STATEMENT OF WORK

The goal of the overall project is to integrate the Phenix software suite, developed and licensed by Lawrence Berkeley National Laboratory (LBNL), and ChimeraX, developed by the Resource for Biocomputing, Visualization, and Informatics (RBVI) at the University of California, San Francisco (UCSF) and copyrighted and licensed by The Regents of the University of California (The Regents). Phenix is a popular software suite for the automated determination of molecular structures using X-ray crystallography and cryo-electron microscopy (cryo-EM) experimental methods, and ChimeraX is an advanced interactive 3D visualization and analysis software application that is also very popular within the same scientific research community. The two software suites are entirely complementary and integrating them using the robust software interfaces we propose to develop will provide for smooth exchange of data and information and thus will enable research scientists to create more efficient visualizations and innovative modifications of large cryo-EM datasets and models.

Specifically, the UCSF ChimeraX project team will collaborate closely with the LBNL Phenix team to design and implement Application Programming Interfaces (APIs) for exchanging data between Phenix and ChimeraX. The overall goal is to enable ChimeraX to run Phenix tools and retrieve output from these tools. Similarly, in the other direction, the API implemented in ChimeraX will enable Phenix to send modified structural models and cryo-EM density maps to ChimeraX for visualization. The work at UCSF will be carried out by current software engineering staff (Mr. Tom Goddard and Mr. Eric Pettersen) and does not require hiring new personnel. RBVI staff member Dr. Elaine Meng will be responsible for developing documentation and tutorials, and for providing user support via a virtual help desk for the integrated Phenix/ChimeraX software. Prof. Thomas Ferrin is co-investigator and will be responsible for the overall administration and direction of the UCSF component of the project.

UCSF tasks and deliverables

1) **Implement ChimeraX/Phenix communication APIs**. In collaboration with LBNL Phenix team members, UCSF will plan, develop, implement, debug, refine, and document intercommunication APIs between Phenix and ChimeraX.
2) **Implement fragment extension ChimeraX user interface.** Allows building atomic models in electron microscopy and X-ray maps using Phenix residue placement combined with ChimeraX manual refinement.
3) **Implement ligand placement ChimeraX user interface.** Allows placing small molecule ligands into electron microscopy and X-ray maps using Phenix algorithms and interactive ChimeraX refinement.
4) **Implement restraint specification ChimeraX user interface.** Allows a user to specify atomic model building constraints in ChimeraX that are then used by Phenix model building algorithms.
5) **Implement validation ChimeraX user interface.** Provides a user interface in ChimeraX for researchers to examine and fix atomic model problems identified by Phenix validation tools.

Timeline

This timeline for the UCSF ChimeraX developments is taken directly from the text of the awarded Phenix NIH R24 grant.

Specific Aim	Year 1	Year 2	Year 3	Year 4	Year 5
Improve program usability and integration	Implement REST interface in Phenix and ChimeraX (including ISOLDE)	Implement interface to fragment extension and ligand placement algorithms	Implement interface to refinement restraints and validation algorithms	Community testing and feedback on Phenix/ChimeraX and iSOLDE interfaces	Implementation of additional interfaces to Phenix algorithms based on community feedback

Deliverable Details

The following extracts from the awarded Phenix R24 grant describe the developments that will be undertaken by the UCSF ChimeraX team.

Fragment extension. In challenging regions of density maps, interactive building can model loops and other marginally resolved structures that are beyond the reach of current automated approaches. Phenix has excellent capabilities to propose fragment extensions at a specific location (Terwilliger 2003a; Terwilliger 2003b; Terwilliger et al. 2018). ChimeraX can display the alternative fragments and allow the user to choose, modify where necessary (e.g. via ISOLDE) and repeat. This would facilitate building large regions at a time, whereas current practice is limited to manually adding one residue at a time, making the task extremely slow and painstaking.

