The blobulator: a toolkit for identification and visual exploration of hydrophobic modularity in protein sequences

Connor Pitman[1], Ezry Santiago-McRae[1], Ruchi Lohia[2], Ryan Lamb[1], Kaitlin Bassi[1], Lindsey Riggs[1], Thomas T. Joseph[3], Matthew E.B. Hansen[4], and Grace Brannigan[1,5,*]

[1]**Center for Computational and Integrative Biology, Rutgers University–Camden, 201 Broadway, 08103, NJ, USA**
[2]**Department of Cancer Biology, Perelman School of Medicine, University of Pennsylvania, 3400 Civic Center Blvd, 19104, PA, USA**
[3]**Department of Anesthesiology and Critical Care, Perelman School of Medicine, University of Pennsylvania, JMB 305, 3620 Hamilton Walk, 19104, PA, USA**
[4]**Department of Genetics, Perelman School of Medicine, University of Pennsylvania, 3400 Civic Center Blvd, 19104, PA, USA**
[5]**Department of Physics, Rutgers University–Camden, 201 Broadway, 08103, NJ, USA**

ABSTRACT

While contiguous subsequences of hydrophobic residues are essential to protein structure and function, as in the hydrophobic core and transmembrane regions, there are no current bioinformatic tools for module identification focused on hydrophobicity. To fill this gap, we created the *blobulator* toolkit for detecting, visualizing, and characterizing hydrophobic modules in protein sequences. This toolkit uses our previously developed algorithm, blobulation, which was critical in both interpreting intra-protein contacts in a series of intrinsically disordered protein simulations [1] and defining the "local context" around disease-associated mutations across the human proteome[2]. The *blobulator* toolkit provides accessible, interactive, and scalable implementations of blobulation. These are available via a webtool, a VMD plugin, and a command line interface. We highlight use cases for visualization, interaction analysis, and modular annotation through three example applications: a globular protein, two orthologous membrane proteins, and an IDP. The *blobulator* webtool can be found at `www.blobulator.branniganlab.org`, and the source code with pip installable command line tool, as well as the VMD plugin with installation instructions, can be found on GitHub at `www.GitHub.com/BranniganLab/blobulator`.

Keywords: sequence-structure relationship, single-nucleotide polymorphism, computational methods, hydrophobicity, protein sequence organization, intrinsically disordered proteins

1 INTRODUCTION

Protein sequences are modular: they consist of contiguous units (such as alpha helices, functional domains, etc.) which are incorporated into a range of analysis pipelines, visualization tools, and conceptual frameworks. Few tools for detecting modulation, however, explicitly incorporate residue hydrophobicity, despite the cooperative nature of the hydrophobic effect; the critical role of hydrophobicity in stabilizing the core of globular proteins; and the commonplace use of residue hydrophobicity in predictors of protein disorder [3–8] and membrane interactions [9, 10]. We recently introduced "blobulation" as a scheme for segmenting protein sequences by contiguous hydrophobicity, and have demonstrated its use in diverse applications such as analyzing simulations [1] of intrinsically disordered proteins (IDPs) and identifying the local sequence context around single-nucleotide polymorphisms (SNPs) [2]. Here we introduce a toolkit designed to make blobulation accessible and convenient for a wide range of users.

Blobulation detects modularity in protein sequences by searching for "h-blobs": subsequences longer than a threshold length in which every residue is more hydrophobic than a threshold hydrophobicity. The remaining subsequences are sorted into "p-blobs" (which satisfy the length, but not the hydrophobicity criterion) and "s-blobs" (which satisfy neither length nor hydrophobicity criteria). While secondary structure detectors typically use a default set of parameters that users rarely change, blobulation settings are meant to be tuned: adjusting the hydropathy and length thresholds in blobulation allows users to gradually shift from detecting many small modules to a few longer ones, bringing the relevant aspects of sequence organization into focus (use-case examples shown in section 4). Furthermore, since blobulation requires only the sequence, it is particularly

Table 1: Features of the blobulator toolkit compared to other available tools for characterizing protein elements by secondary structure or hydropathy. Rows vary by the elements characterized or identified, and columns vary by the type of analysis or visualization. Each cell gives example of tools available for this purpose as well as the target users of the most accessible tool. Features that can be accessed via a GUI are considered accessible to end-users, while features that require significant scripting are noted as requiring power users.

	Secondary Structure			**Hydropathy**		
	Prediction	**Sequence View**	**Structure View**	**Prediction**	**Sequence View**	**Structure View**
Residue Characterization[a]	end user (e.g. PSIPRED,[11] SCRATCH,[12] AlphaFold[13])	end user (e.g. PROTEUS2,[14] PSIPRED,[15] JPred4,[16] NetSurfP,[17] ALIGNSEC[18])	end user (e.g. VMD,[19] UCSF Chimera,[20] POLYVIEW[21])	end user (e.g. EMBL-EBI[22])	end user (e.g. BioJava,[23] Expasy,[24] MPEx,[25] EMBOSS[26])	end user (e.g. VMD,[19] UCSF Chimera,[20] UGENE[27])
Module Identification [b]				power user (**blobulator** CLI)	end user (**blobulator** webtool)	end user (**blobulator** VMD)
Module Characterization[c]	power user (e.g. PSIPRED,[11] SCRATCH,[12] AlphaFold[13])	(not widely available)	power user (e.g. VMD,[19] UCSF Chimera[20])			

[a] predicts and/or visualizes secondary structure or hydropathy by residue
[b] predicts and/or visualizes stretches of contiguous secondary structure or hydropathy
[c] predicts and/or visualizes physicochemical or other properties aggregated over the topological element

appropriate in scenarios where circumventing structure is desirable (including IDPs).

Many existing tools cluster residues into protein modules using secondary structure. While blobulation yields modules analogous to secondary structure elements, we have not previously provided "blob" versions of tools that traditionally rely on secondary structure. On the left half of Table 1, we list the common use-cases for secondary structure identification among currently available tools: residue characterization (secondary structure of each residue), module identification (determining where an alpha helix, beta sheet, or other secondary structure element begins and ends), and module characterization (calculating collective properties of the secondary structure element, for instance, the net charge per residue within a helix or beta sheet). The columns show the properties used to define modularity (secondary structure or hydropathy), and then are subdivided by the functionality of the tools: whether the tool makes modularity predictions (usually via a command-line interface), or provides a graphical interface (via a sequence or structure view). For secondary structure tools, module prediction can be accessed through command-line tools [11–13], graphically displayed alongside the sequence [14–18], or incorporated into visualizations of the structure [19–21]. As shown on the right half of Table 1, tools for characterization [22], annotation [23–26], or coloring by the hydrophobicity of individual residues [19, 20, 27] are well-established. Yet, few tools incorporate "modular" definitions based on contiguous hydrophobicity, and we are unaware of any that allow graphical exploration or module characterization. Plots of residue hydrophobicity may invite the viewer to identify contiguous hydrophobic regions by eye, but do not provide a systematic, automatable version of the same process.

To provide a previously unavailable analogue to secondary structure tools that defines and characterizes modules based on hydropathy (bottom right of Table 1), we developed the *blobulator* toolkit: an accessible, interactive, and scalable suite of tools that includes a webtool, command line interface (CLI), and a Visual Molecular Dynamics (VMD) [19] plugin. The webtool supports a sequence-level view of blobs that includes smooth real-time parameter adjustment, a feature to introduce mutations, and layers of annotations including physicochemical properties and disease-associated mutations. The backend program for the webtool is available as a pip-installable CLI and supports high-throughput batch blobulation of FASTA files. Finally, the VMD plugin allows users to view blobs on protein structures by creating a blobulator interface in VMD, which stores identified blobs for analysis across simulation trajectories and the generation of publication-quality images and movies. As a whole, the *blobulator* toolkit provides previously unavailable functionality for detecting modules defined by hydrophobicity through both prediction and visualization of blobs and their characteristics in protein sequence and structure. Additionally, it goes beyond

secondary structure analogues, enabling end users to access module characterization through graphical user interfaces.

In this paper, we outline the blobulation algorithm and each component of the *blobulator* toolkit. We then demonstrate its utility in three example applications: detecting blobs corresponding to tertiary interactions in a globular protein; comparing blobulated sequences from a structurally conserved membrane protein family; and quantifying the effects of disease-associated mutations in an IDP. We close by describing future applications for these tools, including novel insights that the *blobulator* is positioned to enable.

2 BLOBULATION

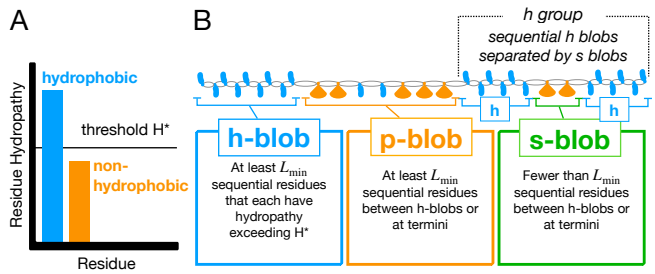

Figure 1. Blobulation algorithm. Figure adapted from Lohia et al. 2022 [2]. A) First, the sequence is digitized. Residues are classified as either hydrophobic (blue) or non-hydrophobic (orange) by comparing their hydropathy to the user-selected threshold, H^*. B) The sequence is then segmented into h-blobs, s-blobs, and p-blobs based on H^* and L_{min}.

Whole-sequence blobulation consists of two steps: digitization and clustering. We provide a description of the algorithm below, as well as a pseudocode version of the algorithm in the Supporting Information.

