

RESEARCH ARTICLE

Using machine learning to study protein–protein interactions: From the uromodulin polymer to egg zona pellucida filaments

Luca Jovine 

Department of Biosciences and Nutrition,
Karolinska Institutet, Huddinge, Sweden

Correspondence

Luca Jovine, Department of Biosciences and
Nutrition, Karolinska Institutet, Blickagången
16, SE-141 83 Huddinge, Sweden.
Email: luca.jovine@ki.se

Funding information

Knut and Alice Wallenberg Foundation
project, Grant/Award Number: 2018.0042;
Swedish Research Council project,
Grant/Award Number: 2020-04936

Abstract

Neural network-based models for protein structure prediction have recently reached near-experimental accuracy and are fast becoming a powerful tool in the arsenal of biologists. As suggested by initial studies using RoseTTAFold or the ColabFold implementation of AlphaFold2, a particularly interesting future development will be the optimization of these computational methods to also routinely yield high-confidence predictions of protein–protein interactions. Here I use AlphaFold2 and ColabFold to investigate the activation and polymerization of uromodulin (UMOD)/Tamm-Horsfall protein, a zona pellucida (ZP) module-containing protein whose precursor and filamentous structures have been previously determined experimentally by X-ray crystallography and cryo-EM, respectively. Despite having no knowledge of the UMOD polymer structure (coordinates for which were neither used for model training nor as template), AlphaFold2/ColabFold are able to recapitulate a crucial conformational change underlying UMOD polymerization, as well as the general organization of protein subunits within the resulting filament. This surprising result is achieved by simply deleting from the input sequence a stretch of residues that correspond to a polymerization-inhibiting C-terminal propeptide. By mimicking in silico the activating effect of propeptide dissociation triggered by site-specific proteolysis of the protein precursor, this example has implications for the assembly of egg coat proteins and the many other molecules that also contain a ZP module. Most importantly, it shows the potential of exploiting machine learning not only to accurately predict the structures of individual proteins or complexes, but also to carry out computational experiments replicating specific molecular events.

KEYWORDS

artificial intelligence, protein polymerization, protein–protein interactions, uromodulin, zona pellucida

Abbreviations: CTP, C-terminal propeptide; GDT_TS, global distance test total score; IHP, internal hydrophobic patch; EHP, external hydrophobic patch; EM, electron microscopy; PAE, predicted aligned error; pLDDT, predicted local distance difference test; pTMScore, predicted template modeling score; SP, signal peptide; VE, vitelline envelope; ZP, zona pellucida.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Molecular Reproduction and Development* published by Wiley Periodicals LLC

1 | INTRODUCTION

From mollusk to human, the egg coat (called zona pellucida [ZP] in mammals and vitelline envelope [VE] in non-mammals) is a specialized extracellular matrix that plays key biological roles during oogenesis, fertilization and—in the case of mammals—preimplantation development (Killingbeck & Swanson, 2018; Litscher & Wassarman, 2020). These functions are intrinsically linked to the architecture of the coat, which in turn depends on the assembly of filaments mediated by the “ZP domain,” a bipartite polymerization module conserved in all egg coat subunits as well as many other extracellular proteins with highly different biological functions (Bork & Sander, 1992; Jovine et al., 2002, 2005).

Structural studies of individual ZP/VE subunits revealed that the ZP module consists of two structurally related immunoglobulin-like domains, ZP-N and ZP-C, that are separated by an interdomain linker (Bokhove & Jovine, 2018). ZP module polymerization is activated upon cleavage-dependent dissociation of a C-terminal propeptide (CTP) that includes a polymerization-blocking external hydrophobic patch (EHP) constituting the last β -strand of ZP-C (Han et al., 2010; Jovine et al., 2004; Schaeffer et al., 2009). Recent X-ray and cryo-electron microscopy (EM) structures of ZP module-containing urinary protein uromodulin (UMOD; also known as Tamm–Horsfall protein) revealed the dramatic conformational changes that are triggered by the dissociation of its CTP, as a result of site-specific cleavage by transmembrane protease hepsin (Bokhove et al., 2016; Stanisich et al., 2020; Stsiapanava et al., 2020). However, the magnitude of the observed rearrangements raises the question of whether ZP module filament assembly may also involve polymerization chaperone(s), and it is unclear if equivalent conformational changes take place in the case of heteromeric egg coat filaments.

