RESEARCH STRATEGY

A. SIGNIFICANCE

Background: Over the past few decades, molecular graphics has moved from the purview of structural biologists into the realm of many life scientists. Originally developed to visualize, interpret and manually manipulate only atomic-scale data, molecular modeling is now also used for presentation, education, and computational studies of molecules, macromolecular assemblies, and even larger-scale data such as from 3D optical microscopy. In recent years, several factors have also greatly affected the development of modeling applications: the rapid increase in size of data sets, the proliferation of web resources, and the rise of parallel computing using commodity hardware[1].

In the early stages of molecular graphics systems, the largest data sets that could be visualized interactively were single molecules, perhaps with density maps from X-ray crystallography. Data size has grown rapidly as new experimental methods have been developed. Cryo-electron microscopy (cryo-EM) has produced density maps for larger molecules and assemblies[2,3]. As microscopy resolution improves, more and more data sets of varying scales (atomic, molecular, even cellular) have become available. Computational methods have also contributed to the growth. For example, molecular dynamics generates thousands of coordinate sets, and integrative hybrid modeling[4] produces large ensembles of candidate macromolecular assembly structures. In our experience, no current molecular modeling package handles all of these new and large data types gracefully.

The number of web resources, both databases and computational services, has increased steadily as well[5-7]. While the Worldwide Protein Data Bank (wwPDB) is still the leading repository for experimentally solved atomic structures, many other repositories are of interest to researchers: ModBase[8] for predicted atomic structures; EMDataBank[9] for electron microscopy density maps; and sequence and domain databases from NCBI[10,11], just to name a few. Computational services are also numerous: BLAST[12] from NCBI for finding similar sequences, DOCK Blaster[13] for structure-based virtual screening, MolProbity[14] for evaluating and repairing protein structures, *etc.* This wide variety of resources creates a problem for molecular graphics applications because each requires its own access method and offers results in its own format.

Finally, hardware improvement has changed in flavor in the past few years[15]. Previously, computing performance improved through faster individual processors. More recently, performance has been improved by increasing the number of processors (CPUs) and taking advantage of the parallelism of graphics processors (GPUs). Faster individual processors benefit all applications, because the same code with no changes will simply run faster. Increasing the number of processors, on the other hand, does not help applications that were designed to run on a single processor. Unfortunately, most visualization applications are single-threaded and therefore do not immediately benefit from these extra processors. There are some standard libraries for parallel computing (OpenMP[16], OpenMM[17]) to take advantage of multiple processing units, but applications must be modified to make use of them.

This proposal is to complete the development of the ChimeraX[18] visualization software projected to replace our 15-year-old widely used Chimera software in ~2020. ChimeraX has been under development for the past two years by our NIH Biotechnology Research Resource (BTRR) center. It has been downloaded 4700 times, cited by prominent Science articles[19,20], and used to render Protein Databank website images. Researchers from other universities are developing sophisticated cryo-EM refinement and validation tools as extensions to ChimeraX. We describe some of the current ChimeraX capabilities in a recent journal article[18]. This proposal seeks funding for development of ChimeraX core features and applications to cryo-EM structure determination. Additional advanced ChimeraX applications such as 3D light-sheet microscopy analysis and virtual-reality user interfaces are in development and will be funded from other sources. ChimeraX is the fourth generation (following Chimera, MidasPlus, and Midas) of highly successful visualization software developed and distributed over 45 years by the same team of six developers, with a cumulative 150 years of experience in this domain. ChimeraX will replace Chimera, the cornerstone visualization program for high-resolution cryo-EM, a program currently being cited in 200 new research publications per month with over 12,500 citations to date. Foundational software components of Chimera such as the Python 2 language (to be retired in 2020), Tk window toolkit (obsolete; development ceased several years ago), and OpenGL before the era of

programmable graphics processors are all updated to current state-of-the-art software components in ChimeraX.

Our vision is to create an innovative and integrated multiscale modeling environment that enables researchers to interactively access, visualize and analyze structural data of varying resolution on scales ranging from atomic to cellular. The centerpiece of the environment is ChimeraX, our new high-performance visualization application for three-dimensional data, including atomic structures, density maps, and microscopy data, built using standard toolkits that utilize the parallel computing capabilities of modern hardware. Other types of data such as protein sequences and modeling restraints may be displayed and analyzed using integrated tools that link the disparate data types. For example, selecting an amino acid in the protein sequence view will also highlight the same amino acid in the 3D graphics view, and vice versa. A wide range of web resources, databases, and computational services will be accessible directly through ChimeraX user interfaces. Services, including interactions with other application suites on the user's computer such as Phenix[21], Rosetta[22] or IMP[23] for model refinement, will be available as ChimeraX tools created by our lab and other labs. This multiscale modeling environment will enable researchers to view, analyze and model biomolecular systems using a suite of integrated tools.

Impact: It is a well-accepted adage that visualization is one of the major tools for interpreting and analyzing scientific data, communicating hypotheses and conclusions to scientific colleagues, and educating students and the public. However, the enormous growth in both size and complexity of biological data sets in recent years due to rapid advances in experimental and computational methods has made scientific visualization and analysis of the underlying models and data more challenging than ever. Researchers require tools that enable them to see and understand collections of large and disparate data. As documented in this proposal, ChimeraX will have a large positive impact in this critical area of biomedical research, allowing users to effectively visualize and analyze a wide range of structural data at multiple scales. This proposal focuses on expanding ChimeraX's initial implementation into a fully functional next-generation visualization and analysis tool that can be used to solve important and challenging research problems, as well as a development platform that individual scientist-users can build on for their own specialized research needs, for example, by implementing algorithms to handle data of novel types.

The large molecular complexes being solved at atomic resolution by cryo-EM present immense challenges to understanding the function of million-atom assemblies. Many of the analysis challenges are the result of existing software not being able to handle the richer and larger scale of data coming from new instruments. A broad array of assessment, model-building, and measuring capabilities are needed, and the project proposed here is explicitly focused on developing such software. In the near future this need will become even more acute as new National Cryo-EM Centers are established as a result of the NIH Transformative High Resolution Cryo-Electron Microscopy program.

Creating a software platform for interactive visualization and analysis of large and complex structural data sets is a significant interdisciplinary challenge. Domain experts are needed to guide algorithm development and define common workflows. Computer science expertise is required for design and implementation of efficient algorithms and user interfaces. And best-practice software engineering principles are required for project tracking, documentation, and bug reporting and tracking. The long-term success of our previous-generation visualization software, UCSF Chimera, is evidence that we are well suited for undertaking this challenging task.

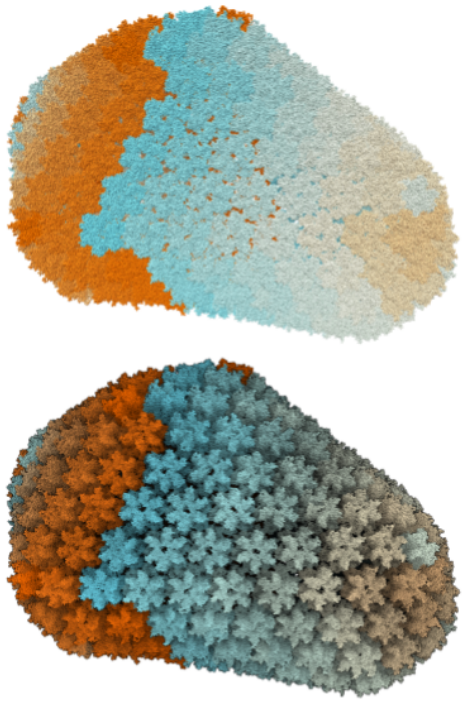

Figure 1: "Simple" model lighting (top) versus ambient occlusion (bottom) of the entire HIV-1 capsid.

B. INNOVATION

Each of aim of this project will develop innovative capabilities vital to molecular research in thousands of labs. Aim 1 builds the foundation for visualizing and analyzing diverse data

types from cryo-EM, X-ray, NMR, cross-linking, multiple sequence alignments, FRET, chromosome conformation capture, light-sheet and super-resolution 3D microscopy, small-angle scattering, molecular dynamics, and others. Combined analysis of multiple types of data is a key strength of ChimeraX. While other visualization packages emphasize specific types of data or specific problems such as drug-receptor interactions, the intent in ChimeraX is to permit simultaneous examination of all data types relevant to structure analysis. Aim 1 also provides new visualization capabilities such as interactive ambient occlusion lighting, which dramatically improves the perception of depth in depictions of large molecular assemblies and 3D microscopy (Figure 1), but has only become possible with recent advances in graphics processors. Ambient shadowing allows more effective communication of research results through publication figures.

