The CSCI2952G Journal, , pp. 1–5

doi:

EnCODON- From Atoms to Abstractions: Geometric Representation Saliency in Protein-Protein Interaction Networks

Charumathi Narayanan,[1] Daniel Posmik[2] and Minh Tran[3]

[1]Department of Computer Science, Brown University, 115 Waterman St, 02906, RI, Country, [2]Department of Biostatistics, Brown University, 121 S Main St,, 02903, RI, USA and [3]The College, Brown University, 1 Prospect St, 02912, RI, USA

Abstract

Predicting protein-protein interactions (PPIs) is fundamental to understanding cellular function and drug discovery. Graph neural networks (GNNs) have emerged as a natural framework for modeling these interactions, with proteins themselves representable as graphs at the atomic or residue level. However, the optimal level of geometric detail for PPI prediction remains an open question. We compare three approaches that differ in their geometric protein representations: (1) a GNN operating on full protein graphs, preserving complete structural information [Gao et al., 2023]; (2) learned latent embeddings that capture underlying geometric structure at the node level; and (3) sequence-derived embeddings from the ESM-2 protein language model [Lin et al., 2023]. For the inter-protein interaction level, we implement GNNs with and without attention mechanisms to isolate the contribution of geometric representation from architectural choices. These results clarify which level of geometric abstraction is most informative for PPI prediction, with implications for computational efficiency and model design in protein interaction networks.We find that the high-quality, sequence-based features from the ESM-2 language model significantly outperform all GNN-derived geometric features, achieving a peak $\mathbf{F1-score}$ of $\mathbf{0.817}$. These results clarify which level of geometric abstraction is most informative for PPI prediction, with implications for computational efficiency and model design in protein interaction networks.

Key words: Protein-Protein Interaction, Geometric learning, Deep learning, Graph Attention Networks, ESM-2, Dimensionality Reduction

Introduction and Data

Proteins are the essential workers of the body. They rarely act alone, relying on other proteins to perform their tasks properly. This web of interactions lays the foundation for the activities and functions ensuring survival at the cellular level [Braun and Gingras, 2012]. Protein-protein interactions (PPIs) are thus a vital source of information about how the body functions. Moreover, since illness is often caused by mutations and disruptions to PPIs, understanding these interactions can greatly inform drug development [Greenblatt et al., 2024].

However, understanding the biochemical mechanisms underlying protein interactions remains a challenging task. One issue is the sheer scale: Given the magnitude and variety of proteins, an estimated 200,000 unique PPIs exist in humans, a significant portion of which remains undiscovered [Shin et al., 2020]. Additionally, a complex network of factors determines the mechanism of interactions, including temperature, the local microenvironment, and signaling molecules. Most notably, biochemical relationships within a protein itself have emergent effects on its interactions. Understanding these dependencies could inform what kinds of interactions might occur.

Deep learning algorithms can address these issues by integrating large-scale data to accurately represent PPIs and perform downstream tasks with these representations, including prediction of unknown protein interactions and identification of targets for drug effectiveness [Soleymani et al., 2022]. Fortunately, as biological experiments uncover more proteins and their interactions, many databases now exist for PPIs, including STRING, KEGG, BioGRID, and BioRank [Rao et al., 2014]. There are also PPI datasets for specific illnesses, such as OncoPPi, a cancer-focused dataset [Li et al., 2017]. Benchmark datasets, such as SHS27k and SHS148k, also exist to validate algorithm performance [Gao et al., 2023].

Related Work

Gao et al. [2023] propose a hierarchical graph neural network (GNN) for protein-protein interaction link prediction. The hierarchical nature of this task stems from the ability to represent proteins as graphs themselves. Thus, in the global (”top-down”) PPI network, each node is itself a graph, with protein graphs constituting exceptionally rich representations at the node level.

At the node level, a sequence of amino acids or residues corresponds to a molecular graph constructed from contact maps, where edges connect physically adjacent

© The Author . Published by Oxford University Press. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

residues (within 10 Å C_α-C_α distance). These molecular graphs carry valuable topological information, including node importance, connectivity, and local binding patterns. The model employs graph convolutional networks (GCNs) in the bottom view to learn protein representations from residue-level physicochemical properties, and graph isomorphism networks (GINs) in the top view to capture PPI network topology. However, as the number of (potentially highly connected) proteins grows, inference in this hierarchical framework becomes prohibitively computationally expensive.