1. In the digitization step, the algorithm defines a given residue as hydrophobic or non-hydrophobic as follows:
   (a) The user selects a normalized (0 to 1) hydrophobicity scale, and sets a threshold hydrophobicity (H^*) on this scale.
   (b) The hydropathy for each residue (i) is assigned based on the normalized hydrophobicity scale. The hydropathies are then smoothed for each residue i and the residues adjacent to it in sequence ($i-1$, i, and $i+1$) , yielding the smoothed hydropathy H_i.
   (c) Residue i is classified as hydrophobic ($H_i > H^*$) or non-hydrophobic ($H_i \leq H^*$), shown in Fig. 1A.
2. In the clustering step, subsequences termed "blobs" are defined as follows:
   (a) The user sets L_{min} equal to a positive integer representing a threshold number of residues required to form an h- or p-blob.
   (b) Subsequences of at least L_{min} sequential hydrophobic residues are classified as hydrophobic blobs (h-blobs) (Fig. 1B).
   (c) All other linking sequences are then classified based on their L_{min} as either non-hydrophobic blobs (p-blobs, $L \geq L_{\text{min}}$) or short blobs (s-blobs, $L < L_{\text{min}}$) (Fig. 1B).

Though there are default settings for the initial blobulation, the user can tune various parameters. Different combinations of H^* and L_{min} will detect blobs with varying properties, and adjusting them can extract blobs that correspond to key hydrophobic regions (some examples are shown in section 4). Alternatively, users may reveal hierarchical layers of organization by blobulating a given sequence under varying parameter settings. Very high thresholds (H^* approaching 1) will return one large p-blob, and very low thresholds (H^* approaching 0) will return one large h-blob, while intermediate thresholds will reveal intrinsic segmentation within the protein sequence. Additionally, high L_{min} will only yield h-blobs if used with a low to moderate H^*.

We will also consider higher-order organization beyond individual blobs: "blob groups" are h-blobs separated only by s-blobs. Examples and further discussion of blob groups can be found in sections 4.1 and 4.3. Blobs are labeled as follows: by their type (h, p, s), group number (1, 2, 3), and for h-blobs within a blob group, by a sub-group letter (a, b, c). This forms a unique label for every blob (h1a, s1, h1b, etc). We refer to the distribution of blobs in a protein sequence as its "blob topology".

3 THE BLOBULATOR TOOLKIT

Here, we present three tools by which an amino acid sequence can be blobulated: a webtool, a command line interface (CLI), and a Visual Molecular Dynamics (VMD) plugin. The webtool can be used to interactively explore blobulated sequences, introduce mutations, and investigate various blob properties. The command line interface is the backend for the webtool, but also can be used to run batch processes and accepts DNA sequences. The VMD plugin can be used with molecular structures for analysis and creating images and movies of blobulated proteins in VMD. All images in this paper reflect the output of the *blobulator* version 1.0b.

3.1 Webtool

Blobulation of any amino acid sequence can be achieved using the *blobulator* webtool found at `www.blobulator.branniganlab.org`. On the "New Query" tab of the homepage, the user can submit either a database ID or manually enter a protein sequence for blobulation. From here, a user can blobulate any protein or amino acid sequence using an accepted ID or manually entering the sequence directly, as well as toggle to other information tabs. Figure 2 illustrates the blobulation of insulin, a peptide hormone that promotes glucose absorption by cells.

After the initial blobulation, the user can interactively tune some of the parameters. For example, while our previous studies [1, 2] used the Kyte-Doolittle scale [30], the user may select the Eisenberg-Weiss scale [31] or the Moon-Fleming scale [32]. Additionally, the user can change the hydrophobicity scale used for digitization (Fig. 2A) or interactively modify the H^* threshold and $L_{\min}$ cutoff (here we use $H^*= 0.4$ and $L_{\min}= 4$, respectively) to segment the protein into modules with varying hydrophobicity and length properties. This can be done by adjusting the respective sliders or manually entering a value (Fig. 2B and C). A user may also introduce a mutation by selecting one of the black triangles or manually entering the position and alternate amino acid in the "mutate residue" field and selecting the checkbox (Fig. 2D). All changes update dynamically. Users can download both an image of the output as a PDF and the raw data (in CSV format) used to generate the webtool output (Fig. 2E).

The smoothed hydropathy of each residue H_i (as defined in section 2) is shown in the first "Results" track (Fig. 2F) along with the H^* threshold (blue line). While a graph displaying the hydrophobicity of individual residues is common to many tools, blobulation clusters adjacent residues by their hydrophobicity, identifying blobs. The next track shows the output of the clustering step (Fig. 2G), where blobs are colored by their blob type. Blobulation of insulin using the settings $H^*= 0.4$ and $L_{\min}= 4$ identifies 8 h-blobs, 5 p-blobs, and 3 s-blobs. If a UniProt ID is provided, known disease-associated mutations are obtained from UniProt and displayed on every track as black triangles (Fig. 2). Hovering over these with the cursor will display the effect of the mutation on the amino acid, its reference SNP cluster ID (rsid), and link to the Single Nucleotide Polymorphism Database (dbSNP).

The webtool includes a series of tracks in which blobs are colored by additional information relating to charge composition, hydrophobicity, length, or some combination of the three. As shown in Figure 2 H-L, these include tracks displaying blob charge, globular tendency, the predicted propensity of each blob to deleterious mutations, and two tracks showing order and disorder predictions for each blob, respectively.

The first two of these tracks are colored by mean net charge per residue and globular tendency, both calculated based solely on charged residues in the blob. (Fig. 2H and 2I). In the mean net charge per residue track (Fig. 2H), blobs are colored by the average charge of all the residues within each blob. The use of this track is demonstrated in the example in section 4.2. In the globular tendency track (Fig. 2I), blobs are colored based on the region they occupy in the Das-Pappu diagram. Traditionally, the Das-Pappu phase is the globular tendency based on the charge content of a whole protein sequence [28], but here we have calculated the Das-Pappu phase of each blob. The use of this track is demonstrated in section 4.1.

The predicted mutation sensitivity track is colored based on an enrichment calculation that includes both blob hydrophobicity and length (Fig. 2J). This track colors blobs based on the enrichment of disease variants in blobs with the same properties found in our previous study [2]. Enrichments were calculated from a dataset containing about 70,000 human disease-associated and non-disease-associated missense mutations.

While many tools to predict disorder focus on predicting intrinsically disordered regions, or whether a whole protein is ordered or disordered, we included two tracks that estimate the intrinsic disorder of each blob (Fig. 2K and 2L). The first of these tracks colors blobs by their signed distance from the order/disorder boundary defined by the Uversky-Gillepse-Fink boundary plot [29] (shown on the right) using its mean hydrophobicity and mean net charge per residue (Fig. 2K). The use of this track is examined in section 4.3. The next track colors blobs by the PV2 predicted disordered fraction [33] (Fig. 2L), stored in the database of disordered prediction (D2P2) and retrieved from UniProt. We note that there is generally a lack of consensus among disorder predictors for whether or not a given group of residues is disordered [34]. Additionally, if the user does not access the protein sequence using a UniProt ID, this track will not appear.