Aim 2 tackles challenges arising from new atomic-resolution cryo-EM data. These maps often have greatly varying resolutions across the structure. Misplaced residues and register shifts at 3Å resolution are common, often with functional implications, for example in insulin receptor[24], chaperone p97[25], complement C4[26], and Aim 2 will provide validation tools to identify errors and modeling tools to correct them. It will improve map visualization to produce the most accurate structures permitted by the data, enabling interactive sharpening, and highlighting cut edges when clipping maps. It will enable working efficiently with larger map sizes, for example, optimizing rendering by putting microscopy map and segmentation data on the graphics card to allow fast changes in display settings.

Aim 3 enables the community developing structure analysis algorithms to share their innovations much more broadly. Novel algorithm developments too often end with publication and minimal use by the community for lack of a distribution mechanism that reaches the potential audience. The combination of built-in visualization and analysis tools for a wide variety of data and the extensible architecture for integration with other applications and web resources offers the unique opportunity to create a scientific visualization platform that serves both researchers and developers of new methods. Together these innovations will significantly advance many important and biomedically relevant research projects.

C. APPROACH

Preliminary Results: As the next-generation visualization application and computing platform from our lab, the Resource for Biocomputing, Visualization and Informatics (RBVI), ChimeraX is designed to respond to many modern biomedically relevant scientific challenges. These include the need to efficiently handle the multimillion-atom experimentally determined structures now found in the Protein Data Bank[27] (PDB) as well as the vastly increased complexity inherent in very large structures and multiprotein complexes. To date, we have primarily focused on four areas: i) graphics quality and the software components necessary to achieve fast interactive performance, including parsing large mmCIF-formatted data files from the PDB and storing the associated data in memory-efficient internal data structures; ii) a preliminary implementation of a mechanism, the “Toolshed,” to facilitate community development of new algorithms and methods by building upon the ChimeraX platform; iii) development of advanced versions of key algorithms, including fast and robust solvent-excluded surfaces with clean patch boundaries, and the ability to show helices as curved-cylinder “tubes”; and iv) the basic infrastructure to visualize and analyze integrative hybrid models. Results in areas i) and iv) are described below, but space limitations prohibit a full discussion of all current capabilities.

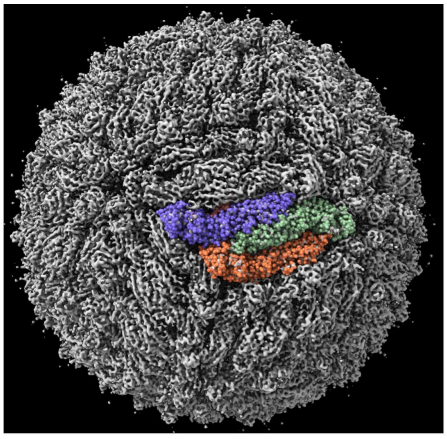

Figure 2: Zika virus capsid cryo-EM map at 3.8 Å and atomic model (EMDB 8116, PDB 5IRE).

With ChimeraX, performance is a primary design goal, both in terms of the size of data sets that may be used and the interactive response speed required to manipulate these data. One of the largest entries currently in the wwPDB is 3J3Q, an atomic-level structure of the entire HIV-1 capsid, consisting of 2.4 million atoms in 1,356 chains (Figure 1). The structure is not available in the old column-oriented “PDB” data format, which can only accommodate 100,000 atoms per file. Instead, the data is available in macromolecular Crystallographic Information File[28,29] (mmCIF) format as a 266-megabyte text file. ChimeraX uses our new high-performance mmCIF parser module, developed by team member Greg Couch, to open and display 3J3Q in a little over 10 seconds, using 1.75 gigabytes of memory.

The new Zika virus PDB entry (5IRE -- Figure 2) is larger than 3J3Q in terms of the number of atoms in the overall assembly, but the raw data contains atomic coordinates for just a single asymmetric unit plus a large number of symmetry operators. ChimeraX opens and displays the complete Zika virus assembly with full symmetry in approximately 10 seconds as well. (Timing measurements done on a 2012 Apple MacBook with Intel graphics. These same structures cannot be displayed at all in our previous-generation program, Chimera.)

ChimeraX includes several major innovations in graphics visualization methods, developed by team member Tom Goddard. As noted above, interactive ambient occlusion lighting provides much more realistic lighting effects for models, and is especially beneficial when viewing very large molecules. (Figures 1 and 2. Also see letters from David Agard and Gabe Lander.) ChimeraX also incorporates "instancing," a memory-efficient method for rendering many copies of the same object in different locations and orientations. Instancing significantly reduces overall memory usage and improves graphics performance, allowing for large molecules to be manipulated interactively. Lastly, ChimeraX contains a completely new surface-generation algorithm, also implemented by Goddard, that is both much faster due to use of multiple processor cores (*i.e.,* multithreading) and completely avoids the numerical instabilities present in Chimera's old MSMS[30] surface-generation algorithm.

Over the past year we have implemented within ChimeraX the ability to visualize and analyze integrative hybrid (I/H) models archived using the mmCIF I/H data dictionary being developed by the Protein Databank[31,32]. Integrative modeling combines experimental inputs beyond the traditional X-ray crystallography, nuclear magnetic resonance, and electron microscopy, such as chemical crosslinks with linked residues identified by mass spectroscopy, Förster resonance energy transfer (FRET), small-angle X-ray scattering, and theoretical approaches such as homology modeling to compute models consistent with all input data[33]. The resulting models are typically ensembles, often combining atomic and coarse-grained elements, and may involve multiple and time-ordered states. Our effort is part of a three-way collaboration between the PDB, primarily responsible for defining the I/H dictionary, Andrej Sali's lab at UCSF, responsible for converting published integrative models into the dictionary format, and our lab, tasked with implementing visualization software to read, display and enable analysis of archived models (see letters from Stephen Burley and Andrej Sali).

Visualization of integrative data utilizes a vast array of ChimeraX capabilities. As an example, nup84, a seven-protein complex that is the primary architectural element of the outer ring of the nuclear pore, was modeled using two X-ray structures, five homology models based on 16 template structures, 300 chemical crosslinks utilizing two different linkers, 2D electron microscopy restraining the projected shape, and coarse grain multi-residue beads representing approximately 100 disordered segments[34] (Figure 3). All of the primary data including multiple sequence alignments and homology templates, as well as two resulting ensembles of 1000 mixed atomic and coarse-grained models characterizing two conformations, are presented, in addition to localization density maps to summarize ensemble variation. The hierarchical data capabilities of ChimeraX are used to organize the 40 different data files of this one study. Graphical tools allow inspection of violated crosslink restraints. New structures can be located by a sequence-search web service provided by the RBVI, BlastPDB[35], and aligned to the modeled complex, all done within ChimeraX to incorporate data acquired after the original model was published. Integrative hybrid modeling (IHM) represents an exciting area of recent development and takes advantage of ChimeraX's capabilities for handling many kinds of diverse data. For this reason, and also because our past and current work in this area is funded together with the rest of ChimeraX, we have described it in this proposal along with

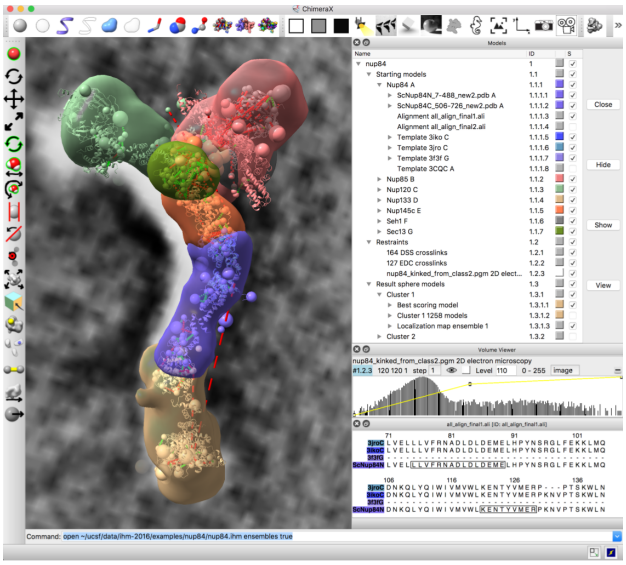

Figure 3: Nuclear pore subcomplex nup84 showing inputs and results, ensemble variability as surfaces, large beads for disordered regions, cross-links, electron microscopy, multiple sequence alignments and templates for homology models.

other preliminary results. However, going forward, we intend for IHM visualization and analysis to be funded separately, and it is the focus of a separate R01 application.