Here, I use AlphaFold2 (Jumper et al., 2021) and ColabFold (Mirdita et al., 2021) to investigate whether the recent advances in the application of machine-learning to protein structure prediction (AlQuraishi, 2021) can provide insights into these kinds of questions, using ZP module activation as an example of the complex conformational changes that can take place during the assembly of polymeric proteins.

2 | RESULTS

2.1 | Modeling of the UMOD ZP module in its polymerization-inhibited state

As a necessary control, I first used AlphaFold2 to model the C-terminal half of UMOD, whose structure was previously determined by X-ray crystallography (Bokhove et al., 2016). This part of the protein, which encompasses its elastase/pronase-resistant fragment and is sufficient for polymerization (Bokhove et al., 2016; Jovine et al., 2002), consists of an epidermal growth factor domain (EGF IV) that is not involved in subunit/subunit interactions within the UMOD filament (Stsiapanava et al., 2020), the ZP module and, in

the nonpolymeric precursor form of UMOD, the C-terminal EHP-containing propeptide (Figure 1a).

Despite the fact that no experimental structures of UMOD were used as templates during modeling (see Section 4.1) and consistent with the corresponding prediction deposited at EMBL-EBI (Tunyasuvunakool et al., 2021), the top five models produced by AlphaFold2 are in very good agreement with the crystallographic information on the protein; this is both at the level of ZP-N (average root mean square deviation [RMSD] over 106 C α : 2.2 Å) and the region that includes the interdomain linker, ZP-C and the CTP (average RMSD over 180 C α : 2.0 Å). Moreover, although the relative orientations of the ZP-N and ZP-C domains of the monomeric AlphaFold2 models are only approximately similar to those observed in the homodimeric crystal structure (average RMSD over 286 C α for models 1–4: 6.8 Å), the interdomain linkers of all models closely adopt the experimentally observed pre-polymerization conformation, consisting of an α -helix (α 1) and a β -strand (β 1) (average RMSD over 24 C α : 1.1 Å) (Figure 1b,c).

As also indicated by Global Distance Test (GDT_TS) scores of 95.3 (ZP_N) and 88.9 (linker + ZP-C + EHP), these results indicate that AlphaFold2 can accurately model the two moieties of the UMOD ZP module. On the other hand, in agreement with the interchain variability observed in the crystal structure (Bokhove et al., 2016), the relative orientation of the ZP-N and ZP-C domains is less defined.

2.2 | Modeling of the polymerization-activated state of the UMOD ZP module

To investigate the state of the protein activated for polymerization, I then used AlphaFold2 to model a variant of the same region of UMOD that was C-terminally truncated at the hepsin cleavage site and thus lacked the EHP. This resulted in a significantly different set of relative ZP-N/ZP-C orientations, with the top three ranked models having an interdomain linker whose α 1 region converted to a β -strand (α 1 β ′) that faces the internal hydrophobic patch of ZP-C (IHP; an EHP-like element that corresponds to β -strand A and is also involved in polymerization) and pairs with its β -strand F to replace the missing EHP (Figure 2a). Strikingly, this conformational change and intermolecular interaction closely mimics one of the key intermolecular interactions observed in the cryo-EM structure of the UMOD filament (Stsiapanava et al., 2020), so that the ZP-N and ZP-C domains of the activated ZP module can be readily superimposed onto the ZP-N domain of a UMOD subunit and the ZP-C domain of the preceding subunit within the filament (RMSD 1.2 Å over 229 C α ; Figure 2b). A reminiscent but different conformation is instead found in the fourth and fifth ranked models, where the ZP-N/ZP-C linker adopts an extended conformation that also contacts the same region of ZP-C; however, in this case the interaction is mediated by the C-terminal part of the linker, which pairs—in reverse orientation compared to models 1–3—with both β A′ (another ZP-C strand involved in UMOD polymerization) and the beginning of β A/IHP (data not shown).

