARE ACTIVE IN THE ONLINE PDF