Performance, both efficient memory usage and fast response times, along with the clarity and aesthetics of model display in ChimeraX, indicate that our design and implementation decisions are sensible and, when complemented by the additional functionality we propose, will result in a software tool able to meet the demands of the biomedical research community for many years to come. "Daily builds" of ChimeraX have been available since November 2016, and we have made three alpha releases in 2017; since November 2016, there have been 4,700 downloads of ChimeraX from our website[36]. We expect to make the first beta release of ChimeraX before the end of 2017.

Research Plan:

To address the scientific challenges and achieve the vision outlined above, we have devised a set of specific aims to focus our development efforts and to achieve maximum impact on the extensive scientific community that depends on our software for successful advancement of their research programs.

Aim 1: Continue the design, implementation, documentation and dissemination of ChimeraX for the interactive visualization and analysis of atomic and cryo-electron microscopy data sets. In our vision for a multiscale modeling environment described above, ChimeraX provides the molecular visualization and analysis component. For its foundation, we have designed and implemented a high-performance core for reading, manipulating and rendering large data sets. These data sets may originate from experiments or computation and range from atomic-level structures to volume data and composite models. Sources of atomic-level structures include X-ray crystallography, high-resolution cryo-EM, molecular simulations and structure-prediction software. Volume data include electron density maps, micrographs from cryo-EM, and simulated maps from atomic-level models. Composite models include hierarchical multi-resolution models from IHM. While these data types may differ significantly in scale and source, the common thread is that they all have spatial aspects that are suitable for three-dimensional visualization. In addition, multiple data sets may represent the same biological system and therefore should be viewed and analyzed concurrently rather than independently.

Many of these data sets also have non-spatial aspects that are important for analysis. For example, IHM platforms such as IMP[23,37] include uncertainty measures for models, and the Protein Data Bank[27] associates results of stereochemical geometry checks, experimental data validation, and ligand validation with atomic structures. ChimeraX will include 2D plots and graphs that may be linked with the 3D visualization, so that user actions in one representation automatically trigger updates in the other (Figure 4). Using a combination of 2D and 3D visualization capabilities in ChimeraX, we will implement a variety of tools to enhance and integrate the display of disparate data.

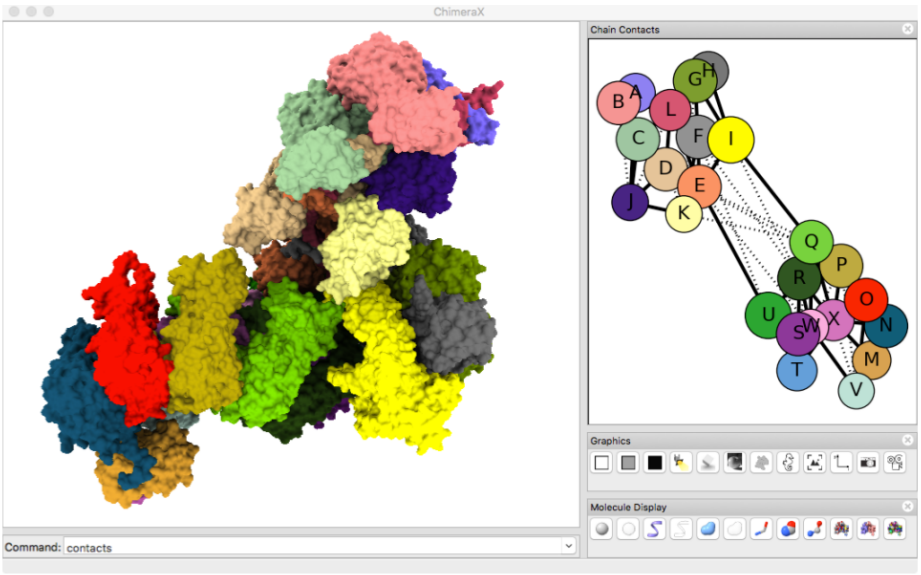

Figure 4: The ChimeraX graphics windows shows the CRISPR cascade complex (PDB 5CD4). The 2D connectivity network shows each chain in the assembly as a circle, with lines between chains in contact. Toolbars on the lower right provide shortcuts to common actions, e.g., changing lighting or model representation.

Finally, we will implement an export mechanism for publishing results on the Web. While ChimeraX will be extremely useful for interactive visualization and analysis, it is also a large application that users may not wish to install simply to look at a particular set of research results. To facilitate dissemination of results and hypotheses generated in ChimeraX, we will allow users to export structures, volumes, and composite models for interactive viewing using standard Web browsers. Specific

research plans in each of these areas is detailed in the sub-aims below.

Aim 1A: Further strengthen ChimeraX's ability to efficiently read, display and manipulate very large homogeneous and heterogeneous data sets. Data sets are getting larger. The largest Protein Data Bank entry with explicit coordinates for all atoms is 3J3Q, the structure of an entire HIV-1 capsid, with 2.4 million atoms (24 times what the original PDB format was designed to support). With high-resolution cryo-EM now a reality[2,38,39], atomic details of even larger molecular assemblies will be deposited in the wwPDB in the near future. To interactively process such data, ChimeraX must efficiently and effectively address three issues: memory usage, graphics rendering and computational speed. As described in the Preliminary Results section, we have successfully achieved impressive early results in these three areas already and intend to build on these results during the proposed period of support with the addition of new algorithms. ChimeraX currently serves as a testbed for algorithms and design choices.

ChimeraX is implemented using the Python programming language and a small amount of C++ (the latter where performance is critical). It efficiently reads and displays atomic structures containing millions of atoms and cryo-EM density maps measured in gigabytes. To achieve this level of performance, much of the molecular data is stored and manipulated using NumPy[40] arrays. We plan to design, implement and concisely document a set of Python-friendly Application Programmer Interfaces (APIs) for accessing these "internal" molecular data structures to support our goal of making the program a community development platform. These APIs will remain unchanged from release to release, thereby providing stable interfaces upon which additional functionality can be added, either by ourselves or by outside developers.

The 3D graphics rendering in ChimeraX has also been implemented from the ground up to take advantage of modern GPU hardware by using the latest version of OpenGL[41]. By using OpenGL and the OpenGL Shading Language (GLSL)[42,43], we have implemented graphics features such as instancing and photorealistic lighting in a platform-independent manner. (ChimeraX runs on MacOS, WIndows7/8/10, and Linux.) Our goal is to provide the best platform-independent graphics performance possible by taking advantage of the latest version of OpenGL (e.g., OpenGL 4.6 released in 2017) while "falling back" to OpenGL 3 when necessary in order to support older GPU hardware. Going forward, our performance goal is to render molecular graphics scenes at 100 frames/second for models of modest size, and thus support virtuality-reality headsets. (ChimeraX provides some support for VR headsets now[44].)

Aim 1B: Integrate two-dimensional diagrams with the three-dimensional graphics for linking non-spatial data with molecular visualization. Many relevant types of data are not spatial and therefore not optimally represented in three dimensions. These include model uncertainty estimates from IHM, phi-psi angles for protein structure validation, and energy fluctuations over time in molecular dynamics (MD). For such data, 2D diagrams such as Ramachandran plots, temporal plots, histograms, *etc.* are more appropriate. Even some inherently spatial information such as distances between residues may be more easily interpretable in 2D, for example, as a chain contact network (Figure 4 above). The 2D display is, however, most definitely tied to the 3D data. For example, in a 2D plot of energy *vs.* time for a MD simulation, each data point corresponds to a particular "frame" or snapshot in the 3D trajectory. Our vision of providing an *integrated* modeling environment means that when users interact with one type of data, the corresponding data in another view is automatically updated in the appropriate way based on the underlying data relationships; *i.e.,* clicking on a data point in the 2D energy *vs.* time plot should show the matching trajectory conformation in the 3D graphics window.

For ChimeraX, the goal of this sub-aim is to design and implement an application programmer interface (API) to which visualization and analysis tools can "subscribe" to and then send and receive events such as "new atom selected" or "different MD time step requested". The first use for this notification mechanism will be the molecular graphics component, which will send events when users alter what is selected. We will then implement notification-aware versions of 2D diagrams based on libraries such as Matplotlib[45] (a popular Python library commonly used to render 2D plots) and Cytoscape.js[46] (a Javascript script library for visualizing networks). The notification-aware 2D diagrams can then be used as building blocks for tools that need to display non-spatial data linked with molecular graphics.

Aim 1C: Implement novel analysis and modeling functionality and user interfaces, integrate existing web resources, and migrate selected Chimera tools to ChimeraX. One of the reasons that our older

Chimera application has been so successful within the scientific community is the broad suite of tools it offers covering a wide range of scientific needs, including structure superposition, H-bond detection, molecular docking, MD, cryo-EM and EM tomography. Using the same successful approaches as we have in the past, we will add tools to ChimeraX to address both current and new needs as determined by our collaborators and the scientific community. Our goal is for ChimeraX to serve as a component in an ecosystem that includes external services and interactive applications.

