

**INTRODUCTION**

# Tools 2020: A compilation of tools for protein science

In recognition of the ever-increasing importance of databases, computational methods and experimental techniques, 2 years ago *Protein Science* published the special issue “Tools for Protein Science.” It has been exceptionally well received and has prompted us to publish its successor “Tools 2020.” In the future, we plan to offer such issues on a yearly basis.

## 1 | PROTEIN STRUCTURE

Given the extraordinary impact and ongoing value of the Protein Data Bank (PDB) it has to have pride of place in this issue (Goodsell et al. <https://doi.org/10.1002/pro.3730>). We should not forget that following the introduction of the PDB there were a number of well-known labs which strongly resisted the proposition that the coordinates and supporting data for all published structures be publicly released. Ever-increasing peer pressure from those who did deposit their data led to the expectation that “everyone needs to release their data.” Regrettably this is not the case in some other areas of science.

Of the 155,000 or so atom-level 3D structures of biomolecules in the PDB, about 90% were determined by X-ray crystallography. It is often stated that X-ray crystallography gives “static” information, although this must be qualified. Even before the structure of any protein was determined it was known that exposure of oxygen-free crystals of hemoglobin to air caused the crystals to crack. Thus, the binding of oxygen to hemoglobin was associated with some conformational change which was of sufficient strength and magnitude to overcome the protein–protein contacts which stabilized the crystal. Once the first protein crystal structures were determined it was apparent from the electron density that side-chains on the surface of the protein were often much more mobile than those in the core, and so on.

The wealth of structural information in the PDB has made it possible to better define the stereochemistry of biomolecules. This, in turn, has provided guidelines against which newly determined structures can be evaluated. Here, Jane and David Richardson and coworkers (Prisant

et al. <https://doi.org/10.1002/pro.3786>) describe the latest developments in the MolProbity web service for model validation. Their report is directed especially toward needs in cryo-electron microscopy.

Also, Gert Vriend and colleagues (Lange et al. <https://doi.org/10.1002/pro.3788>) describe a wide variety of facilities which support and exploit the use of PDB-derived structural data in bioinformatics studies.

The DALI web server was introduced by Liisa Holm and Chris Sander 25 years ago and has been an invaluable resource ever since. Here Dr. Holm (<https://doi.org/10.1002/pro.3749>) updates DALI and gives insights into the extraordinary insights which it can provide. With the benefit of experience she also contrasts what DALI can and cannot do.

ConSurf is a bioinformatics tool for accurately estimating the evolutionary rate of each protein in a protein family. Ben-Tal and coworkers (<https://doi.org/10.1002/pro.3779>) show that these evolutionary relationships can reveal structural and functional importance of specific positions.

## 2 | STRUCTURE VISUALIZATION

The visualization of biomolecular structures and the communication of biological insights play an ever-increasing role in protein science.

PyMOL is one of the most popular molecular graphics programs for making publication-quality images of macromolecular structures. Here, Blaine Mooers (<https://doi.org/10.1002/pro.3781>) describes a wide variety of Python functions which are “shortcuts” that not only save time but also, for example, generate molecular representations not available in PyMOL.

Traditional approaches to teaching the principles of protein structure have included two-dimensional (stereo) images on paper, physical models, and various video approaches. To provide a more realistic and engaging teaching environment, Jane Allison and coworkers (<https://doi.org/10.1002/pro.3752>) have developed a virtual reality application, Peppy. It provides a novel, dynamic, and fun tool that allows exploration of the basics of protein structure.

Visualization of complex biomolecular electrostatic properties can be challenging for two-dimensional graphics systems. To help address this problem, Nathan Baker and colleagues (<https://doi.org/10.1002/pro.3773>) have modified the Adaptive Poisson-Boltzmann Solver software to visualize and compare electrostatic information.

### 3 | SEQUENCE ANALYSIS

The rapid expansion of structural data in the PDB has, if anything, been dwarfed by the explosion of sequence information. KEGG is an extraordinarily powerful reference knowledge base which facilitates the interpretation of genome sequences. Here, Minoru Kanehisa and Yoko Sato (<https://doi.org/10.1002/pro.3711>) describe the use of KEGG Mapper in the interpretation of cellular functions and other high-level features.

DRSASP, the Dundee Resource for Sequence Analysis and Structure Prediction, has been developed by Geoffrey Barton and colleagues (<https://doi.org/10.1002/pro.3783>). Its uses include secondary structure and solvent accessibility prediction, disorder prediction, and specificity determining site prediction, among others.

Lukasz Kurgan and coworkers (<https://doi.org/10.1002/pro.3756>) have developed DISOselect which combines information from a number of different procedures to optimize the use of sequence information in predicting regions of disorder. Also, Birthe Kragelund's group (<https://doi.org/10.1002/pro.3754>) has developed IDDomainSpotter which uses compositional sequence bias to assess and visualize domain organization in long intrinsically disordered regions.