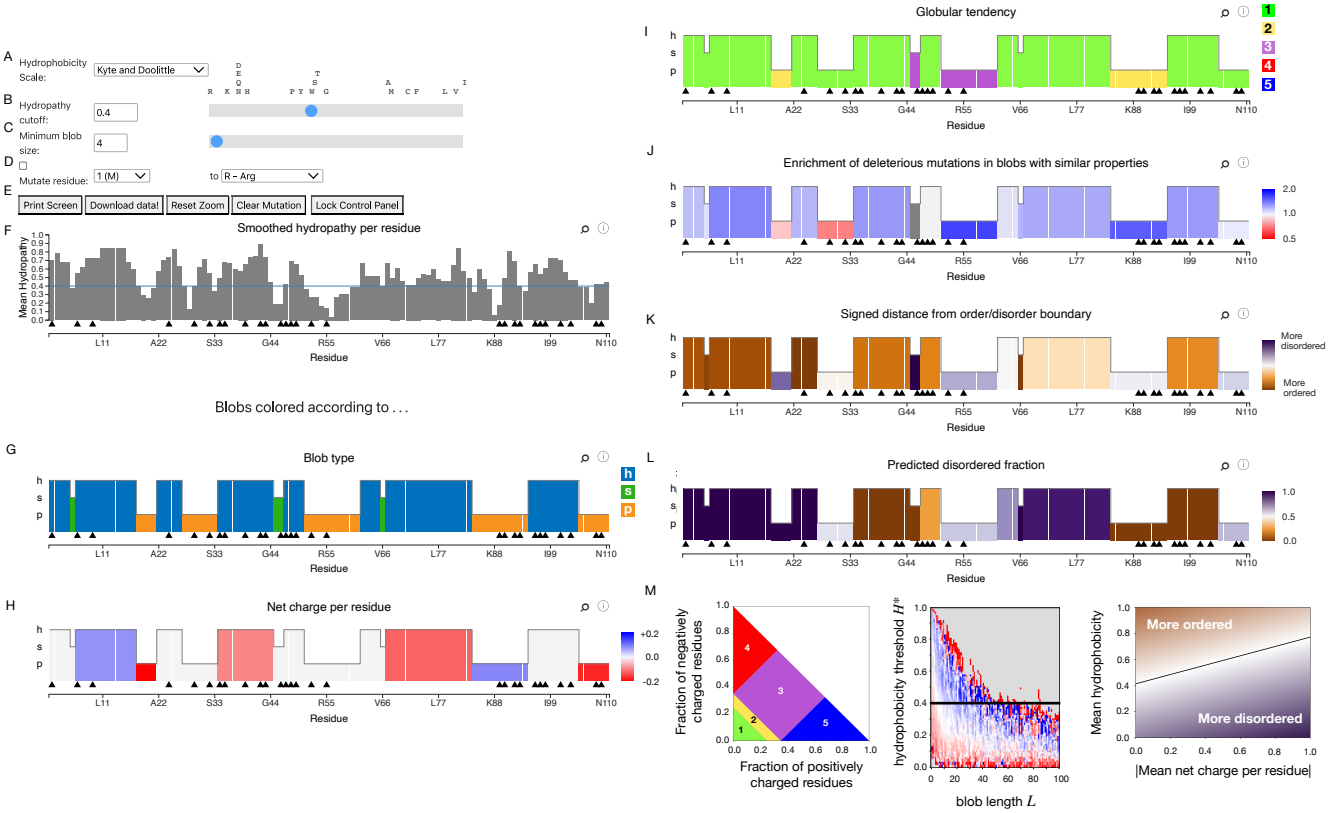

Figure 2. Screenshots from the blobulator webtool results page. Parameter and sequence adjustments are made via: a dropdown menu for selecting a hydropathy scale (A), hydropathy cutoff H^* adjustment (numerical entry, slider, and 1-letter amino acid code) (B), and minimum h- or p-blob length adjustment $L_{\min}$ (numerical entry and slider) (C). Introducing mutations is done using the "mutate residue" panel (D). Additional data and display options: print screen, download data, reset zoom, clear mutation, and lock control panel (E). The "smoothed hydropathy per residue" track displays both residue hydropathy (gray bars) and H^* (blue, horizontal line) (F). Clickable black triangles introduce known disease-associated mutations (only available via ID entry of a human sequence), and zooming is supported on all tracks. Blob type is represented by both bar height (y-axis) and color on the "blob type" track (height is preserved in subsequent tracks) (G). Blobs are then colored according physicochemical properties: H) net charge per residue: positive (blue) or negative (red); I) Das-Pappu phase [28] (plot shown in M): globular (1, green), Janus/boundary (2, yellow), strong polyelectrolyte (3, purple), strong polyanion (4, red), strong polycation (5, blue); J) dSNP enrichment (blue) or depletion (red) based on analysis using human SNPs [2] (plot shown in M); K) position on the Uversky-Gillepse-Fink boundary plot [29] (plot shown in M); and L) predicted fraction of disordered residues, according to PV2 via D2P2. M) Associated plots for blob coloring schemes in I, J, and K. The example protein shown in these tracks is insulin (UniProt ID: P01308).

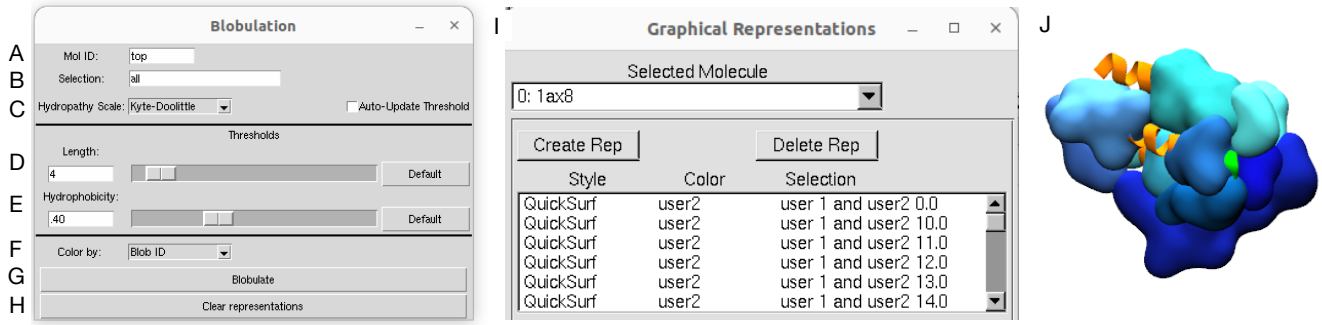

Figure 3. Screenshot of the VMD blobulator plugin. The user sets the Mol ID of the protein (A), provides an atomselection for the display (B), selects a hydrophobicity scale and may select the auto update option to snap the threshold to a scale-dependent default (C), sets the minimum blob length $L_{\min}$ (D) and hydrophobicity cutoff H^* (E), selects how blobs are colored (F). The "Blobulate" button initiates blobulation and creates graphical representations within VMD (G). All representations are deleted by clicking the "Clear representations" button (H). The "Graphical Representations" control panel contains the created graphical representations, which can be further modified (I), and are eventually displayed in the example viewer window (J) showing Leptin (PDB: 1ax8).

3.2 Python Scripting and Command Line Interface

For use in high-throughput applications, like those presented in Ref. 2, we also provide a stand-alone Python package for blobulation [35]. This package is also the backend for the blobulator webtool, and is called when a protein is submitted via that interface. The package can be used either in Python scripts or directly via a command line interface (CLI) [35]. The input to the CLI can be either amino acid or DNA sequences in plain text or FASTA format. The outputs of the CLI are CSV files (one for each input sequence) with each amino acid labeled by blob type and the blob index, identical to the downloadable blobulation output of the webtool (see 3.1). Example scripts are also provided on the blobulator GitHub [35].

3.3 VMD Plugin

While the webtool allows users to view the blobulation of the sequence and various averaged physicochemical properties, the plugin tool for Visual Molecular Dynamics (VMD) [19] allows users to incorporate blob information into the visualization and analysis of protein structure and dynamics. Users can assign blobs with the same parameters and algorithm in the webtool, while taking advantage of the existing VMD tools for scriptable analyses and creation of high-quality images and movies. Additionally, the blobulation algorithm was reimplemented in TCL, the native VMD scripting language, for maximum compatibility. The TCL script extracts the sequence information from the provided structure and then detects blobs based on user-provided parameters. The user, user2, and user3 fields associated with each atom store the type of blob (h, p, s), the blob ID, or the group ID, respectively. This information can then be used in atom selections or user-field coloring schemes provided by the VMD software; the molecular images in this manuscript were created by creating an atom selection for each h-blob, representing it with "quick-surf", and coloring each h-blob a slightly different shade of blue. Examples of images created using this plugin tool can be found throughout applications in this paper (Figs. 4, 5, 7, 9). For proteins with many blobs, this approach can require many separate graphical representations; we include an automated approach to create the representations in a batch process.

Since users of VMD typically create representations through GUIs, we also provide a GUI through which the user can provide and adjust parameters, as well as call the scripts that blobulate the sequence and create the representations. This GUI is similar to the control panel found on the blobulator webtool's output page, with some additional features. There is an option to blobulate only part of a protein using the Selection field, which may be useful for large proteins or to "zoom in" on a given set of residues (Fig. 3B). Adjacent to the threshold sliders, there are buttons to restore each field to its default value: 4 and the equivalent of 0.4 on the selected scale for the length and hydrophobicity thresholds, respectively (Fig. 3D and E). Additionally, in the viewer window, users are able to adaptively tune parameters and view how this changes the blobs found on a protein's structure. Finally, using a dropdown menu (Fig. 3F), users can color blobs by their type (all h-blobs are colored the same shade of blue), or by their ID (all h-blobs are colored distinct shades of blue). Scripts and installation instructions for the VMD plugin can be found in the blobulator GitHub repository: `https://github.com/BranniganLab/blobulator`.

4 EXAMPLE APPLICATIONS

In this section, we present the blobulation of three different types of proteins: a globular protein, a membrane protein, and an IDP. For each protein, we demonstrate how blobulation can provide a hydropathy-based framework for identifying protein modularity using only the sequence. We have chosen example proteins with previously established features, such as binding sites and disease-associated mutations, and illustrate the context that blobulation provides to these known features and how future research might utilize the *blobulator* toolkit for proteins with less established features.