We will create new user interfaces (UI) to address issues arising from novel data types. For example, because hierarchical models may use different representations at different resolutions, users need a means to identify the components of the model to be displayed as well as the representations to use. Also, many experimentally determined structures in the wwPDB include symmetry or subassemblies (Figure 4), and an innovative UI is required to easily navigate the related visual representations and substructures. For instance, with a whole-virus model of HIV, we envision that users will be able to perform operations such as “hide the membrane”, “show only the capsid”, “add the RNA”, *etc.* because ChimeraX will extract the hierarchy information and labels, and present them to the user in an intuitive way that focuses on the virus itself rather than how the model is stored.

Several existing Chimera tools have been developed by outside collaborators and we plan to bring forward tools such as ISOLDE and GaudiMM[47]. (See letters from Tristan Croll, J.D. Marechal, Paul Adams, and the Richardsons.) An early version ISOLDE, which use the GPU-based OpenMM[45] package for its MD engine, already works with ChimeraX and is available in the preliminary implementation of the ChimeraX Toolshed.

We will also develop tools that utilize web-based resources to provide additional functionality. Heavily used web tools from Chimera such as homology modeling (MODELLER[48]), pocket identification (CASTp[49]), ligand docking (AutoDock Vina[50]), and sequence-based structure search (BLAST[12]) will also be modernized in ChimeraX to use the current most effective algorithms. Direct fetching of structures and annotations from web databases will be greatly extended (Chimera accesses 12 web databases; see Huang *et.al*, 2014[35]), providing a simplified procedure for users to add databases using URL patterns without having to write new code, and to contribute these capabilities through the ChimeraX Toolshed (see Aim 3).

In addition, we will create custom user interfaces for targeted web resources. For example, DOCK Blaster provides a public access service for structure-based ligand discovery and presents its results in a series of web pages. The results are predicted binding modes and are best visualized in an application such as ChimeraX, which can combine the 3D binding modes, 2D scoring plots, and tabular prediction data into a single integrated interface. We will develop graphical interfaces that facilitates both the invoking and analyzing the results from these types of web resources. We will also document the interface code so that community developers interested in other resources can straightforwardly incorporate them into ChimeraX (see Aim 3 below).

Finally, we will be migrating certain significant and highly used features in Chimera to ChimeraX. For example, MultAlign Viewer[51] displays sequence alignments alongside the molecular graphics, automatically linking structures with their sequences. We will add an enhanced version of MultAlign Viewer to ChimeraX. In addition to displaying sequence alignments in the standard matrix form, we will incorporate a ProfileGrids[52] -style layout for handling large alignments containing thousands of sequences.

Aim 1D: Create an export mechanism to facilitate disseminating results and hypotheses generated in ChimeraX via a WebGL-based viewer. ChimeraX is designed as a full-featured visualization and analysis application for desktop and laptop computers. This obviously means that users must first install it in order to use all those features. However, in some instances users may only need to access a small subset of features. For example, for viewing models created by other researchers, it may be sufficient to provide a simple interactive viewer and perhaps a few basic analysis tools.

Today, the most common way to disseminate graphically oriented results is to publish images or videos. However, this unnecessarily limits what scientists can see and their ability to interactively explore the data. Recently, a number of WebGL-based molecular viewers, *e.g.*, GLmol[53] and NGL[54], have been released that allow users to display molecules interactively in a Web browser. These viewers offer a path to make results generated in ChimeraX available to a wide audience. As a first step, we have recently added the capability to export ChimeraX scenes using the GL Transmission Format[55] for subsequent import into WebGL-based applications. As we continue this effort, we will identify a WebGL-based molecular viewer and further augment

ChimeraX to export data in a format optimized for this viewer. This prototype implementation may not include support for all of the data types in ChimeraX, but will export common types such as atomic structures and volumetric data surfaces. We will then either enhance the chosen viewer ourselves or, preferably, work with the viewer's authors to add support for formats and visual representations for more data types. For example, most current viewers focus on the display of models because WebGL provides the novel functionality of displaying interactive 3D graphics in a Web browser window. However, for dissemination and communication of scientific results, including other data such as sequence alignments, 2D plots (Ramachandran, energy *vs.* time, *etc.*) greatly enhances the information presented. We will include additional information through embedding the WebGL viewer in web pages so that narratives may be easily displayed with the 3D graphics, as well as using technologies such as Scalar Vector Graphics[56] (SVG) for integrating 2D data.

Aim 2: Within the ChimeraX framework, develop tools for atomic-resolution modeling from cryo-electron microscopy data sets. The goal of Aim 2 is to develop interactive visualization and analysis software for data from state-of-the-art cryo-EM instruments. This will enable researchers to analyze larger data sets containing the richer views of molecular and cellular structures made possible by recent leaps in instrumentation technology. Our software will help build and validate atomic models as well as coarser models of molecular environments within cells, and allow measurements and comparisons of models that unravel the dynamic behavior and functions of molecular machines and cellular processes. The tools we create are *interactive*, using real-time 3D visualization to reveal every aspect of data and models, in combination with fast calculations (typically taking less than a second) to do alignments, fitting, measurements, morphing, clash detection, and dozens of other tasks where human judgment is needed. These operations help researchers set up up long-running calculations such as model refinements, molecular docking, and simulations, identify and correct errors, and produce illustrations and animations for communicating new biological results. All of the proposed capabilities are built on the core ChimeraX visualization platform described in Aim 1 above.

Aim 2A: Develop analyses of map quality for model building. Calculate and show spatial variations in resolution, find domain motions, identify and quantify approximate symmetries, analyze difference maps, perform segmentation. New map analysis capabilities in ChimeraX will help achieve the highest resolutions from acquired micrographs and survey spatial variation in resolution. We describe a few specific targeted areas for development below, but many others are planned, such as finding approximate symmetries for averaging to improve resolution, analysis of difference maps, and segmentation capabilities.

Local resolution: Variations in resolution for different regions of cryo-EM maps are common and affect the detail of model building that will be possible in different portions of the map. The variation can be computed with the ResMap[57] program. We collaborated with ResMap developer Alp Kucukelbir to have ResMap output a Chimera script that shows colored map surfaces (Figure 5), and we plan to include ResMap directly in ChimeraX.

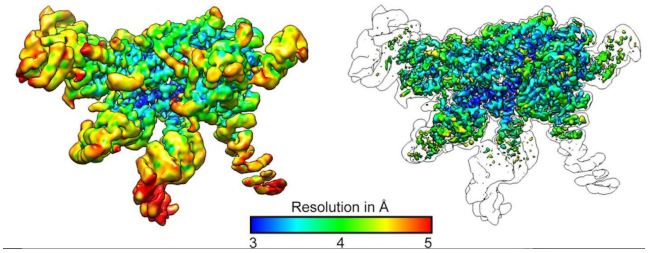

Figure 5: Proteasome lid colored by map resolution using ResMap shown at two threshold levels.

Interactive map sharpening: To enhance high-resolution map detail when fitting side-chain atoms in density, it is useful to apply B-factor sharpening, which boosts the high frequencies. This also increases the noise, and many small blobs of density appear as the map is sharpened. We will implement real-time sharpening of local regions by providing a slider control with immediate map updating to control detail and noise as the researcher examines different parts of the map.

Customizable bandpass Fourier filters: In transmission electron microscopy, different spatial wavelengths are amplified in a complicated pattern described by the contrast transfer function. We will add the ability to apply custom Fourier space filters in collaboration with experts such as Adam Frost (see letter of support) to evaluate, isolate and mitigate frequency-shell-dependent effects at very coarse (<4nm) and very fine resolutions.

Map projections: To assess the quality of single-particle cryo-EM maps, we will add the ability to project maps from three to two dimensions along specified directions. The results can then be compared to micrograph class-averaged 2D images to assess the consistency of the 3D reconstruction with acquired 2D images.

Aim 2B: Create interactive model building and measurement tools. Locate template structures from within the PDB, calculate domain and secondary structure motions for optimal fits, utilize *de novo* backbone tracing algorithms. Compare related models with real-time animation of conformational changes. Measure motions, distances, variation in contact areas, and create schematic 3D illustrations of measured quantities. At resolutions from 2 to 15 Å, different model-building methods are needed to avoid overfitting. At low resolutions, whole domains of X-ray and NMR structures are fit, at intermediate resolutions, helix and sheet positions can be adjusted, and at the highest resolutions, side-chains can be matched to density. Our proposed software development for this sub-aim will focus on the intermediate and high resolutions. Modeling procedures include identifying template structures from the Protein Data Bank, homology modeling to correct differences in sequence, calculating domain and secondary-structure motions for optimal fits, and, for regions lacking templates, utilizing *de novo* backbone-tracing algorithms. To study molecular function, we plan new and improved tools for real-time animation of conformational changes, measurement of motions, distances, and variations in contact area, and making schematic 3D illustrations of measured quantities. Below we describe a few of the planned developments.