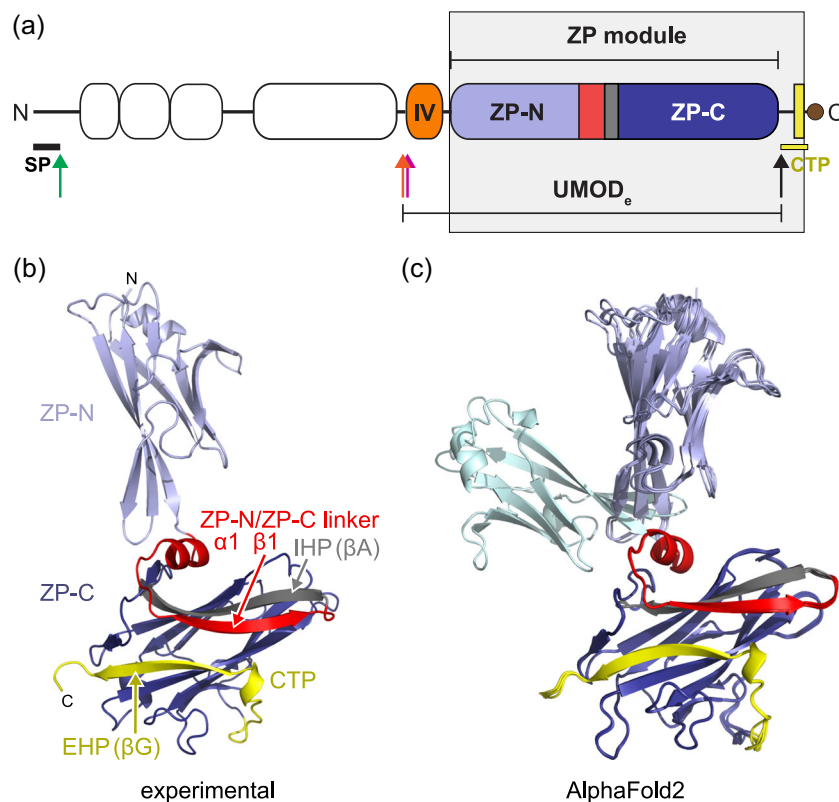


FIGURE 1 AlphaFold2 accurately predicts the structure of the zona pellucida (ZP) module of uromodulin (UMOD) in its polymerization-incompetent state. (a) Scheme of the domain architecture of the human UMOD precursor. Domains within the N-terminal branch region of UMOD, not discussed in this manuscript, are represented by white open rounded rectangles; EGF domain IV is indicated by its roman numeral. The N-terminal signal peptide (SP) and the CTP lost upon precursor processing are indicated by thick horizontal black and yellow lines, respectively, with the external hydrophobic patch (EHP)/ZP-C β -strand G represented by a vertical yellow rectangle and the membrane-anchoring glycosylphosphatidylinositol tail depicted as a brown circle. The ZP-N/ZP-C linker is red, whereas the internal hydrophobic patch (IHP) is dark gray. Vertical green, red, magenta, and black arrows indicate the cleavage sites of signal peptidase, elastase, pronase, and hepsin, respectively. The extents of the ZP module and the elastase-resistant fragment of UMOD (UMOD_e) are indicated, with a light gray rectangle enclosing the part of the protein corresponding to the models shown in (b) and (c). (b) Crystal structure of the ZP module of UMOD and its CTP (Protein Data Bank (PDB) ID 4WRN). For simplicity and comparison with the output of AlphaFold2 (which consists of a single chain), only chain B of the protein homodimer found in the crystal is shown. Elements are colored as in (a). (c) Top five ranked AlphaFold2 models of the UMOD ZP module and CTP, shown as in the previous panel and superimposed over the residues belonging to the interdomain linker, ZP-C and the CTP. The ZP-N domain of model 5, which differs significantly from both models 1-4 and the experimental structure in terms of its orientation relative to ZP-C, is colored light cyan

2.3 | Modeling of the interaction between multiple activated subunits of UMOD

Since the previous computational experiments only included a single molecule of UMOD, I used the homooligomer option of ColabFold (Mirdita et al., 2021) to model a system containing two copies of the activated polymerization region of UMOD. Although the resulting models are significantly different from each other in terms of overall shape, four of them—including the top ranked model—share a common circular arrangement where the ZP-N/ZP-C linker of each subunit is engaged in different types of interactions with the other subunit, all of which mimic contacts observed in the experimental structure of the UMOD polymer (Stsiapanava et al., 2020). These include (1) antiparallel pairing of $\alpha 1\beta'$ of the linker with βF of the

ZP-C domain of the other subunit; (2) parallel pairing of $\alpha 1\beta''$ of the linker with $\beta A''$ of the ZP-C domain of the other subunit; and (3) parallel pairing of $\beta 1$ of the linker with βG of the ZP-N domain of the other subunit (Figure 3a). Notably, none of these linker-mediated interactions was observed upon parallel modeling of a system consisting of two molecules that still contained the EHP (data not shown); without exceptions, this produced models where the interdomain linkers of both subunits adopted the $\alpha 1 + \beta 1$ pre-polymerization conformation shown in Figure 1b,c and made no intermolecular interaction.