4.1 Globular Protein: Lysozyme

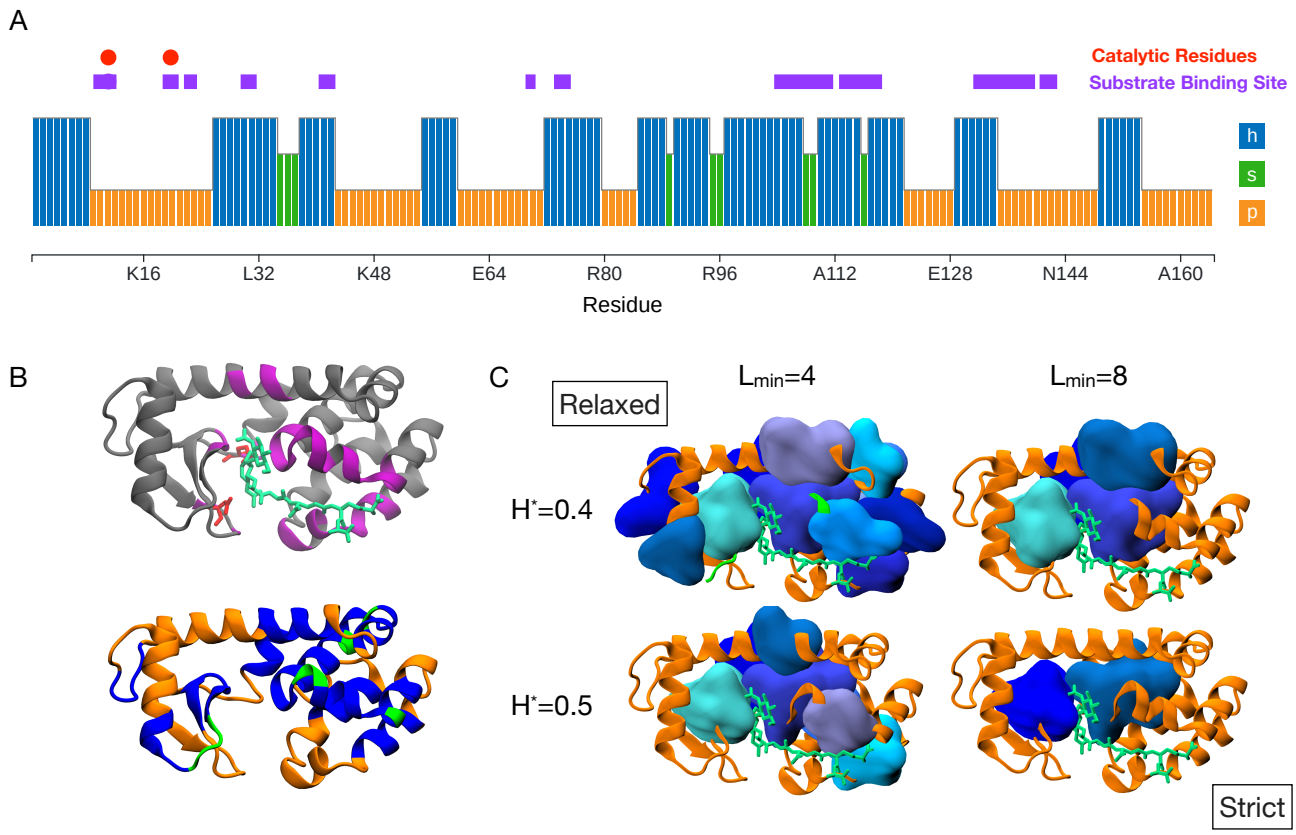

Figure 4. Blobulation of lysozyme. A) Blobs colored according to blob-type, as outputted from the blobulator webtool, and produced using default settings (H^*= 0.4, $L_{\min}$= 4). Annotations indicate catalytic residues (red) and the substrate binding site (purple). B) Molecular image of lysozyme (PDB:148L) with peptidoglycan (green) colored by the substrate binding site (left, residues found within 7Å of peptidoglycan) or by blob type (right, h-blob: blue, p-blob: orange, s-blob: green). C) Blobulation under increasingly stringent settings (H^*= 0.4 and 0.5, $L_{\min}$= 4 and 8). H-blobs are shown as surfaces. Molecular images were generated in VMD [36, 37] using the VMD plugin introduced in 3.3.

In an aqueous environment, most globular proteins have a highly hydrophobic core surrounded by a solvent-accessible surface. One such protein is lysozyme, which cleaves the sugar and peptide components of peptidoglycan. To provide an example of how one might detect hydrophobic blobs that correspond to the hydrophobic core of a globular protein, we blobulated bacteriophage T4 lysozyme and varied both the hydrophobicity threshold (H^*) and minimum length ($L_{\min}$) (Fig. 4). Lysozyme has two catalytic residues near the N-terminus and several substrate contact sites along the sequence (Fig. 4A). Higher hydrophobicity and length thresholds eliminate the h-blobs that are detected at the surface of the protein when using more relaxed settings (Fig. 4B). Blobulation using the most stringent settings shown here (H^*= 0.5, $L_{\min}$= 8) reveals two h-blobs at the protein's center, away from the solvent-accessible surface. By gradually increasing parameter thresholds, we can isolate the components of a globular protein that correspond to the core, as well as the shorter and less hydrophobic blobs that interact at the protein's surface. This is consistent with previous findings that h-blobs tend to be buried in structured proteins [2].

To investigate whether the substrate binding site is composed of h-blobs, we identified blobs containing contact residues within 7Å of peptidoglycan (shown in Fig. 4). When using a shorter $L_{\min}$ ($L_{\min}$= 4), h-blobs are found contacting peptidoglycan across its entire length. However, when using a longer $L_{\min}$ ($L_{\min}$= 8), the only detected h-blobs in the substrate binding site are those in contact with the peptidoglycan sugar component. Stabilization of this sugar ring is vital for the enzyme's ability to cleave the peptidoglycan [38], and we find this ring wedged between two long hydrophobic blobs (Fig. 4B), providing an example of long hydrophobic blobs with interactions critical for function.

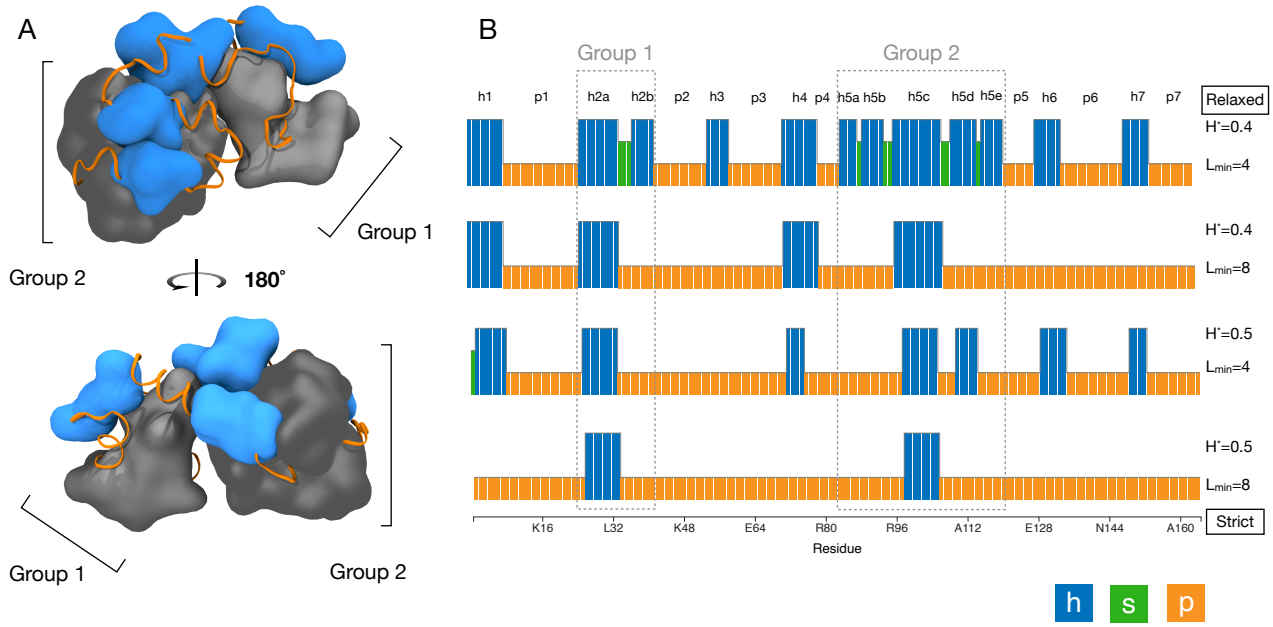

Figure 5. Blob groups in T4 lysozyme. A) Structural view (PDB:2LZM) blobulated using default settings (H^*= 0.4, $L_{\min}$= 4). Groups are gray, ungrouped h-blobs are blue, and p-blobs are orange ribbons. B) Blobulation of T4 lysozyme under increasingly stringent settings. Annotations indicate blob identifiers and blob groups. Molecular images were generated in VMD [36, 37].

Blobulating lysozyme using "relaxed" settings (H^*= 0.4 and $L_{\min}$= 4) reveals two examples of "blob groups": sets of h-blobs separated only by s-blobs (as defined in section 2, shown in gray in Fig. 5A). Both blob groups are found near the center of lysozyme, surrounded by ungrouped h-blobs and p-blobs. Additionally, each group contains an h-blob that remains detected under increasingly stringent settings and is also found at the core of lysozyme (Fig. 5B, shown also in the molecular image in 4). The groups detected here are akin to tertiary elements formed from a network of secondary interacting structure elements, such as alpha-helical bundles. This is an example of hierarchical blob clustering, where hydrophobic blobs detected at restrictive parameters are found in the protein core surrounded by less hydrophobic blobs within the same group, which are in turn surrounded by individual h-blobs.

To provide an example of how blobulation links known mutations with an effect on their local hydrophobic context, we blobulated the lysozyme mutants S117V and T157I. Both mutations affect the temperature on stability of the lysozyme by altering intraprotein hydrophobic interactions. S117 makes the protein more stable at higher temperatures by altering hydrophobic residue interactions in the substrate binding cleft [39], while T157I makes the protein less stable at higher temperatures and disrupts hydrogen bonding at the periphery of the protein [40]. We find that both mutations change the blob topology (Fig. 6): S117V merges two h-blobs, and T157I creates a new h-blob four residues in length. We have previously found that mutations that split, dissolve, or merge h-blobs are enriched for deleterious mutations [2], and this result is consistent with that finding. Additionally, the h-blob introduced by the T157I mutation is classified as a Janus region on the Das-Pappu phase diagram [28]. Janus proteins often have degenerate conformations and switch between ordered and disordered depending on their environment [28], which may cause a change in this blob at higher temperatures and lead to less overall stability for the protein. Finally, it is noteworthy that blobulation is able to detect a change from both mutations, whereas tools such as AlphaFold [13] (which predict atomic coordinates and not necessarily interactions) do not detect structural differences between these mutants and the wildtype (average backbone root mean squared deviation of predicted wildtype and mutant structures is 0.09Å).