Zaydman et al. [2022] utilize spectral methods to recover latent representations at the node level. Specifically, they assess node-level spectral correlations across three distinct spectral windows chosen to differentiate between phylogenetic, indirect PPI, and direct PPI information. Although the authors do not compare their methods to a deep learning framework, the spectral representation of node-level information offers an intriguing compromise between full molecular graph resolution and amino acid sequence representation.

Lin et al. [2023] develop ESM-2, a protein language model trained on evolutionary-scale datasets of protein sequences. The transformer architecture learns rich representations directly from amino acid sequences by leveraging patterns in evolutionary data across millions of proteins. ESM-2 embeddings capture not only local sequence context but also higher-order structural and functional properties without explicit structural input. This approach suggests that protein sequences alone, when processed through a sufficiently powerful language model, may encode much of the information traditionally extracted from molecular graphs, offering a computationally efficient alternative for node-level representations in PPI prediction tasks.

This paper's objective centers on the question of which node-level representation is sufficiently rich for PPI prediction. We compare three approaches representing different levels of geometric abstraction: the full molecular graph from Gao et al. [2023], lower-dimensional manifold embeddings via multidimensional scaling algorithm (MDS) [Hout et al., 2013], and sequence-derived embeddings from the ESM-2 transformer [Lin et al., 2023]. By implementing a consistent GNN architecture at the inter-protein level across all three node-level representations, we isolate the contribution of geometric information to PPI prediction performance. Figure 1 illustrates our framework.

If reduced representations match or outperform the full graph-based approach, we have evidence that explicit molecular graph topology may be redundant when sufficiently powerful embeddings are available. This has critical implications for the accessibility and scalability of PPI studies, as computational costs scale dramatically with the complexity of node-level representations. Understanding the minimal geometric information required for accurate PPI prediction can guide future method development toward approaches that balance predictive performance with computational efficiency.

Methods

We now turn to the three levels of geometric abstraction. Gao et al. [2023] implemented the full molecular graph in an attentive GNN framework. This rich but computationally intensive approach shall serve as our performance baseline. We summarize this first. Second, we outline our transformer-encoder approach using ESM-2 embeddings. Third, we detail our latent representation approach using MDS.

The Hierarchical GNN Approach in HIGH-PPI

HIGH-PPI models PPIs through a hierarchical graph structure with two views: a bottom inside-of-protein view and a top outside-of-protein view. In the bottom view, proteins are represented as graphs where nodes are amino acid residues and edges connect physically adjacent residues (within 10 Å C_α-C_α distance). Each residue node is characterized by seven physicochemical properties: isoelectric point, polarity, acidity and alkalinity, hydrogen bond acceptor/donor capability, octanol-water partition coefficient, and topological polar surface area.

For a protein with n residues, the protein graph is $g_b = (V_b, A_b, X_b)$ where V_b is the node set, $A_b \in \{0,1\}^{n\times n}$ is the adjacency matrix derived from the contact map, and $X_b \in \mathbb{R}^{n\times 7}$ contains residue-level features. The bottom-view GNN employs two GCN blocks to learn protein representations:

$$H^{(1)} = \mathrm{ReLU}(\tilde{D}^{-1/2}\tilde{A}\tilde{D}^{-1/2}X_b W^{(1)}) \tag{1}$$

where $\tilde{A} = A_b + I_n$, $\tilde{D}$ is the corresponding degree matrix, and $W^{(1)}$ is a learnable weight matrix. A second GCN block and self-attention graph pooling produce a fixed-length protein embedding $x \in \mathbb{R}^{1\times d}$.