While even more variability in the relative position of subunits was observed upon modeling of three cleaved copies of UMOD (data not shown), it is remarkable that one of the resulting models reproduces the general arrangement observed experimentally, with the

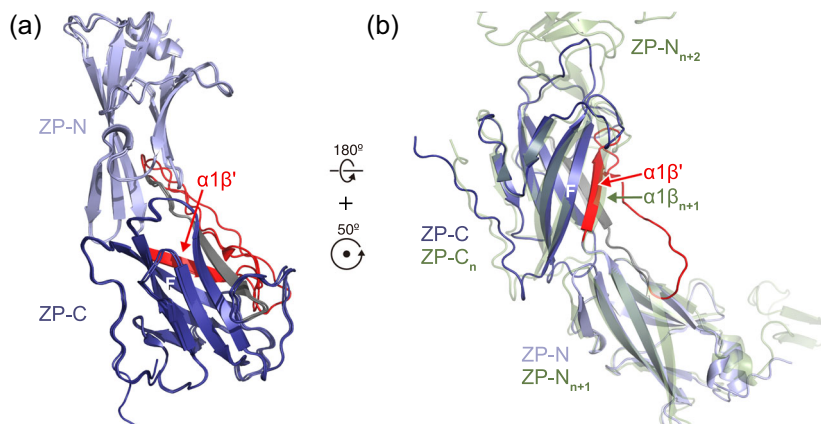


FIGURE 2 Absence of the external hydrophobic patch (EHP) results in a different zona pellucida (ZP) module conformation that mimics a subunit/subunit interaction found in the uromodulin (UMOD) filament. (a) The three highest-ranked AlphaFold2 models of the activated form of the UMOD polymerization region adopt a completely different conformation than that observed in the presence of the EHP-containing propeptide (Figure 1c). The color of the elements of the ZP module and the orientation of its ZP-N domain are the same as in Figure 1b,c. A red arrow indicates the interdomain linker $\alpha 1\beta'$ strand that corresponds to helix $\alpha 1$ in Figure 1b,c. (b) Superimposition of the top model from panel a onto the cryo-EM structure of polymeric UMOD (PDB ID 6TQK; light green). In the UMOD filament, the N-terminal half of the ZP-N/ZP-C linker of a subunit forms a long β -strand ($\alpha 1\beta$) which can also be described as two distinct strands, $\alpha 1\beta'$ and $\alpha 1\beta''$, that respectively pair with βF and $\beta A''$ of the ZP-C domain of the previous subunit. Only the former interaction is recapitulated in the monomeric AlphaFold2 models, due to the fact that their interdomain linker must fold back to connect with the beginning of ZP-C

extended interdomain linker of one molecule wrapping around the ZP-C domain of the preceding subunit and the ZP-N domain of the ensuing one (Figure 3b).

2.4 | Extension to egg coat protein filaments

Since ColabFold is able to produce models that recapitulate part of the main protein–protein interactions stabilizing the UMOD homopolymer, what kind of interactions does it suggest between the different subunits that make up egg coat filaments? To answer this question, I modeled complexes of ZP2 and ZP3, the two major subunits of the mouse ZP that are thought to form heterodimers repeating along the filaments (Litscher & Wassarman, 2020). As shown in Figure 4, the outcome of this prediction essentially mirrored what was observed in the case of UMOD by suggesting that the interdomain linkers of ZP2 and ZP3, which are largely disordered in crystal structures of the individual proteins (Bokhove et al., 2016; Han et al., 2010), also adopt a β -strand conformation that allows them to pair with the ZP-C and ZP-N domains of the adjacent subunits within the ZP filament.