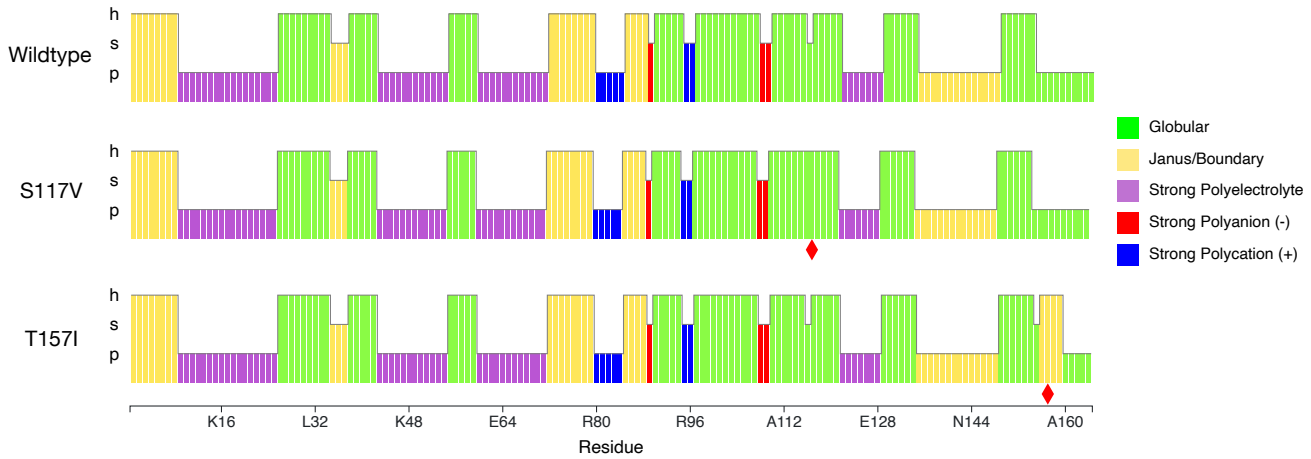

Figure 6. Globular tendency of T4 lysozyme blobs. Blobulation using default parameters (H^*= 0.4, $L_{\min}$= 4). Blobs for each sequence are colored by Das-Pappu phase [28], as in Fig. 2I. Red diamonds indicate mutated residues. S117V, which increases thermostability, joins two h-blobs into one. T157I, which decreases thermostability, creates a new h-blob (as well as an s-blob).

4.2 Membrane Proteins: Pentameric Ligand-Gated Ion Channels (pLGICs)

Unlike soluble globular proteins, which enclose their hydrophobic residues inside a hydrophilic shell, integral membrane proteins use exposed hydrophobic regions as anchors in the cell membrane. The pentameric ligand-gated ion channel (pLGIC) protein family is responsible for neurotransmitter reception in the nervous system of bilateria, but representatives are also found in many other eukaryotes and some prokaryotes [41]. Despite relatively low sequence identity and diverse environmental and ligand requirements, pLGICs are highly structurally conserved [42–46]. The N-terminal extracellular domain consists of a beta-sandwich, which is responsible for ligand sensing and much of the substrate selectivity (Fig. 7A-B). The transmembrane domain consists of four transmembrane helices per subunit (Fig 7D)[41]. When the ligand is bound in the ECD, a signal is transmitted to the transmembrane helical bundle (Fig. 7D), which changes conformation to permit ions to pass through the channel formed by the five pore-lining helices. In this example, we will compare the blobulation of two structural models of the pLGIC family: first, the *Caenorhabditis elegans* glutamate-gated chloride channel (GluCl); and second, the proton-gated *Gloeobacter* ligand-gated cation channel (GLIC).

Starting from the N-terminus, both GluCl and GLIC contain disordered N-terminal extensions which include signal sequences (Fig. 8). GluCl has an additional hydrophobic blob in the N-terminal extension. P-blobs dominate the extracellular beta-sandwich in both cases. The two additional hydrophobic blobs in GluCl correspond to $\beta 1$ and $\beta 5$, which contain two of the four residues that contact glutamate during binding. GLIC, in contrast, is proton-gated via titratable residues mostly in the ECD [43] and lacks these extracellular h-blobs. The transmembrane helices are clearly visible using $H^* = 0.33$ and $L_{\min} \approx 19$. This length corresponds to the typical thickness of a plasma membrane[47]. The extended p-blob seen in the blobulation of GluCl (Fig. 7) corresponds to a disordered intracellular domain common in eukaryotic pLGICs [43].

As ion channels, the electrostatic potential of a pLGIC pore plays a central role in both conductivity and selectivity. The blobs along the ion conduction pathway (in the ECD and the M2 helix) of GluCl are primarily positive or neutral. GLIC, in contrast, has a net negative charge in these regions (Fig. 8). This is consistent with the anion conductivity of GluCl and the cation conductivity of GLIC. The charges in the M2 blob are on the intracellular side, close to the M1-M2 loop. GluCl has a single positive charge (R306), while GLIC has several negative charges in this region. GLIC has an additional positive charge near the M2-M3 loop due to K290, which can be seen by using the blobulator's zoom feature. Finally, the M4 helix of each protein has a net positive charge, consistent with experimental observations that M4 binds anionic lipids [48, 49]. These results show the blobulator's potential for extracting structural information from membrane proteins either individually or within a larger evolutionary context.

4.3 IDP: α-synuclein

Intrinsically disordered proteins (IDPs) fulfill critical roles despite lacking stable tertiary structure (as reviewed in Ref. 50). One such IDP is α-synuclein. Although its function is not fully understood, it forms aggregates implicated in Parkinson's disease [51, 52]. It contains a helix-turn-helix motif at its N-terminal and a random coil near its C-terminal. Although residues throughout the helical region of this protein interact with other α-synucleins, residues 71 to 82 are the necessary aggregating

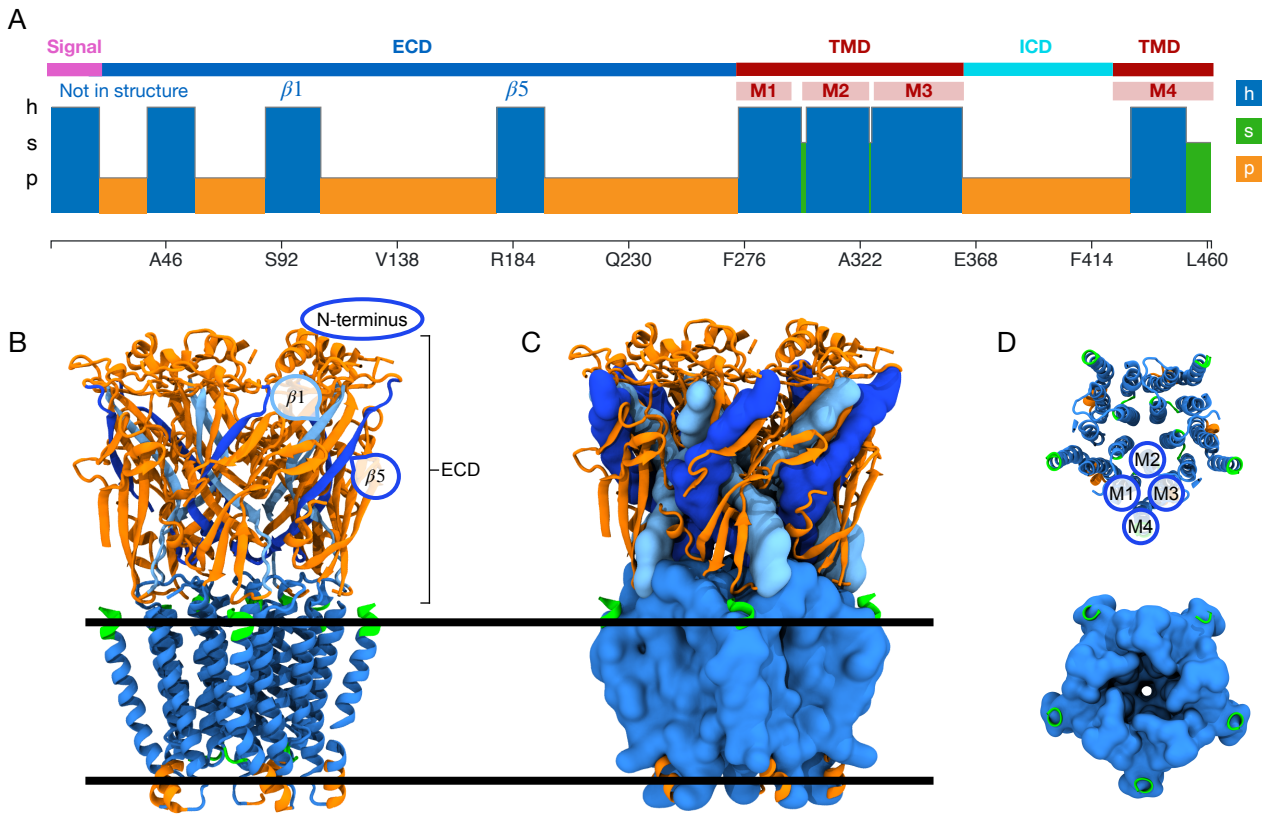

Figure 7. Blobulation of GluCl, a pentameric ligand-gated ion channel. A) Blobulation of GluCl (UniProt ID: G5EBR3) using settings to detect transmembrane regions ($H^* = 0.33$, $L_{\min} = 19$). B) A cartoon representation of GluCl (PDB: 3RHW) colored according to blob type and ID: $\beta 1$ blob (light blue), $\beta 5$ blob (dark blue), transmembrane blobs (intermediate blue), terminal s-blobs (green), p-blobs (orange). The first two blobs in the sequence are absent from the structure. The black lines represent the lipid membrane. C) GluCl showing h-blobs in surface view to better show blob-blob contacts. Colored as in B. D) Extracellular views of the TMDs of the proteins shown in B (upper) and C (lower). Molecular images were generated in VMD [36, 37].