In the top outside-of-protein view, proteins become nodes in the PPI graph $g_t = (V_t, A_t, X_t)$, where $A_t \in \{0,1\}^{m\times m}$ encodes known interactions and X_t contains the protein embeddings from the bottom-view GNN. The top-view GNN employs three Graph Isomorphism Network (GIN) blocks to update protein representations via neighborhood aggregation:

$$x_v^{(k)} = \mathrm{ReLU}(\mathrm{MLP}^{(k)}((1+\epsilon)x_v^{(k-1)} + \sum_{u\in\mathcal{N}(v)} x_u^{(k-1)})) \tag{2}$$

For a protein pair (i,j), the final prediction is obtained by concatenating their learned representations and passing through a MLP.

The end-to-end model is trained using multi-task binary cross-entropy loss across all PPI types. This hierarchical architecture allows HIGH-PPI to simultaneously capture intra-protein structural information and inter-protein network topology.

A Transformer-Encoder Approach with ESM-2

The proposed model is a hybrid architecture designed to leverage pre-trained protein language model representations in place of traditional molecular graph-based feature extraction at the node level. The model retains a hierarchical Graph Neural Network (GNN) framework for modeling inter-protein relationships and calculating interaction scores.

Our architecture consists of two main components. First, the protein feature encoder. We utilize the pre-trained ESM-2 model [Lin et al., 2023] to transform raw protein sequences into fixed-length vector representations, replacing the bottom-level intra-protein GNN used in HIGH-PPI [Gao et al., 2023]. Second, the inter-protein interaction module. A multi-layered Graph Attention Network (GAT) [Veličković et al., 2018] operates on the protein interaction graph, where nodes are ESM-generated protein feature vectors. This module learns the propensity for interaction between protein node pairs.

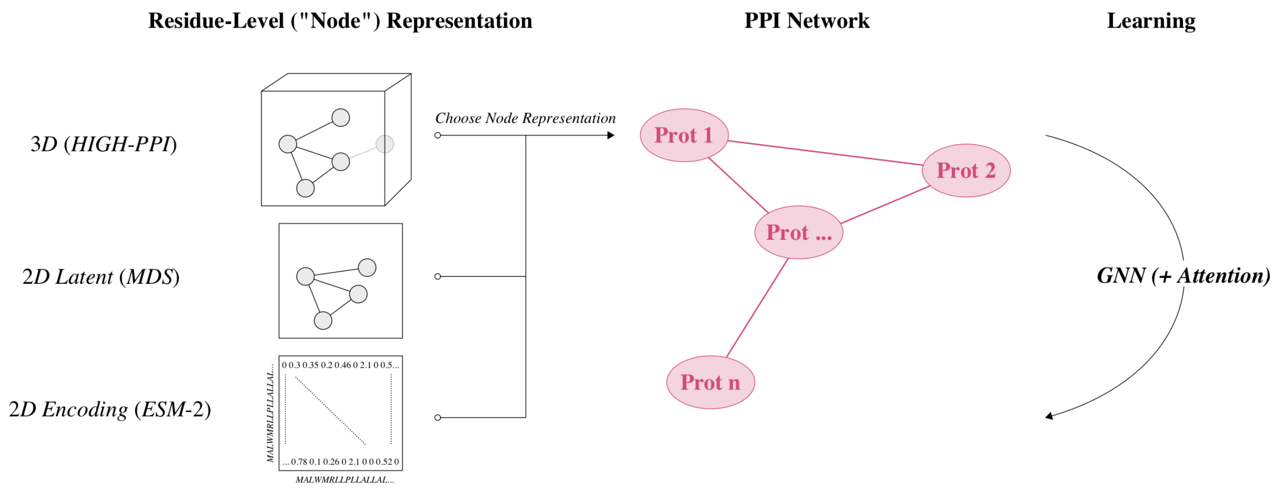

Fig. 1. Overview of the three node-level representation approaches and their integration into the PPI network. The top path shows the full 3D molecular graph (HIGH-PPI), the middle shows the 2D latent manifold representation via MDS, and the bottom shows the 2D encoding from ESM-2 transformer embeddings. Each representation is fed into a consistent architecture for inter-protein link prediction.

To capture structural and functional information embedded within protein sequences, we employ the ESM-2 model (35M parameter version) as the primary feature extractor. Raw protein sequences are tokenized and processed by ESM-2 to obtain residue-level embeddings from the final encoder layer. These residue-level embeddings are pooled to produce a single fixed-dimensional representation $\mathbf{P}_i$ for each protein i.