NAGbinder, introduced by Gajendra Raghava and colleagues (<https://doi.org/10.1002/pro.3761>) uses primary sequence alone to predict *N*-acetylglucosamine binding sites in proteins of interest.

### 4 | STRUCTURE DETERMINATION

The past 10 years could well be described as the decade of cryo-EM. There are clear parallels between information obtained by cryo-EM and that obtained by X-ray crystallography. But there are also distinct differences. In turn, tools that have been extraordinarily successful in the context of X-ray crystallography are being adapted for use in cryo-EM. Here, Tom Terwilliger's group (<https://doi.org/10.1002/pro.3740>) describe the use of iterative map segmentation in cryo-EM map interpretation. Wladek Minor's group (<https://doi.org/10.1002/pro.3747>) have developed Molstack for the presentation and interpretation of electron density and cryo-EM maps. HEMNMA, developed by Slavica Jonic (<https://doi.org/10.1002/pro.3772>) computes

the normal modes of an atomic structure or a density map and uses these to extract information on conformational variability.

Small-angle scattering of X-rays is an important tool to study biological macromolecules in solution. To support such studies, Svergun et al. (<https://doi.org/10.1002/pro.3731>) have established SASBDB, a comprehensive repository of SAS experimental data with over 1,000 entries.

Also in the context of structural studies in solution, Marius Clore's group (<https://doi.org/10.1002/pro.3745>) has developed a three-dimensional potential of mean force to improve backbone and side-chain hydrogen bond geometry in Xplor-NIH protein structure determination. As well, Elizabeth Meiering et al. (<https://doi.org/10.1002/pro.3785>) have developed a computational procedure including automated tracking that propagates initial (single temperature) NMR cross peak assignments to spectra collected over a range of temperatures.

### 5 | MUTANT PROTEINS

The ever-increasing wealth of clinical, genetic, and structural information holds the promise of understanding and addressing disease states that have heretofore been poorly understood. Roman Laskowski, Christine Orengo, Janet Thornton, and their team describe VarSite (<https://doi.org/10.1002/pro.3746>), a web server which maps known disease-associated variants, together with natural variants, onto known structures from the PDB. The associated information and visualization can help differentiate between pathogenic and benign variants. In a related report, Arun Pandurangan and Tom Blundell (<https://doi.org/10.1002/pro.3774>) discuss the prediction of the impact of mutations on protein structure and interactions.

One technique to experimentally analyze protein stability and interactions is differential scanning fluorimetry. Here Changye Sun et al. (<https://doi.org/10.1002/pro.3703>) describe SimpleDSFviewer which is easy to use and amenable to high-throughput applications.

### 6 | PROTEIN-PROTEIN AND PROTEIN-LIGAND INTERACTION

Protein-protein and protein-ligand interaction have central importance in all aspects of protein science from basic biochemistry to disease treatment. There are contributions in this special issue that span the gamut. Tony Kossiakoff and coworkers (<https://doi.org/10.1002/pro.3751>) introduce a novel experimental approach in which an engineered Fab scaffold and a domain of an

immunoglobulin binding protein, protein G, can be linked to produce multivalent and bi-specific entities. It is shown, for example, that such bi-specific modules can induce cell killing by crosslinking T-cells to cancer cells.

On the computational side, Sebastian Kmiecik (<https://doi.org/10.1002/pro.3771>) discusses the flexible docking of peptides to proteins using CABS-dock. Also, Ivet Bahar's group (<https://doi.org/10.1002/pro.3732>) introduce Pharmed for pharmacophore modeling and hit identification based on druggability simulations. Chris Bahl and colleagues (<https://doi.org/10.1002/pro.3721>) discuss modification to the Rosetta macromolecular modeling suite which enable users who have limited computational infrastructure to perform state-of-the-art molecular simulation and design with Rosetta.

To help improve molecular docking techniques, Carlos Camacho and associates (<https://doi.org/10.1002/pro.3784>) have developed DISCO, a directory of structures for cross docking and a benchmark for automated pose and ranking prediction of ligand binding.

There has been longstanding interest in the antimicrobial peptides. Here, Guangshun Wang (<https://doi.org/10.1002/pro.3702>) describes the Antimicrobial

Peptide Database initially online in 2003. The focus is on the peptide design parameters for each of the four unified classes of antimicrobial peptides. Also, Faiza Hanif Waghlu and Susan Idicula-Thomas (<https://doi.org/10.1002/pro.3714>) describe the CAMP, CAMPSign, and ClassAMP resources that provide comprehensive information on antimicrobial peptides and machine learning-based predictive models.

## 7 | CONCLUSION

This issue makes clear the ever-increasing strength and relevance of all aspects of protein science. It also is a testament to the seminal contributions of the participating authors and their colleagues.

**How to cite this article:** Tools 2020: A compilation of tools for protein science. *Protein Science*. 2020;29:5–7. <https://doi.org/10.1002/pro.3799>