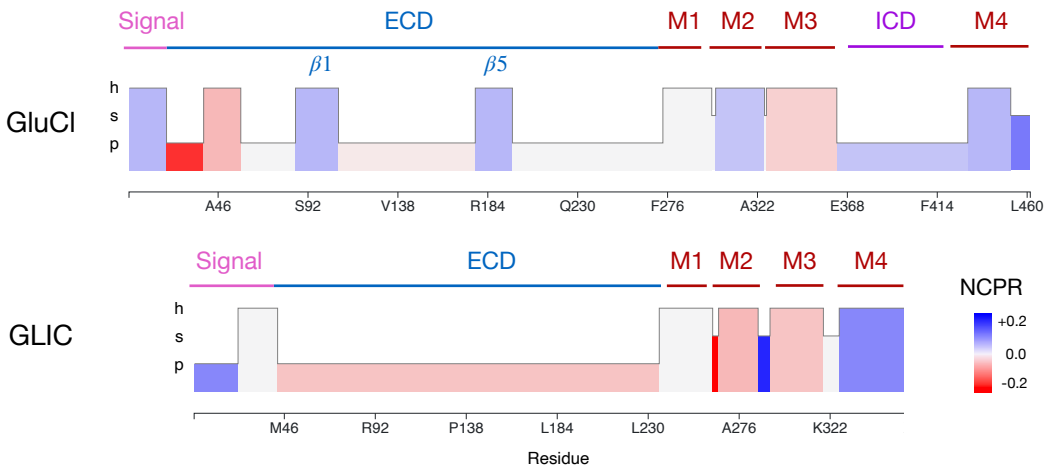

Figure 8. Comparative blobulation of GluCl (as in Fig. 7), an anion-conducting pLGIC, and GLIC, a cation-conducting pLGIC. Net Charge Per Residue (NCPR) tracks for GluCl (UniProt: G5EBR3) and GLIC (UniProt: Q7NDN8) blobulated using settings to detect transmembrane regions ($H^* = 0.33$, $L_{\min} = 19$). Tracks are aligned by the beginning of the h-blob containing the M2 (pore-lining) helix. Annotations indicate important sequence features available on UniProt and indicated in Fig. 7. Signal sequences are indicated in pink, ECD regions are indicated in blue, and transmembrane helices (M1 to M4) are indicated in red.

motif [53] (Fig. 9 B). To identify whether h-blobs overlap with this known motif, we blobulated α-synuclein using the settings H^*= 0.4 and $L_{\min}$= 4 and found that the h-blob group in α-synuclein coincides with this aggregating motif (Fig. 9 A). This result demonstrates that the blobulation of IDPs, where secondary structure elements are often transient, can reveal hydrophobic clusters that in some cases coincide with tertiary interactions.

One application for blobulation is in quantifying the effects of mutated residues on their local sequence context. The helix-turn-helix motif of α-synuclein (Fig. 9B) also interacts with cell membranes [54], but the mechanism driving this interaction is unknown [54]. Here, we consider the known disease-causing mutations A30P and A53T. When expressed in yeast, A53T and wildtype α-synuclein initially bind to the membrane before forming cytoplasmic aggregates, whereas A30P α-synuclein is dispersed throughout the cytoplasm [55]. The A30P mutant has a shorter and less ordered third h-blob than A53T and wildtype α-synuclein (Fig. 10 C). Because the order prediction considers residue hydrophobicity, and proline is less hydrophobic than alanine, a decrease in order is expected because predicted order increases with hydrophobicity on the Uversky-Gillepse-Fink boundary plot [29]. This is also consistent with the finding that while the helical domain of the mutant is partially disrupted, its N-terminus is more dynamic than the wildtype [56]. In contrast to the wildtype, no blobs are shortened in the A53T mutant protein, and the order of the blob containing the mutated residue is only minimally altered.

5 CONCLUSION

Here we presented the *blobulator* toolkit, consisting of a webtool, a command line interface, and a VMD plugin, for users to define and characterize modules based on contiguous hydrophobicity. To demonstrate the utility of the toolkit in detecting the underlying sequence organization, we provided example applications across three classic protein families. We showed how one can use the blobulator to detect blobs corresponding to the hydrophobic core and transmembrane domains, and provided examples where known disease-causing mutations alter blob topology.

The *blobulator* shares a similar aim with other tools built to incorporate tertiary interaction information into bioinformatics analysis (Refs. [13, 57–64]). In contrast to many of these, the underlying algorithm (blobulation) is transparent, straightforward, deterministic, and grounded entirely in biophysical principles. While the toolkit does not currently incorporate post-translational modifications, multiple sequences on the same page, or a structure viewer for sequences blobulated using the webtool, these features are currently under development. The approaches outlined in section 4 can be extended to downstream applications, including: predicting protein segments with critical roles in tertiary interactions, defining local order and disorder with clear edges, utilizing *blobulator* CLI for much larger datasets, and predicting protein features from sequence.

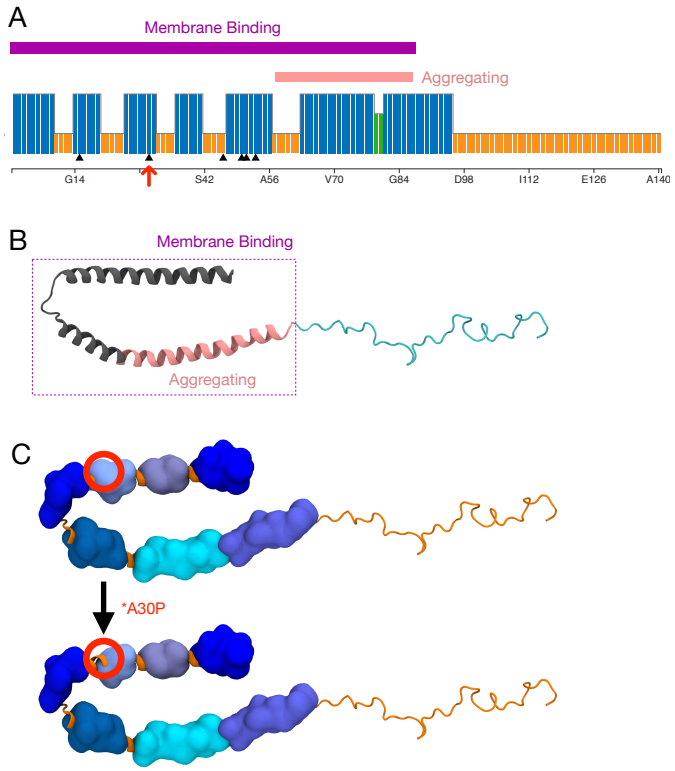

Figure 9. Blobulation of α-synuclein, an intrinsically disordered protein. A) Webtool "blob type" track of α-synuclein (UniProt ID: P37840) blobulated using default settings (H^*= 0.4, $L_{\min}$= 4). Annotations indicate the membrane-interacting region (purple) and the protein-interacting region (pink). B) α-synuclein structure (PDB: 1XQ8), labeled by the membrane-interacting region in cartoon. Aggregating in pink, non-aggregating in purple. C) Changes to blobs (blue surfaces, change indicated by red circles) caused by the A30P mutant compared to the wildtype. Molecular images were generated in VMD [36, 37].

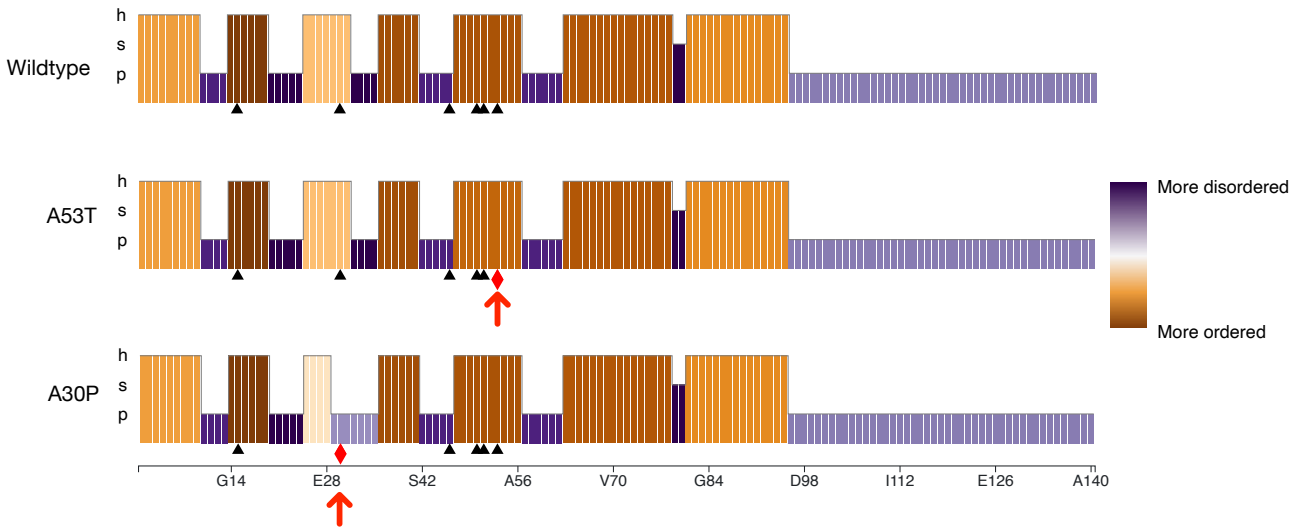

Figure 10. Order predictions for blobs of α-synuclein. Wildtype, A53T, and A30P mutants colored by blob disorder as calculated using each blob's signed distance from the order/disorder boundary of the Uversky-Gillepse-Fink boundary plot [29]. Mutations are indicated: A53T and A30P (red diamonds and arrows), and other known disease-associated mutations (black triangles). Blobulation uses default settings (H^*= 0.4, $L_{\min}$= 4).