3 | DISCUSSION

The recent developments that culminated with the release of open-source code for both AlphaFold2 (Jumper et al., 2021) and RoseTTAFold (Baek et al., 2021) brought protein structure prediction to a level where it can both rival and facilitate experimental structure

determination. This is highlighted by a growing number of reports that models produced by both systems can be successfully used to phase by molecular replacement native X-ray diffraction data for the corresponding proteins (Baek et al., 2021; Flower & Hurley, 2021; Millán et al., 2021; Pereira et al., 2021), as well as to fit maps obtained by cryo-EM (Baek et al., 2021; Gupta et al., 2021). The availability of a database of high-quality structure predictions for the proteomes of several major experimental systems, including human (Tunyasuvunakool et al., 2021), is bound to significantly expand these and other applications in the near future. At the same time, it will no doubt inform functional studies of a plethora of biological problems for which experimental structural data is either limited or missing. Against this background, a future development of major interest will be the extension of these computational approaches to the reliable prediction of protein–protein interactions. Indeed, even though both the AlphaFold2 and RoseTTAFold networks were originally developed to predict individual protein structures (and, thus, trained on monomeric proteins rather than complexes), RoseTTAFold was already shown to be able to successfully predict a number of known protein complexes (Baek et al., 2021). Moreover, a comparable functionality was added to the ColabFold implementation of AlphaFold2 (Mirdita et al., 2021).

In this study, I explored the possibility of using AlphaFold2 and ColabFold to gain insights into the polymerization mechanism of ZP module-containing proteins, a large family of extracellular molecules with highly variable architecture and biological functions (Bork & Sander, 1992; Jovine et al., 2005). In particular, I focused on UMOD, the most abundant protein in human urine and a major player in the defense against urinary tract infections (Schaeffer et al., 2021),