Finding PDB templates: To build a model, we first find structures with similar sequences in the PDB using an RBVI-provided web service -- BlastPDB. We plan to provide a tool to find motions of domains and secondary structure present in available structures and rigidly fit these pieces to produce a starting model for high-resolution refinement. To locate related PDB structures having no discernible sequence similarity, we will use Phenix structure-based searching using marked helices and strands from the density.

Modeling building with Phenix, Rosetta, and XLink Analyzer: Atomic-resolution model refinement involves complex algorithms where visualization is part of the iterative cycle of refinement and correcting mistakes. We are collaborating with several labs creating algorithmic software to develop ChimeraX tools to set up refinements and permit interactive modeling. For high resolution (<3Å) we will interface to Phenix (see letter from Paul Adams), a widely used package for X-ray structure refinement, to optimize side-chain packing. At resolutions better than 6Å, secondary structure is easily resolved, and we will provide a ChimeraX interface to optimize the positions of secondary structure segments using the Rosetta[58] modeling package. The Agard, Cheng, and Frost cryo-EM labs at UCSF all recommend Rosetta for this optimization, but together rely on the expertise of only one postdoc that developed this Rosetta capability, because the calculation is difficult to set up. Our new tool will allow wider use of Rosetta for fitting at this resolution. We also plan to provide capabilities at lower resolutions such as using crosslink constraints in combination with cryo-EM map fitting in collaboration with Jan Kosinski at the Centre for Structural Systems Biology to model very large structures such as the nuclear pore[59].

Map scale calibration: Small inaccuracies in the physical scale of density maps can lead to distorted atomic models. A common solution when fitting independently solved atomic models into maps is to try altering map scaling in one-percent steps to determine which size scaling produces the best fit. We will add a tool to automate this tedious process, quickly testing and optimizing the fit over a finer range of scales, and plotting correlation versus scale factor.

Aim 2C: Devise new model validation tools, including atomic model validation, comparing models to maps, detecting atomic clashes and unfavorable geometries, manually correcting problems, and launching automated refinement methods. Highlight conformational and sequence differences in models derived from different experimental conditions or mutants. Cryo-EM resolutions down to 2Å have revealed structures that could not be crystallized, but these resolutions are low by X-ray standards, in a range where errors in side-chain packing and backbone path are common. We we will provide a range of tools to validate models, compare models to maps, detect atomic clashes and unfavorable geometries, and allow manually correcting problems and launching automated refinement programs. To investigate unique features of cryo-EM structures of assemblies, we will develop visualization tools to highlight conformational and sequence

differences and missing residues compared to X-ray and NMR structures of the component molecules. (See letter from Sriram Subramaniam.)

Visualizing density around residues: A thorough comparison of thousands of residues in an assembly with high-resolution density is time-consuming to carry out. We will streamline this visualization process by showing individual secondary structure segments together with density within a specified distance. The view direction will be chosen to minimize map overlaps, and moving a slider will smoothly transition from one segment to the next for rapid surveying of fit quality. We will also provide an inset view showing where the secondary structure element is in the overall structure to help the researcher account for variations in map resolution between the core and periphery of the entire molecular assembly. Color-coded sequences to highlight poor residue fits and bad geometry will be shown for quick navigation to problematic parts of structures.

Highlighting map zone boundaries: To assess and correct atomic models, we look at map meshes cropped to within some distance of side-chain atoms. To better show where nearby undisplayed density exists, we plan to highlight the boundaries of the cropped density by showing colored bands at density cuts. We've made a crude test of this idea (Figure 6), but finer quality will be obtained by adjusting the mesh near the boundary to make the colored bands thin and smooth so that they will not visually clutter the depiction of atoms.

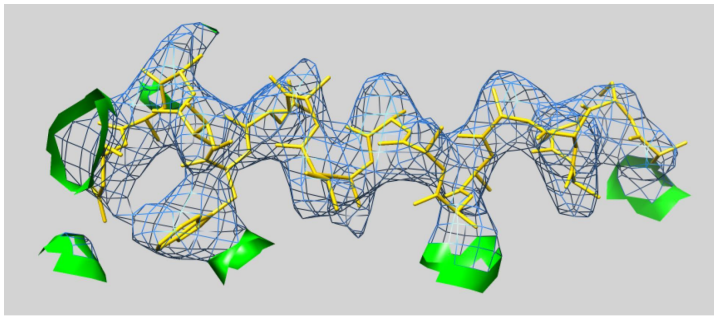

Figure 6: Highlighting density cuts with colored bands when showing a zone around atoms.

MolProbity structure validation: We will add a ChimeraX user interface to the MolProbity web service developed by collaborators Jane and David Richardson (see letter) that identifies clashes, strained geometry, poor hydrogen-bonding patterns, poor fit to density, and a broad range of other validation criteria. The interface will allow a researcher to step through problems and use simple tools to manually improve and then computationally refine (*e.g.*, with Phenix) atomic models.

<u>Aim 3:</u> Design, implement and document software interfaces to facilitate community development of new algorithms and methods by building upon the ChimeraX platform. Community development is a paradigm where access to a software platform is made publicly available so that novel applications leveraging the platform's functionality can be created more easily by independent developers. An example of a successful community development biocomputing project is Cytoscape[60,61], where contributors have written and released hundreds of plug-ins or "apps"[62]. Over the years Chimera has served as the development platform for several projects like this, including Xlink Analyzer[63], Molecular Dynamics Simulations, 3DIANA[64], and JalView[65]. These excellent tools were developed by talented programmers using the relatively sparse Chimera API documentation and examples; in addition, the tools are distributed directly by the research labs and must be verified to work with each new Chimera release.

Our vision for ChimeraX is to make community software development much easier and the resulting tools more widely available and less dependent on which version of ChimeraX was available at the time the code was written. (See letters from Yifan Cheng, Tristan Croll, Paul Adams, the Richardsons, Steve Ludtke, and Maya Topf.) Two critical ingredients for successful community development in any environment are a solid software platform and a suitable distribution channel. For ChimeraX, these translate into the need for a stable API to the program's underlying functionality and a mechanism to interactively install tools from a repository of available tools, what we call the ChimeraX Toolshed. The tools need not be part of the ChimeraX base distribution. In addition to the ChimeraX API, community developers will have access to standard, well-maintained, modern cross-platform libraries used to build ChimeraX, such as the graphical user interface toolkit (PyQt[66]), standard 3D graphics (OpenGL), and fast numerical algorithms (NumPy).

Aim 3A: Create and document semantically versioned application programming interfaces (API) for ChimeraX libraries. Semantic versioning is a software development convention where the compatibility between different versions of software releases may be deduced from their version numbers. We will release ChimeraX using version numbers with major and minor release numbers. Releases of ChimeraX with different major version numbers are not guaranteed to be compatible; *e.g.*, one of the APIs may change. Releases with

the same major version number but different minor version numbers are guaranteed to be compatible, that is, a tool implemented for the release with the older minor version number is guaranteed to work in the release with the newer minor version number. Semantic versioning is essentially a promise to developers that, once written, their software will continue to work until the next major release of ChimeraX. Compatibility is determined by the documentation of the ChimeraX API. Once an API is documented for a particular release, it is effectively a contract that the API will be available unchanged until the next major release. Major releases should be infrequent, because *additions* to the ChimeraX API are permissible with semantic versioning; a major release is only necessary when the existing API must be *altered* rather than extended.

The ChimeraX API will give programmatic access to the functionality discussed in Specific Aims 1 and 2, including molecular data management, graphics display, multi-threading and web services. The documentation will be made available both with each ChimeraX distribution and via the web. The semantically versioned ChimeraX API will provide a stable platform for the development community.

Aim 3B: Design and implement a mechanism for interactive installation of new tools and updates to existing tools in ChimeraX from an RBVI-provided repository accessible via the Web. One of the challenges of distributing software is ensuring that users have (or at least have access to) the "latest and greatest" release. However, users often don't update software unless they have a specific reason to, because the process is perceived as too disruptive or troublesome. To facilitate updates and make them less burdensome, ChimeraX will offer interactive installation of new software releases.