6 DATA AND SOFTWARE AVAILABILITY STATEMENT

The backend for the *blobulator* toolkit can be found on GitHub at `https://github.com/BranniganLab`. Additionally, the webtool is available as a graphical user interface at `https://www.blobulator.branniganlab.org/`.

7 ACKNOWLEDGEMENTS

The authors acknowledge the Office of Advanced Research Computing (OARC) at Rutgers, The State University of New Jersey, for providing access to the Amarel cluster and associated research computing resources. TJ is supported by NIH 1K08GM139031, MEBH is supported by NIH 1R35GM134957, and LR, ESM, CP, and GB are supported by NSF DGE 2152059. We are also grateful to Dr. Jérôme Hénin, Mr. Jesse Sandberg, and Mr. Jahmal Ennis for testing and feedback.

REFERENCES

[1] Ruchi Lohia, Reza Salari, and Grace Brannigan. Sequence specificity despite intrinsic disorder: How a disease-associated val/met polymorphism rearranges tertiary interactions in a long disordered protein. *PLOS COMPUTATIONAL BIOLOGY*, 15(10), OCT 2019.

[2] Ruchi Lohia, Matthew E. B. Hansen, and Grace Brannigan. Contiguously hydrophobic sequences are functionally significant throughout the human exome. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, 119(12), MAR 22 2022.

[3] Kana Shimizu. Poodle: Tools predicting intrinsically disordered regions of amino acid sequence. In D Kihara, editor, *PROTEIN STRUCTURE PREDICTION, 3RD EDITION*, volume 1137 of *Methods in Molecular Biology*, pages 131–145. 2014.

[4] David T. Jones and Domenico Cozzetto. Disopred3: precise disordered region predictions with annotated protein-binding activity. *BIOINFORMATICS*, 31(6):857–863, MAR 15 2015.

[5] Gabor Erdos, Matyas Pajkos, and Zsuzsanna Dosztanyi. Iupred3: prediction of protein disorder enhanced with unambiguous experimental annotation and visualization of evolutionary conservation. *NUCLEIC ACIDS RESEARCH*, 49(W1):W297–W303, JUL 2 2021.

[6] R Linding, LJ Jensen, F Diella, P Bork, TJ Gibson, and RB Russell. Protein disorder prediction: Implications for structural proteomics. *STRUCTURE*, 11(11):1453–1459, NOV 2003.

[7] Ian Walsh, Flavio Seno, Silvio C. E. Tosatto, and Antonio Trovato. Pasta 2.0: an improved server for protein aggregation prediction. *NUCLEIC ACIDS RESEARCH*, 42(W1):W301–W307, JUL 1 2014.

[8] Alex S. Holehouse, Rahul K. Das, James N. Ahad, Mary O. G. Richardson, and Rohit V. Pappu. Cider: Resources to analyze sequence-ensemble relationships of intrinsically disordered proteins. *BIOPHYSICAL JOURNAL*, 112(1):16–21, JAN 10 2017.

[9] Lukas Käll, Anders Krogh, and Erik L.L Sonnhammer. A combined transmembrane topology and signal peptide prediction method. *Journal of Molecular Biology*, 338(5):1027–1036, 2004.

[10] Jeppe Hallgren, Konstantinos D. Tsirigos, Mads Damgaard Pedersen, José Juan Almagro Armenteros, Paolo Marcatili, Henrik Nielsen, Anders Krogh, and Ole Winther. Deeptmhmm predicts alpha and beta transmembrane proteins using deep neural networks. *bioRxiv*, 2022.

[11] LJ McGuffin, K Bryson, and DT Jones. The psipred protein structure prediction server. *BIOINFORMATICS*, 16(4):404–405, APR 2000.

[12] J Cheng, AZ Randall, MJ Sweredoski, and P Baldi. Scratch: a protein structure and structural feature prediction server. *NUCLEIC ACIDS RESEARCH*, 33(2):W72–W76, JUL 1 2005.

[13] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Zidek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstein, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli, and Demis Hassabis. Highly accurate protein structure prediction with alphafold. *NATURE*, 596(7873):583+, AUG 26 2021.

[14] Scott Montgomerie, Joseph A. Cruz, Savita Shrivastava, David Arndt, Mark Berjanskii, and David S. Wishart. Proteus2: a web server for comprehensive protein structure prediction and structure-based annotation. *NUCLEIC ACIDS RESEARCH*, 36(S):W202–W209, JUL 2008.

[15] Daniel W. A. Buchan, Federico Minneci, Tim C. O. Nugent, Kevin Bryson, and David T. Jones. Scalable web services for the psipred protein analysis workbench. *NUCLEIC ACIDS RESEARCH*, 41(W1):W349–W357, JUL 2013.

[16] Alexey Drozdetskiy, Christian Cole, James Procter, and Geoffrey J. Barton. Jpred4: a protein secondary structure prediction server. *NUCLEIC ACIDS RESEARCH*, 43(W1):W389–W394, JUL 1 2015.

[17] Michael Schantz Klausen, Martin Closter Jespersen, Henrik Nielsen, Kamilla Kjaergaard Jensen, Vanessa Isabell Jurtz, Casper Kaae Sonderby, Morten Otto Alexander Sommer, Ole Winther, Morten Nielsen, Bent Petersen, and Paolo Marcatili. Netsurfp-2.0: Improved prediction of protein structural features by integrated deep learning. *PROTEINS-STRUCTURE FUNCTION AND BIOINFORMATICS*, 87(6):520–527, JUN 2019.

[18] Gilbert Deleage. Alignsec: viewing protein secondary structure predictions within large multiple sequence alignments. *BIOINFORMATICS*, 33(24):3991–3992, DEC 15 2017.

[19] W Humphrey, A Dalke, and K Schulten. Vmd: Visual molecular dynamics. *JOURNAL OF MOLECULAR GRAPHICS & MODELLING*, 14(1):33–38, FEB 1996.

[20] EF Pettersen, TD Goddard, CC Huang, GS Couch, DM Greenblatt, EC Meng, and TE Ferrin. Ucsf chimera - a visualization system for exploratory research and analysis. *JOURNAL OF COMPUTATIONAL CHEMISTRY*, 25(13):1605–1612, OCT 2004.

[21] AA Porollo, R Adamczak, and J Meller. Polyview: a flexible visualization tool for structural and functional annotations of proteins. *BIOINFORMATICS*, 20(15):2460–2462, OCT 12 2004.

[22] Fabio Madeira, Nandana Madhusoodanan, Joonheung Lee, Alberto Eusebi, Ania Niewielska, Adrian R. N. Tivey, Rodrigo Lopez, and Sarah Butcher. The embl-ebi job dispatcher sequence analysis tools framework in 2024. *NUCLEIC ACIDS RESEARCH*, 52(W1):W521–W525, APR 10 2024.

[23] Andreas Prlic, Andrew Yates, Spencer E. Bliven, Peter W. Rose, Julius Jacobsen, Peter V. Troshin, Mark Chapman, Jianjiong Gao, Chuan Hock Koh, Sylvain Foisy, Richard Holland, Gediminas Rimsa, Michael L. Heuer, H. Brandstaetter-Mueller, Philip E. Bourne, and Scooter Willis. Biojava: an open-source framework for bioinformatics in 2012. *BIOINFORMATICS*, 28(20):2693–2695, OCT 15 2012.

[24] Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., and Bairoch A. *The Proteomics Protocols Handbook*, chapter Protein Identification and Analysis Tools on the Expasy Server, pages 571–607. Humana Press, 2005.

[25] Craig Snider, Sajith Jayasinghe, Kalina Hristova, and Stephen H. White. Mpex: A tool for exploring membrane proteins. *PROTEIN SCIENCE*, 18(12):2624–2628, DEC 2009.

[26] P Rice, I Longden, and A Bleasby. Emboss: The european molecular biology open software suite. *TRENDS IN GENETICS*, 16(6):276–277, JUN 2000.

[27] Konstantin Okonechnikov, Olga Golosova, Mikhail Fursov, and UGENE Team. Unipro ugene: a unified bioinformatics toolkit. *BIOINFORMATICS*, 28(8):1166–1167, APR 15 2012.

[28] Rahul K. Das and Rohit V. Pappu. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, 110(33):13392–13397, AUG 13 2013.

[29] VN Uversky, JR Gillespie, and AL Fink. Why are “natively unfolded” proteins unstructured under physiologic conditions? *PROTEINS-STRUCTURE FUNCTION AND GENETICS*, 41(3):415–427, NOV 15 2000.

[30] J KYTE and RF DOOLITTLE. A simple method for displaying the hydropathic character of a protein. *JOURNAL OF MOLECULAR BIOLOGY*, 157(1):105–132, 1982.