The interaction graph $G = (V, E)$ is constructed based on known PPIs from the training dataset, where V consists of all proteins and E represents known interactions. Each node $i \in V$ is initialized with its ESM-2 feature vector $\mathbf{P}_i$.

The GAT layers update each protein's feature vector by aggregating information from its neighbors $\mathcal{N}(i)$, weighted by attention coefficients α_{ij}:

$$\mathbf{h}_i^{(k)} = \sigma\left(\sum_{j\in\mathcal{N}(i)\cup\{i\}} \alpha_{ij}^{(k)} \mathbf{W}^{(k)} \mathbf{h}_j^{(k-1)}\right) \tag{3}$$

After the final GAT layer, the learned representations $\mathbf{h}_i^{(K)}$ and $\mathbf{h}_j^{(K)}$ for protein pair (i, j) are combined and passed through a Multi-Layer Perceptron (MLP) to output the interaction score:

$$\hat{y}_{ij} = \mathrm{Sigmoid}(\mathrm{MLP}(\mathbf{h}_i^{(K)} \odot \mathbf{h}_j^{(K)})) \tag{4}$$

where $\odot$ denotes element-wise multiplication, capturing the edge feature representation between source and target nodes.

A Latent Representation Approach via MDS

As an intermediate level of geometric abstraction between full molecular graphs and ESM-2 sequence embeddings, we employ MDS to obtain low-dimensional representations of protein structure. This approach preserves the essential geometric relationships for inter-residue distances while reducing dimensionality.

For each protein, we obtained predicted AlphaFold [Jumper et al., 2021] structures from UniProt. Each structure provides 3D coordinates for all residues, which we represent as a distance matrix $\mathbf{D} \in \mathbb{R}^{n\times n}$, where n is the number of residues and d_{ij} denotes the Euclidean distance between residues i and j. Our goal is to embed these 3D structural relationships into a 2D representation that preserves pairwise distances as faithfully as possible.[1]

For a square, symmetric, positive semi-definite distance matrix $\mathbf{D}$, classical MDS proceeds as follows:

1. Double-center: $\mathbf{B} = -\frac{1}{2}\mathbf{H}\mathbf{D}^2\mathbf{H}$ where $\mathbf{H} = \mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T$
2. Eigendecompose: $\mathbf{B} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$
3. Return coordinates: $\mathbf{X} = \mathbf{U}_k\mathbf{\Lambda}_k^{1/2}$ (top k eigenvectors/-values)

The squared distances in this low-dimensional space optimally approximate the original squared distances. Classical MDS seeks a configuration of points $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]^T \in \mathbb{R}^{n\times p}$ in p dimensions (here $p = 2$) that minimizes the discrepancy between original distances d_{ij} and embedded distances $\|\mathbf{x}_i - \mathbf{x}_j\|$ [Hout et al., 2013]. The objective function, known as the stress criterion, is defined as:

$$\mathrm{Stress}(\mathbf{X}) = \sqrt{\frac{\sum_{i<j}(d_{ij} - \|\mathbf{x}_i - \mathbf{x}_j\|)^2}{\sum_{i<j} d_{ij}^2}} \tag{5}$$

Lower stress values indicate better preservation of the original distance structure. Across our protein dataset, we find that overall stress is relatively low, suggesting reasonable fidelity of the 2D embeddings. However, preservation quality degrades for larger inter-residue distances, which are disproportionately compressed toward zero in the lower-dimensional space. This is an inherent limitation of dimensionality reduction: distances spanning greater Euclidean extent in the original space cannot be faithfully preserved when the target manifold has fewer degrees of freedom. See Appendix **??** for the full stress distribution.

[1] While the dimension reduction from 3D to 2D is modest, our primary objective is to evaluate whether a lower-dimensional representation suffices for PPI prediction. The practical utility of this specific dimensionality choice is secondary to testing our central hypothesis concerning geometric abstraction levels.