In addition, rather than releasing ChimeraX as a monolithic package containing all available tools, the standard distribution will contain only the most commonly used ones; the remainder will be placed in an RBVI-provided repository accessible from the Internet. The tools are packaged individually or as bundles to make them installable independently. New versions of standard tools may also be added to the repository. To install tools, the standard ChimeraX distribution will include a mechanism that communicates with the RBVI repository via a web service to obtain a list of new or updated tools. Users can then selectively install these, with new tools being available immediately and core updates available after restarting ChimeraX.

Aim 3C: Create a Web interface for external developers to deposit and distribute their tools using the RBVI repository. Streamlining the software update process not only simplifies delivering RBVI-developed tools, but also offers a means to distribute community- developed tools. The toolshed repository will initially be populated with tools written by us. However, it will also provide a convenient central location for distributing tools developed by the community. We will create a web interface customized for submitting ChimeraX tools for inclusion in the repository. To provide a level of verification and security, submitters will be required to create a password-protected account and have a valid email address. With a registered account, developers can submit new tools or updates to tools they previously placed in the repository.

Timeline: Most of the specific aims can be split roughly into two phases, Initial Implementation and Enhancements (see table below). Implementation provides the reference functionality that may be used as a foundation for other aims. Enhancements include adding functionality when new requirements arise (for example, when other specific aims are implemented), bug fixes, and changes necessitated by new versions of platform operating systems.

Potential Problems, Alternative Strategies, Benchmarks for Success:

Potential problems revolve around building a large user base, having adequate programmer resources for rapid development, incorporating the best available technologies, and maintaining the software beyond the period of funded development. We aim to convert as many as possible of our 15,000 registered (and many more unregistered) Chimera users to ChimeraX. Long-term users of complex software are notoriously reluctant to upgrade, but our early releases have shown that the numerous more powerful capabilities of ChimeraX compared to Chimera are compelling. The ChimeraX user interface and commands will be somewhat familiar to Chimera users, making the transition easier. Our extensive presentations of tutorials at workshops and development of online tutorial materials, now entirely focused on ChimeraX, will aid in recruiting existing Chimera users as well as graduate students new to molecular visualization. While most academic software projects are the work of a single graduate student, our software is developed by a team with deep experience in the problem domain, and we believe the scope of the work is appropriate given the human resources available. ChimeraX is based on the most widely used and stable components: the Python 3

Activity	Year 1	Year 2	Year 3	Year4
Specific Aim 1: Interactive visualization and analysis platform (ChimeraX)				
SA 1A: Support very large homogeneous and heterogeneous data sets				
SA 1B: Integrate visualization of non-spatial 2D data with 3D graphics				
SA 1C: Implement novel analysis and modeling tools				
SA 1D: Export ChimeraX data for web-based visualization				
Specific Aim 2: Atomic resolution modeling from cryo-electron microscopy datasets				
SA 2A: Develop analyses of map quality for model building				
SA 2B: Create interactive model building and measurement tools				
SA 2C: Devise new model validation tools				
Specific Aim 3: Community development platform for new algorithms and methods				
SA 3A: Create and document APIs for ChimeraX libraries				
SA 3B: Design and implement web-based ChimeraX bundle repository				
SA 3C: Develop bundle contribution method for community developers				

Color Key: Implementation Enhancements

programming language, Qt 5 window toolkit, and the latest GPU programming APIs of the OpenGL graphics library. We believe that these components will sustain a 15-year software lifetime (the current age Chimera). Community effort will be important for maintenance of the software beyond the funded development period, and a primary goal of the ChimeraX design (all of Aim 3) is to engage significant numbers of community developers.

Alternative strategies to deliver multiscale visualization to researchers include using web-based and cloud-based tools and mobile devices (phones, tablets). While many computing applications have migrated to these technologies, we believe that in-depth analysis and visualization still require the larger screens, high graphics performance, and fast access to large data attainable with desktop and laptop computers. As discussed, ChimeraX will leverage web and cloud computation. Separate from this proposal, ChimeraX is the leading virtual-reality molecular and cellular visualization tool, and new interactive VR technology demands the most powerful desktop graphics available, 50 times more powerful than current mobile devices.

We will use unambiguous quantitative measures of success for ChimeraX as we have for Chimera, which is currently cited about 200 times per month in research journals, with over 12,500 citations to date. Four hundred new Chimera users register per month, with cumulative unique registrations of 15,000, and 50 extensions to Chimera have been written by outside developers. These numbers reflect lower bounds, as many research articles do not cite the visualization software used, and many users do not register. Our registration system asks a user to register only after 15 distinct days of use, and there are no software limitations for failure to register. This gives a conservative measure of serious users, unlike typical raw download numbers, which grossly overestimate user counts by factors of 30 or more. We expect numbers of citations and registrations per month for ChimeraX to be at half this level when ChimeraX supersedes all Chimera functionality by the end of the proposed funding period, and beyond current rates after a few additional years, when users have migrated to the new software. Our current Chimera software has likely reached most researchers doing intensive molecular structure analysis. The stable programming APIs, Toolshed app store, and active support provided with ChimeraX are expected to enable hundreds of contributed tools from numerous research labs over the projected 15-year lifespan of the software, with an estimated 50 contributed tools during the proposed 4-year development period.

REFERENCES

1. O'Donoghue SI, Goodsell DS, Frangakis AS, Jossinet F, Laskowski RA, Nilges M, Saibil HR, Schafferhans A, Wade RC, Westhof E, Olson AJ. Visualization of macromolecular structures. *Nat Methods.* 2010 March; 7(3 Suppl): 42. [PubMed: 20195256] doi:10.1038/nmeth.1427.
2. Baker ML, Zhang J, Ludtke SJ, Chiu W. Cryo-EM of macromolecular assemblies at near-atomic resolution. *Nat Protoc.* 2010 September; 5(10): 1697-1708. [PubMed: 20885381] PMCID: PMC3107675 doi:10.1038/nprot.2010.126.
3. Baker ML, Hryc CF, Zhang Q, Wu W, Jakana J, Haase-Pettingell C, Afonine PV, Adams PD, King JA, Jiang W, Chiu W. Validated near-atomic resolution structure of bacteriophage epsilon15 derived from cryo-EM and modeling. *Proc Natl Acad Sci U S A.* 2013 July 23; 110(30): 12301-12306. [PubMed: 23840063] PMCID: PMC3725109 doi:10.1073/pnas.1309947110.
4. Sali A, Berman HM, Schwede T, Trewhella J, Kleywegt G, Burley SK, Markley J, Nakamura H, Adams P, Bonvin AM, Chiu W, Peraro MD, Di Maio F, Ferrin TE, Grunewald K, Gutmanas A, Henderson R, Hummer G, Iwasaki K, Johnson G, Lawson CL, Meiler J, Marti-Renom MA, Montelione GT, Nilges M, Nussinov R, Patwardhan A, Rappsilber J, Read RJ, Saibil H, Schroder GF, Schwieters CD, Seidel CA, Svergun D, Topf M, Ulrich EL, Velankar S, Westbrook JD. Outcome of the first wwPDB hybrid/integrative methods task force workshop. *Structure.* 2015 July 07; 23(7): 1156-1167. [PubMed: 26095030] PMCID: PMC4933300 doi:10.1016/j.str.2015.05.013.
5. Rose PW, Beran B, Bi C, Bluhm WF, Dimitropoulos D, Goodsell DS, Prlic A, Quesada M, Quinn GB, Westbrook JD, Young J, Yukich B, Zardecki C, Berman HM, Bourne PE. The RCSB protein data bank: Redesigned web site and web services. *Nucleic Acids Res.* 2011 January; 39(Database issue): 392. [PubMed: 21036868] PMCID: PMC3013649 doi:10.1093/nar/gkq1021.
6. McWilliam H, Valentin F, Goujon M, Li W, Narayanasamy M, Martin J, Miyar T, Lopez R. Web services at the european bioinformatics institute-2009. *Nucleic Acids Res.* 2009 July; 37(Web Server issue): 6. [PubMed: 19435877] PMCID: PMC2703973 doi:10.1093/nar/gkp302.
7. Bhagat J, Tanoh F, Nzuobontane E, Laurent T, Orlowski J, Roos M, Wolstencroft K, Aleksejevs S, Stevens R, Pettifer S, Lopez R, Goble CA. BioCatalogue: A universal catalogue of web services for the life sciences. *Nucleic Acids Res.* 2010 July; 38(Web Server issue): 689. [PubMed: 20484378] PMCID: PMC2896129 doi:10.1093/nar/gkq394.
8. Pieper U, Webb BM, Barkan DT, Schneidman-Duhovny D, Schlessinger A, Braberg H, Yang Z, Meng EC, Pettersen EF, Huang CC, Datta RS, Sampathkumar P, Madhusudhan MS, Sjolander K, Ferrin TE, Burley SK, Sali A. ModBase, a database of annotated comparative protein structure models, and associated resources. *Nucleic Acids Res.* 2011 January; 39(Database issue): 465. [PubMed: 21097780] PMCID: PMC3013688 doi:10.1093/nar/gkq1091.
9. Lawson CL, Patwardhan A, Baker ML, Hryc C, Garcia ES, Hudson BP, Lagerstedt I, Ludtke SJ, Pintilie G, Sala R, Westbrook JD, Berman HM, Kleywegt GJ, Chiu W. EMDataBank unified data resource for 3DEM. *Nucleic Acids Res.* 2016 January 4,; 44(D1): 396. [PubMed: 26578576] PMCID: PMC4702818 doi:10.1093/nar/gkv1126.
10. Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, Geer RC, He J, Gwadz M, Hurwitz DI, Lanczycki CJ, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C, Bryant SH. CDD: NCBI's conserved domain database. *Nucleic Acids Res.* 2015 January; 43(Database issue): 222. [PubMed: 25414356] PMCID: PMC4383992 doi:10.1093/nar/gku1221.
11. NCBI Resource Coordinators. Database resources of the national center for biotechnology information. *Nucleic Acids Res.* 2016 January 4; 44(D1): 7. [PubMed: 26615191] PMCID: PMC4702911 doi:10.1093/nar/gkv1290.
12. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990 October 5,; 215(3): 403-410. [PubMed: 2231712] doi:10.1016/S0022-2836(05)80360-2.
13. Irwin JJ, Shoichet BK, Mysinger MM, Huang N, Colizzi F, Wassam P, Cao Y. Automated docking screens: A feasibility study. *J Med Chem.* 2009 September 24; 52(18): 5712-5720. [PubMed: 19719084] PMCID: PMC2745826 doi:10.1021/jm9006966.
14. Chen VB, Arendall WB,3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. MolProbity: All-atom structure validation for macromolecular crystallography. *Acta*