[31] D EISENBERG, RM WEISS, TC TERWILLIGER, and W WILCOX. Hydrophobic moments and protein-structure. *FARADAY SYMPOSIA OF THE CHEMICAL SOCIETY*, (17):109–120, 1982.

[32] C. Preston Moon and Karen G. Fleming. Side-chain hydrophobicity scale derived from transmembrane protein folding into lipid bilayers. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, 108(25):10174–10177, JUN 21 2011.

[33] Matt E. Oates, Pedro Romero, Takashi Ishida, Mohamed Ghalwash, Marcin J. Mizianty, Bin Xue, Zsuzsanna Dosztanyi, Vladimir N. Uversky, Zoran Obradovic, Lukasz Kurgan, A. Keith Dunker, and Julian Gough. D^2p^2 database of disordered protein predictions. *NUCLEIC ACIDS RESEARCH*, 41(D1):D508–D516, JAN 2013.

[34] Yumeng Liu, Xiaolong Wang, and Bin Liu. A comprehensive review and comparison of existing computational methods for intrinsically disordered protein and region prediction. *BRIEFINGS IN BIOINFORMATICS*, 20(1):330–346, JAN 2019.

[35] Ruchi Lohia, Connor Pitman, Ezry Santiago-McRae, Ryan Lamb, Lindsey Riggs, Kaitlin Bassi, Matthew E.B. Hansen, Thomas T. Joseph, and Grace Brannigan. https://github.com/branniganlab/blobulator, December 2024.

[36] William Humphrey, Andrew Dalke, and Klaus Schulten. VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14:33–38, 1996.

[37] John Stone. *An Efficient Library for Parallel Ray Tracing and Animation*. Master's thesis, Computer Science Department, University of Missouri-Rolla, April 1998.

[38] R KUROKI, LH WEAVER, and BW MATTHEWS. A covalent enzyme-substrate intermediate with saccharide distortion in a mutant t4 lysozyme. *SCIENCE*, 262(5142):2030–2033, DEC 24 1993.

[39] BK SHOICHET, WA BAASE, R KUROKI, and BW MATTHEWS. A relationship between protein stability and protein function. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, 92(2):452–456, JAN 17 1995.

[40] T ALBER, DP SUN, K WILSON, JA WOZNIAK, SP COOK, and BW MATTHEWS. Contributions of hydrogen-bonds of thr-157 to the thermodynamic stability of phage-t4 lysozyme. *NATURE*, 330(6143):41–46, NOV 5 1987.

[41] Mariama Jaiteh, Antoine Taly, and Jérôme Hénin. Evolution of pentameric ligand-gated ion channels: Pro-loop receptors. *PLOS ONE*, 11(3), MAR 17 2016.

[42] Marco Cecchini and Jean-Pierre Changeux. The nicotinic acetylcholine receptor and its prokaryotic homologues: Structure, conformational transitions & allosteric modulation. *Neuropharmacology*, 96:137–149, September 2015.

[43] Rebecca J. Howard. Elephants in the dark: Insights and incongruities in pentameric ligand-gated ion channel models. *Journal of Molecular Biology*, 433(17):167128, August 2021.

[44] Ákos Nemecz, Marie S. Prevost, Anaïs Menny, and Pierre-Jean Corringer. Emerging molecular mechanisms of signal transduction in pentameric ligand-gated ion channels. *Neuron*, 90(3):452–470, May 2016.

[45] Ludovic Sauguet, Azadeh Shahsavar, and Marc Delarue. Crystallographic studies of pharmacological sites in pentameric ligand-gated ion channels. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1850(3):511–523, March 2015.

[46] Corrie J.B. daCosta and John E. Baenziger. Gating of pentameric ligand-gated ion channels: Structural insights and ambiguities. *Structure*, 21(8):1271–1283, August 2013.

[47] Carlos Baeza-Delgado, Marc A. Marti-Renom, and Ismael Mingarro. Structure-based statistical analysis of transmembrane helices. *EUROPEAN BIOPHYSICS JOURNAL WITH BIOPHYSICS LETTERS*, 42(2-3, SI):199–207, MAR 2013.

[48] Casey L. Carswell, Camille M. Hénault, Sruthi Murlidaran, J.P. Daniel Therien, Peter F. Juranka, Julian A. Surujballi, Grace Brannigan, and John E. Baenziger. Role of the fourth transmembrane α helix in the allosteric modulation of pentameric ligand-gated ion channels. *Structure*, 23(9):1655–1664, September 2015.

[49] Thorsten Althoff, Ryan E. Hibbs, Surajit Banerjee, and Eric Gouaux. X-ray structures of glucl in apo states reveal a gating mechanism of cys-loop receptors. *NATURE*, 512(7514):333+, AUG 21 2014.

[50] Peter E. Wright and H. Jane Dyson. Intrinsically disordered proteins in cellular signalling and regulation. *NATURE REVIEWS MOLECULAR CELL BIOLOGY*, 16(1):18–29, JAN 2015.

[51] MG Spillantini, RA Crowther, R Jakes, M Hasegawa, and M Goedert. α-synuclein in filamentous inclusions of lewy bodies from parkinson's disease and dementia with lewy bodies. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, 95(11):6469–6473, MAY 26 1998.

[52] JE Galvin, K Uryu, VMY Lee, and JQ Trojanowski. Axon pathology in parkinson's disease and lewy body dementia hippocampus contains α-, β-, and γ-synuclein. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, 96(23):13450–13455, NOV 9 1999.

[53] Ricardo Guerrero-Ferreira, Nicholas M. I. Taylor, Ana-Andreea Arteni, Pratibha Kumari, Daniel Mona, Philippe Ringler, Markus Britschgi, Matthias E. Lauer, Ali Makky, Joeri Verasdonck, Roland Riek, Ronald Melki, Beat H. Meier, Anja Bockmann, Luc Bousset, and Henning Stahlberg. Two new polymorphic structures of human full-length alpha-synuclein fibrils solved by cryo-electron microscopy. *ELIFE*, 8, DEC 9 2019.

[54] Tapojyoti Das and David Eliezer. Membrane interactions of intrinsically disordered proteins: The example of alpha-synuclein. *BIOCHIMICA ET BIOPHYSICA ACTA-PROTEINS AND PROTEOMICS*, 1867(10, SI):879–889, OCT 2019.

[55] Katherina Vamvaca, Michael J. Volles, and Peter T. Lansbury, Jr. The first n-terminal amino acids of α-synuclein are essential for α-helical structure formation in vitro and membrane binding in yeast. *JOURNAL OF MOLECULAR BIOLOGY*, 389(2):413–424, JUN 5 2009.

[56] Luisel R. Lemkau, Gemma Comellas, Kathryn D. Kloepper, Wendy S. Woods, Julia M. George, and Chad M. Rienstra. Mutant protein a30p α-synuclein adopts wild-type fibril structure, despite slower fibrillation kinetics. *JOURNAL OF BIOLOGICAL CHEMISTRY*, 287(14):11526–11532, MAR 30 2012.

[57] Wei Zheng, Qiqige Wuyun, Yang Li, Quancheng Liu, Xiaogen Zhou, Chunxiang Peng, Yiheng Zhu, Lydia Freddolino, and Yang Zhang. Deep-learning-based single-domain and multidomain protein structure prediction with d-i-tasser. *NATURE BIOTECHNOLOGY*, 2025 MAY 23 2025.

[58] Michel van Kempen, Stephanie S. Kim, Charlotte Tumescheit, Milot Mirdita, Jeongjae Lee, Cameron L. M. Gilchrist, Johannes Soeding, and Martin Steinegger. Fast and accurate protein structure search with foldseek. *NATURE BIOTECHNOLOGY*, 42(2), FEB 2024.

[59] Lupeng Kong, Fusong Ju, Wei-mou Zheng, Jianwei Zhu, Shiwei Sun, Jinbo Xu, and Dongbo Bu. Proalign: Directly learning alignments for protein structure prediction via exploiting context-specific alignment motifs. *JOURNAL OF COMPUTATIONAL BIOLOGY*, 29(2, SI):92–105, FEB 2022.

[60] Zongyang Du, Hong Su, Wenkai Wang, Lisha Ye, Hong Wei, Zhenling Peng, Ivan Anishchenko, David Baker, and Jianyi Yang. The trrosetta server for fast and accurate protein structure prediction. *NATURE PROTOCOLS*, 16(12):5634–5651, DEC 2021.

[61] Fandi Wu and Jinbo Xu. Deep template-based protein structure prediction. *PLOS COMPUTATIONAL BIOLOGY*, 17(5), MAY 2021.

[62] Liam J. McGuffin, Recep Adiyaman, Ali H. A. Maghrabi, Ahmad N. Shuid, Danielle A. Brackenridge, John O. Nealon, and Limcy S. Philomina. Intfold: an integrated web resource for high performance protein structure and function prediction. *NUCLEIC ACIDS RESEARCH*, 47(W1):W408–W413, JUL 2 2019.

[63] Sheng Wang, Wei Li, Shiwang Liu, and Jinbo Xu. Raptorx-property: a web server for protein structure property prediction. *NUCLEIC ACIDS RESEARCH*, 44(W1):W430–W435, JUL 8 2016.

[64] Badri Adhikari, Debswapna Bhattacharya, Renzhi Cao, and Jianlin Cheng. Confold: Residue-residue contact-guided ab initio protein folding. *PROTEINS-STRUCTURE FUNCTION AND BIOINFORMATICS*, 83(8):1436–1449, AUG 2015.