We implemented MDS using the `smacof` package in R, which employs iterative majorization algorithms to minimize stress. The resulting 2D coordinates for each residue provide a compact yet geometrically meaningful representation of protein structure, capturing the essential geometry while eliminating potentially redundant structural details. These 2D node-level representations are then used as input features for both inter-protein GNN modules.

Experiments and Results

The experimental setup is designed to systematically evaluate the impact of different architectural choices and hyperparameters on the performance of the Protein-Protein Interaction (PPI) prediction model. The methodology adopts a modular approach, separating the model into components responsible for intra-protein feature generation and inter-protein topological learning. This matrix evaluates **32** distinct configurations.

Model Component Architecture

The model is defined by combining one Intra-Protein module (feature embedding) with one Inter-Protein module (topological aggregation).

Table 1. Experimental Matrix Grouped by Embedding and Batch Size (All 100 Epochs)

Model Components		**Hyperparameters**		**F1**
Intra-Protein	**Inter-Protein**	**Emb. $(D)^a$ Size**	**Batch Size**	**Score**
GCN[b]	**GIN**[c]	**128**	**64**	0.498
			128	0.510
		256	**64**	0.630
			128	0.629
GAT[d]	**GIN**	**128**	**64**	0.635
			128	0.627
		256	**64**	0.614
			128	0.601
ESM-2[e]	**GIN**	**128**	**64**	**0.817**
			128	0.655
		256	**64**	0.794
			128	0.630
ESM-2	**GAT**	**128**	**64**	0.568
			128	0.575
		256	**64**	0.556
			128	0.576

[a] D is the embedding dimension size.

[b] **GCN**: Graph Convolutional Network.

[c] **GIN**: Graph Isomorphism Network.

[d] **GAT**: Graph Attention Network.

[e] **ESM-2**: Evolutionary Scale Modeling, version 2 (Pre-trained Protein Language Model).

[*] Highlighted in red is the original High-PPI setup.

Hyperparameter Variations

The experiment systematically explores three key hyperparameters across all architectural combinations:

1. **Embedding Size** (D): Tested at **128** and **256**. This defines the latent dimension of the node features passed to the final link prediction head. A larger dimension increases model capacity.
2. **Batch Size**: Tested at **64** and **128**. This controls the number of samples processed per iteration.
3. **Epochs**: Tested at **100** and **500**. This explores the necessary training duration, especially for models leveraging complex features like ESM-2, which often require extensive fine-tuning.

The highest priority comparison is the value of high-quality features, seen by evaluating **ESM-2** versus **GCN/GAT** across the same Inter-Protein aggregation module, as this directly addresses the required level of geometric abstraction.

Discussion and Conclusion

The experimental matrix yields two primary findings with critical implications for PPI model design: the indispensable value of rich sequence features** and the failure of the attention mechanism when paired with these features.

Feature Saliency and Efficiency

The data definitively shows that the predictive power for PPIs stems from the feature encoder. The best $\mathbf{ESM-2}$ configuration ($\mathbf{0.817}$) provides a $\mathbf{29.9}\%$ relative improvement over the best $\mathbf{GCN}$ configuration ($\mathbf{0.629}$), confirming that explicit molecular graph topology is largely redundant when powerful sequence-derived embeddings are used. This finding validates the shift away from computationally intensive residue-level GNNs toward feature lookups from pre-trained language models, significantly increasing the scalability of PPI prediction. Furthermore, the optimal $\mathbf{ESM-2}$ configuration used the smaller $\mathbf{D = 128}$ embedding size and $\mathbf{BatchSize = 64}$, demonstrating that efficiency and performance are well-balanced in a compact setup.

GIN vs. GAT: Failure of Attention

While our primary research interest focused on the $\mathbf{ESM-2+}$ $\mathbf{GAT}$ model to analyze the contribution of attention, the results show a large, consistent performance gap: the best $\mathbf{ESM-2+GIN}$ model ($\mathbf{0.817}$) drastically outperformed the best $\mathbf{ESM-2+GAT}$ model ($\mathbf{0.521}$). This result is a key finding for GNN research in the protein domain. It suggests that the GAT attention mechanism is unstable or detrimental when aggregating the highly compressed and abstract features generated by ESM-2. The simpler, permutation-invariant aggregation of GIN is demonstrably more robust for learning the PPI network topology atop these rich embeddings. We hypothesize that GAT may over-smooth the already high-quality ESM-2 features, leading to poorer discriminative ability, whereas GIN successfully preserves local structure. Future work will complete the $\mathbf{500-epoch}$ matrix to fully assess the long-term convergence properties of the best-performing ESM-2 models.