Crystallogr D Biol Crystallogr. 2010 January; 66(Pt 1): 12-21. [PubMed: 20057044] PMCID: PMC2803126 doi:10.1107/S0907444909042073.
15. Waldrop MM. The chips are down for moore's law. *Nature.* 2016 February 11; 530(7589): 144-147. [PubMed: 26863965] doi: 10.1038/530144a
16. Muddukrishna A, Jonsson PA, Brorsson M. Characterizing task-based OpenMP programs. *PLoS One.* 2015 April 10,; 10(4): e0123545. [PubMed: 25860023] PMCID: PMC4393318 doi:10.1371/journal.pone.0123545.
17. Eastman P, Pande VS. OpenMM: A hardware independent framework for molecular simulations. *Comput Sci Eng.* 2015 July 1; 12(4): 34-39. [PubMed: 26146490] PMCID: PMC4486654 doi:10.1109/MCSE.2010.27.
18. Goddard TD, Huang CC, Meng EC, Pettersen EF, Couch GS, Morris JH, Ferrin TE. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci.* 2017 July 14 [PubMed: 28710774] PMCID: in-progress doi:10.1002/pro.3235.
19. Verba KA, Wang RY, Arakawa A, Liu Y, Shirouzu M, Yokoyama S, Agard DA. Atomic structure of Hsp90-Cdc37-Cdk4 reveals that Hsp90 traps and stabilizes an unfolded kinase. *Science.* 2016 June 24; 352(6293): 1542-1547. [PubMed: 27339980] PMCID: PMC5373496 doi:10.1126/science.aaf5023.
20. McCullough J, Clippinger AK, Talledge N, Skowyra ML, Saunders MG, Naismith TV, Colf LA, Afonine P, Arthur C, Sundquist WI, Hanson PI, Frost A. Structure and membrane remodeling activity of ESCRT-III helical polymers. *Science.* 2015 December 18; 350(6267): 1548-1551. [PubMed: 26634441] PMCID: PMC4684769 doi:10.1126/science.aad8305.
21. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH. PHENIX: A comprehensive python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr.* 2010 February 01; 66(Pt 2): 213-221. [PubMed: 20124702] PMCID: PMC2815670 doi:10.1107/S0907444909052925.
22. Leaver-Fay A, Tyka M, Lewis SM, Lange OF, Thompson J, Jacak R, Kaufman K, Renfrew PD, Smith CA, Sheffler W, Davis IW, Cooper S, Treuille A, Mandell DJ, Richter F, Ban YE, Fleishman SJ, Corn JE, Kim DE, Lyskov S, Berrondo M, Mentzer S, Popovic Z, Havranek JJ, Karanicolas J, Das R, Meiler J, Kortemme T, Gray JJ, Kuhlman B, Baker D, Bradley P. ROSETTA3: An object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol.* 2011; 487: 545-574. [PubMed: 21187238] PMCID: PMC4083816 doi:10.1016/B978-0-12-381270-4.00019-6.
23. Russel D, Lasker K, Webb B, Velazquez-Muriel J, Tjioe E, Schneidman-Duhovny D, Peterson B, Sali A. Putting the pieces together: Integrative modeling platform software for structure determination of macromolecular assemblies. *PLoS Biol.* 2012 January; 10(1): e1001244. [PubMed: 22272186] PMCID: PMC3260315 doi:10.1371/journal.pbio.1001244.
24. Croll TI, Smith BJ, Margetts MB, Whittaker J, Weiss MA, Ward CW, Lawrence MC. Higher-resolution structure of the human insulin receptor ectodomain: Multi-modal inclusion of the insert domain. *Structure.* 2016 March 01; 24(3): 469-476. [PubMed: 26853939] PMCID: PMC4860004 doi:10.1016/j.str.2015.12.014.
25. Banerjee S, Bartesaghi A, Merk A, Rao P, Bulfer SL, Yan Y, Green N, Mroczkowski B, Neitz RJ, Wipf P, Falconieri V, Deshaies RJ, Milne JL, Huryn D, Arkin M, Subramaniam S. 2.3 A resolution cryo-EM structure of human p97 and mechanism of allosteric inhibition. *Science.* 2016 February 19; 351(6275): 871-875. [PubMed: 26822609] doi:10.1126/science.aad7974.
26. Croll TI, Andersen GR. Re-evaluation of low-resolution crystal structures via interactive molecular-dynamics flexible fitting (iMDFF): A case study in complement C4. *Acta Crystallogr D Struct Biol.* 2016 September 01; 72(Pt 9): 1006-1016. [PubMed: 27599733] doi:10.1107/S2059798316012201.
27. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The protein data bank. *Nucleic Acids Res.* 2000 January 1,; 28(1): 235-242. [PubMed: 10592235] PMCID: PMC102472.
28. Westbrook JD, Fitzgerald PM. The PDB format, mmCIF, and other data formats. *Methods Biochem Anal.* 2003; 44: 161-179. [PubMed: 12647386].
29. Westbrook JD, Bourne PE. STAR/mmCIF: An ontology for macromolecular structure. *Bioinformatics.* 2000 February; 16(2): 159-168. [PubMed: 10842738].
30. Sanner MF, Olson AJ, Spehner JC. Reduced surface: An efficient way to compute molecular surfaces.