Ligand placement. Flexible ligands can be especially difficult to place correctly in moderate-resolution density, requiring subjective human analysis. Correct ligand modeling is critical to understanding the function of enzymes and optimizing drugs. Phenix has tools to flexibly dock ligands into density (Terwilliger et al. 2006; Terwilliger et al. 2007), but these are currently primarily used in a non-interactive manner and cannot easily take advantage of detailed prior knowledge about binding sites. We will make it possible for researchers to drag a box delimiting the potential binding site or click to specify a density blob using the existing ChimeraX map segmentation tools, and then call to Phenix to supply several alternative ligand placements to choose from, modify, and send back to Phenix for further refinement.

Restraints. Interactive specification of refinement restraints, e.g. symmetry, metal locations, protein-ligand covalent bonds, and other settings, e.g. definitions of rigid groups, is needed for automated structure determination. Currently in Phenix, for most of these parameters either Phenix defines them automatically (Moriarty et al. 2009), or the researcher specifies them manually. Automatic methods can often make mistakes in the case of a difficult or poor starting model at low resolution, where for example bad clashes can be turned into bonds inadvertently by automatic linking algorithms. Another example is tight symmetry restraints acting to minimize the real differences between related copies of a macromolecular chain. The ChimeraX/Phenix interface will allow intuitive assignment and direct checking of these restraints on the model. The ISOLDE

refinement suite in ChimeraX has a rich framework for user-driven addition and removal of positional, torsion, and distance restraints - all of which will be readily translated to a form readable by Phenix.

Validation. Phenix has extensive model and map validation capabilities (Afonine et al. 2010; Urzhumtsev et al. 2014; Moriarty et al. 2018; Afonine, Klaholz, et al. 2018; Williams et al. 2018; Prisant et al. 2020) that include quality metrics for identifying incorrectly modeled residues. Ideally, every problem area would be inspected and corrected interactively. ChimeraX can request validation information from Phenix and provide a simple interface to move from one problem area to the next, while using its rich visualization and editing capabilities to assess and resolve problems, request refinement in Phenix, and repeat, quickly iterating to produce the best achievable structure. Current methodology for checking structures containing thousands of residues is sufficiently onerous that most large structures deposited in the Protein Data Bank based on 3 Å or worse resolution data include many local mistakes, some trivial and some serious. These structures are often used as starting models for related structures, and the mistakes are propagated. For example, four recent SARS-CoV-2 RNA polymerase structures from different research labs (PDB IDs 7bv2, 7bv1, 7btf, 6m71) were all deposited with a helix incorrectly shifted by nine residues, a mistake inherited from older SARS-CoV structures (PDB IDs 6nur, 6nus) which had excellent traditional validation scores. That mistake, and others, were discovered by a combination of rebuilding with the ISOLDE program and new multi-residue validation criteria from Phenix/MolProbity and were subsequently re-versioned in the PDB by their depositors. We expect better validation and correction, especially at low resolution, to be one of the most important outcomes of the Phenix/ChimeraX integration.

Phenix and ChimeraX Communication Method

For each of the four examples above, Phenix and ChimeraX have well-developed existing capabilities that the resource will couple for more efficient and accurate structure determination. Enabling all of these improvements through integration will be a communication protocol that allows Phenix and ChimeraX to bidirectionally pass arguments and return values using the https protocol, with each program running independently. This communication method is described here, as it is central to the other developments and will require some effort to implement.

Phenix will implement an Application Programming Interface (API) based on Representational State Transfer (REST) for external software (Fielding 2000). The REST API is a standard architecture for defining how computers communicate with one another over a network. However, the same interface can be used for communicating between software programs on the same computer. This interface will enable ChimeraX to execute Phenix tools and retrieve the output. Similarly, in the other direction, a REST API implemented on ChimeraX and ISOLDE will enable Phenix to send modified models and maps for visualization. This bidirectional communication between ChimeraX and Phenix will improve the model-building process by letting scientists focus on working with their models instead of managing which files to load into which program and juggling multiple graphical user interfaces for running tools.