Ultimately, we conclude that the explicit retention of 3D molecular graph geometry is **not necessary** for state-of-the-art PPI prediction. The $\mathbf{ESM-2}$ (intra-protein feature encoder) $\mathbf{+GIN}$(inter-protein topological aggregator) architecture offers the optimal balance of predictive power ($\mathbf{0.817}$ F1-score) and computational efficiency.

References

P. Braun and A.-C. Gingras. History of protein–protein interactions: From egg-white to complex networks. *PROTEOMICS*, 12(10):1478–1498, 2012. doi: https://doi.org/10.1002/pmic.201100563.

Z. Gao, C. Jiang, J. Zhang, X. Jiang, L. Li, P. Zhao, H. Yang, Y. Huang, and J. Li. Hierarchical graph learning for protein-protein interactions. *Nature Communications*, 2023. doi: https://doi.org/10.1038/s41467-023-36736-1.

J. F. Greenblatt, B. M. Alberts, and N. J. Krogan. Discovery and significance of protein-protein interactions in health and disease. *Cell*, 187(23):6501–6517, 2024. doi: https://doi.org/10.1016/j.cell.2024.10.038.

M. C. Hout, M. H. Papesh, and S. D. Goldinger. Multidimensional scaling. *Wiley Interdisciplinary Reviews: Cognitive Science*, 4(1):93–103, 2013. doi: 10.1002/wcs.1203.

J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, and D. Hassabis. Highly accurate protein structure prediction with alphafold. *Nature*, 596:583–589, 2021. doi: 10.1038/s41586-021-03819-2.

Z. Li, A. A. Ivanov, R. Su, V. Gonzalez-Pecchi, Q. Qi, S. Liu, P. Webber, E. McMillan, L. Rusnak, C. Pham, X. Chen, X. Mo, B. Revennaugh, W. Zhou, A. Marcus, S. Harati, X. Chen, M. A. Johns, M. A. White, C. Moreno, L. A. D. Cooper, Y. Du, F. R. Khuri, and H. Fu. The OncoPPi network of cancer-focused protein-protein interactions to inform biological insights and therapeutic strategies. *Nat. Commun.*, 8(1):14356, 2017.

Z. Lin, H. Akin, R. Rao, B. Hie, Z. Zhu, W. Lu, N. Smetanin, R. Verkuil, O. Kabeli, Y. Shmueli, A. dos Santos Costa, M. Fazel-Zarandi, T. Sercu, S. Candido, and A. Rives. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637): 1123–1130, 2023. doi: 10.1126/science.ade2574.

V. S. Rao, K. Srinivas, G. N. Sujini, and G. N. S. Kumar. Protein-protein interaction detection: Methods and analysis. *International Journal of Proteomics*, 2014(1):147648, 2014. doi: https://doi.org/10.1155/2014/147648.

W.-H. Shin, K. Kumazawa, K. Imai, T. Hirokawa, and D. Kihara. Current challenges and opportunities in designing protein-protein interaction targeted drugs. *Adv. Appl. Bioinform. Chem.*, 13:11–25, 2020.

F. Soleymani, E. Paquet, H. Viktor, W. Michalowski, and D. Spinello. Protein-protein interaction prediction with deep learning: A comprehensive review. *Comput. Struct. Biotechnol. J.*, 20:5316–5341, 2022.

P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio. Graph attention networks, 2018.

M. A. Zaydman, A. S. Little, F. Haro, V. Aksianiuk, W. J. Buchser, A. DiAntonio, J. I. Gordon, J. Milbrandt, and A. S. Raman. Defining hierarchical protein interaction networks from spectral analysis of bacterial proteomes. *eLife*, 11: e74104, 2022. doi: 10.7554/eLife.74104.