Biopolymers. 1996 March; 38(3): 305-320. [PubMed: 8906967].
31. Sali A, Berman HM, Schwede T, Trewhella J, Kleywegt G, Burley SK, Markley J, Nakamura H, Adams P, Bonvin AM, Chiu W, Peraro MD, Di Maio F, Ferrin TE, Grunewald K, Gutmanas A, Henderson R, Hummer G, Iwasaki K, Johnson G, Lawson CL, Meiler J, Marti-Renom MA, Montelione GT, Nilges M, Nussinov R, Patwardhan A, Rappsilber J, Read RJ, Saibil H, Schroder GF, Schwieters CD, Seidel CA, Svergun D, Topf M, Ulrich EL, Velankar S, Westbrook JD. Outcome of the first wwPDB hybrid/integrative methods task force workshop. *Structure.* 2015 July 07; 23(7): 1156-1167. [PubMed: 26095030] PMCID: PMC4933300 doi:10.1016/j.str.2015.05.013.
32. Burley SK, Kurisu G, Markley JL, Nakamura H, Velankar S, Berman HM, Sali A, Schwede T, Trewhella J. PDB-dev: A prototype system for depositing integrative/hybrid structural models. *Structure.* 2017 September 05; 25(9): 1317-1318. [PubMed: 28877501].
33. Ward AB, Sali A, Wilson IA. Biochemistry. integrative structural biology. *Science.* 2013 February 22,; 339(6122): 913-915. [PubMed: 23430643] PMCID: PMC3633482 doi:10.1126/science.1228565.
34. Shi Y, Fernandez-Martinez J, Tjioe E, Pellarin R, Kim SJ, Williams R, Schneidman-Duhovny D, Sali A, Rout MP, Chait BT. Structural characterization by cross-linking reveals the detailed architecture of a coatomer-related heptameric module from the nuclear pore complex. *Mol Cell Proteomics.* 2014 November 01; 13(11): 2927-2943. [PubMed: 25161197] PMCID: PMC4223482 doi:10.1074/mcp.M114.041673.
35. Huang CC, Meng EC, Morris JH, Pettersen EF, Ferrin TE. Enhancing UCSF chimera through web services. *Nucleic Acids Res.* 2014 July; 42(Web Server issue): 478. [PubMed: 24861624] PMCID: PMC4086125 doi:10.1093/nar/gku377.
36. ChimeraX homepage. Available from: http://www.rbvi.ucsf.edu/chimerax.
37. Lasker K, Velazquez-Muriel JA, Webb BM, Yang Z, Ferrin TE, Sali A. Macromolecular assembly structures by comparative modeling and electron microscopy. *Methods Mol Biol.* 2012; 857: 331-350. [PubMed: 22323229] doi:10.1007/978-1-61779-588-6_15.
38. Wang RY, Kudryashev M, Li X, Egelman EH, Basler M, Cheng Y, Baker D, DiMaio F. De novo protein structure determination from near-atomic-resolution cryo-EM maps. *Nat Methods.* 2015 April; 12(4): 335-338. [PubMed: 25707029] PMCID: PMC4435692 doi:10.1038/nmeth.3287.
39. Chen M, Baldwin PR, Ludtke SJ, Baker ML. De novo modeling in cryo-EM density maps with pathwalking. *J Struct Biol.* 2016 December 01; 196(3): 289-298. [PubMed: 27436409] PMCID: PMC5118137.
40. van der Walt S, Colbert SC, Varoquaux G. The NumPy array: A structure for efficient numerical computation. *CISE-M.* 2011; 13(2): 22-30. doi:10.1109/MCSE.2011.37.
41. Reina G, Muller T, Ertl T. Incorporating modern OpenGL into computer graphics education. *IEEE Comput Graph Appl.* 2014; 34(4): 16-21. [PubMed: 25051566] doi:10.1109/MCG.2014.69.
42. Bailey M. Using GPU shaders for visualization, part 3. *IEEE Comput Graph Appl.* 2013; 33(3): 5-11. [PubMed: 24807985] doi:10.1109/MCG.2013.49.
43. Chavent M, Levy B, Krone M, Bidmon K, Nomine JP, Ertl T, Baaden M. GPU-powered tools boost molecular visualization. *Brief Bioinform.* 2011 November; 12(6): 689-701. [PubMed: 21310717] doi:10.1093/bib/bbq089.
44. ChimeraX VR. Available from: http://www.rbvi.ucsf.edu/chimerax/docs/user/vr.html.
45. Hunter JD. Matplotlib: A 2D graphics environment. *CISE-M.* 2007; 9(3): 90-95. doi:10.1109/MCSE.2007.55.
46. Franz M, Lopes CT, Huck G, Dong Y, Sumer O, Bader GD. Cytoscape.js: A graph theory library for visualisation and analysis. *Bioinformatics.* 2016 January 15,; 32(2): 309-311. [PubMed: 26415722] PMCID: PMC4708103 doi:10.1093/bioinformatics/btv557.
47. Rodriguez-Guerra Pedregal J, Sciortino G, Guasp J, Municoy M, Marechal JD. GaudiMM: A modular multi-objective platform for molecular modeling. *J Comput Chem.* 2017 September 15; 38(24): 2118-2126. [PubMed: 28605037] PMCID: in-progress doi:10.1002/jcc.24847.
48. Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol.* 1993 December 5,; 234(3): 779-815. [PubMed: 8254673].
49. Dundas J, Ouyang Z, Tseng J, Binkowski A, Turpaz Y, Liang J. CASTp: Computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Res.* 2006 July 01; 34(Web Server issue): 116. [PubMed: 16844972] PMCID: PMC1538779.
50. Trott O, Olson AJ. AutoDock vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem.* 2010 January 30; 31(2): 455-461.

[PubMed: 19499576] PMCID: PMC3041641 doi:10.1002/jcc.21334.
51. Meng EC, Pettersen EF, Couch GS, Huang CC, Ferrin TE. Tools for integrated sequence-structure analysis with UCSF chimera. *BMC Bioinformatics.* 2006 July 12; 7: 339. [PubMed: 16836757] PMCID: PMC1570152.
52. Roca AI. ProfileGrids: A sequence alignment visualization paradigm that avoids the limitations of sequence logos. *BMC Proc.* 2014 August 28,; 8(Suppl 2 Proceedings of the 3rd Annual Symposium on Biologica): S6. eCollection 2014. [PubMed: 25237393] PMCID: PMC4155610 doi:10.1186/1753-6561-8-S2-S6.
53. GLmol: WebGL molecular viewer. Available from: https://www.glmol.com/.
54. Rose AS, Hildebrand PW. NGL viewer: A web application for molecular visualization. *Nucleic Acids Res.* 2015 July 1,; 43(W1): 576. [PubMed: 25925569] PMCID: PMC4489237 doi:10.1093/nar/gkv402.
55. Wikipedia entry for glTF (GL transmission format). Available from: https://en.wikipedia.org/wiki/GlTF.
56. Neumann A. Scalable vector graphics (SVG); 2008.
57. Swint-Kruse L, Brown CS. Resmap: Automated representation of macromolecular interfaces as two-dimensional networks. *Bioinformatics.* 2005 August 1; 21(15): 3327-3328. [PubMed: 15914544].
58. DiMaio F, Tyka MD, Baker ML, Chiu W, Baker D. Refinement of protein structures into low-resolution density maps using rosetta. *J Mol Biol.* 2009 September 11; 392(1): 181-190. [PubMed: 19596339] PMCID: PMC3899897 doi:10.1016/j.jmb.2009.07.008.
59. Kosinski J, Mosalaganti S, von Appen A, Teimer R, DiGuilio AL, Wan W, Bui KH, Hagen WJ, Briggs JA, Glavy JS, Hurt E, Beck M. Molecular architecture of the inner ring scaffold of the human nuclear pore complex. *Science.* 2016 April 15,; 352(6283): 363-365. [PubMed: 27081072] doi:10.1126/science.aaf0643.
60. Cline MS, Smoot M, Cerami E, Kuchinsky A, Landys N, Workman C, Christmas R, Avila-Campilo I, Creech M, Gross B, Hanspers K, Isserlin R, Kelley R, Killcoyne S, Lotia S, Maere S, Morris J, Ono K, Pavlovic V, Pico AR, Vailaya A, Wang PL, Adler A, Conklin BR, Hood L, Kuiper M, Sander C, Schmulevich I, Schwikowski B, Warner GJ, Ideker T, Bader GD. Integration of biological networks and gene expression data using cytoscape. *Nat Protoc.* 2007; 2(10): 2366-2382. [PubMed: 17947979] PMCID: PMC3685583.
61. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003 November; 13(11): 2498-2504. [PubMed: 14597658] PMCID: PMC403769 doi:10.1101/gr.1239303.
62. Lotia S, Montojo J, Dong Y, Bader GD, Pico AR. Cytoscape app store. *Bioinformatics.* 2013 May 15; 29(10): 1350-1351. [PubMed: 23595664] PMCID: PMC3654709 doi:10.1093/bioinformatics/btt138.
63. Kosinski J, von Appen A, Ori A, Karius K, Muller CW, Beck M. Xlink analyzer: Software for analysis and visualization of cross-linking data in the context of three-dimensional structures. *J Struct Biol.* 2015 March; 189(3): 177-183. [PubMed: 25661704] PMCID: PMC4359615 doi:10.1016/j.jsb.2015.01.014.
64. Segura J, Sanchez-Garcia R, Tabas-Madrid D, Cuenca-Alba J, Sorzano CO, Carazo JM. 3DIANA: 3D domain interaction analysis: A toolbox for quaternary structure modeling. *Biophys J.* 2016 February 23; 110(4): 766-775. [PubMed: 26772592] PMCID: PMC4775853 doi:10.1016/j.bpj.2015.11.3519.
65. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. Jalview version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics.* 2009 May 01; 25(9): 1189-1191. [PubMed: 19151095] PMCID: PMC2672624 doi:10.1093/bioinformatics/btp033.
66. Bouda P. PyQt und PySide. München: Open Source Press; 2012.