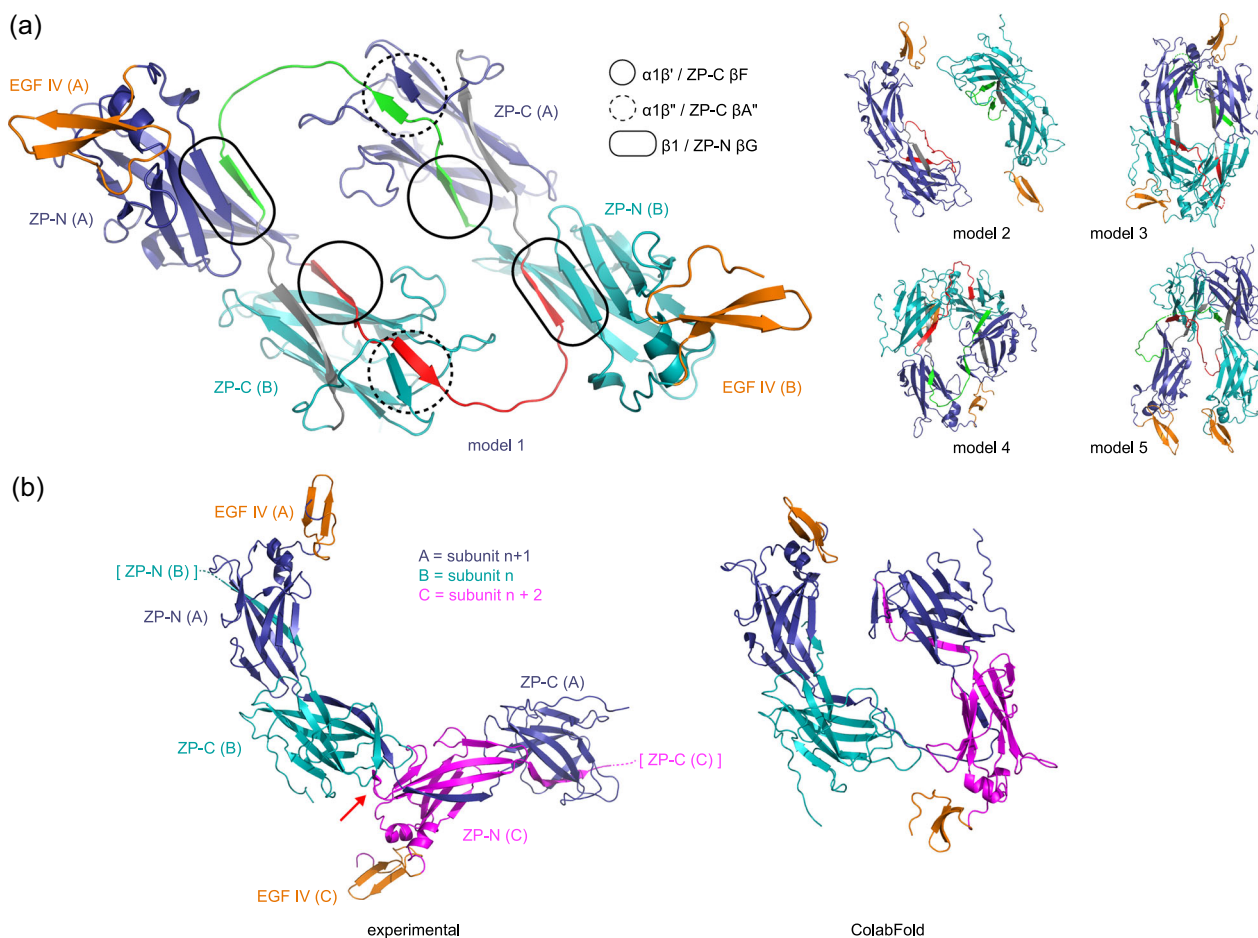


FIGURE 3 ColabFold modeling of multisubunit systems reproduces the three main interactions made by the zona pellucida (ZP)-N/ZP-C linker in the uromodulin (UMOD) filament. (a) Models of two copies of the UMOD polymerization region truncated at the hepsin cleavage site. Except for model 2, which consists of two noninteracting molecules where the ZP-N/ZP-C linker forms the same $\alpha 1\beta'$ /ZP-C βF intramolecular contact shown in Figure 2b, all other models show intermolecular interactions mediated by the interdomain linker. Molecules A and B are colored blue and teal, respectively, with the EGF IV domains shown in orange to clarify the position of the corresponding ZP-N domains. The ZP-N/ZPC linkers are red (molecule A) and green (molecule B), the IHPs are gray. (b) Side-by-side comparison of the cryo-EM structure of native UMOD (PDB ID 6TQK; left) with a three-subunit model generated by ColabFold. Molecules A, B, and C are blue, teal, and magenta, respectively, and their position within the filament is indicated. A red arrow in the left panel indicates the subunit/subunit interface including the two-strand β -sheet made by ZP-N $\beta F'$ and ZPC $\alpha E\beta$, which is not found in the ColabFold model of the complex

because it is the only member of the family for which experimental structural information is available for both precursor and polymeric states of the molecule (Bokhove et al., 2016; Stanisch et al., 2020; Stsiapanava et al., 2020). The outcome of this exercise, however, is not only interesting for what it suggests about UMOD itself, but also for what it implies for other structurally related molecules, such as egg coat proteins. More generally, it supports the idea that—like RoseTTAFold (Baek et al., 2021)—AlphaFold2 has potentialities that go well beyond the prediction of the structure of individual proteins (Jumper et al., 2021; Mirdita et al., 2021).

The results reported in Figures 1c and 2 clearly show that, despite being limited to single-chain prediction, AlphaFold2 was able to at least partially recapitulate (albeit, necessarily, in an intramolecular way) the UMOD conformational change that is triggered by hepsin cleavage-dependent dissociation of the EHP (Bokhove et al., 2016;

Stsiapanava et al., 2020). This is striking, considering that—unlike in the case of the precursor state of the protein (whose experimental structure, PDB ID 4WRN, was excluded as a template during modeling but included in the network's training set)—AlphaFold2 has no knowledge of the polymeric structure of UMOD. This is because all currently available cryo-EM structures of the UMOD filament core (PDB IDs 6TQK, 6TQL, 6ZS5, and 6ZYA) were released after April 30, 2018, the maximum release date of the entries used for model training (Jumper et al., 2021); moreover, no structures of other ZP module protein polymers are found in the PDB. From a biological point of view, the model of a single copy of the polymerization region of UMOD in its activated form (Figure 2a) immediately suggests that dissociation of the EHP from the ZP-C domain is sufficient to trigger the conformational change of ZP-N/ZP-C interdomain linker required for filament assembly (Stsiapanava et al., 2020). At the same time, it

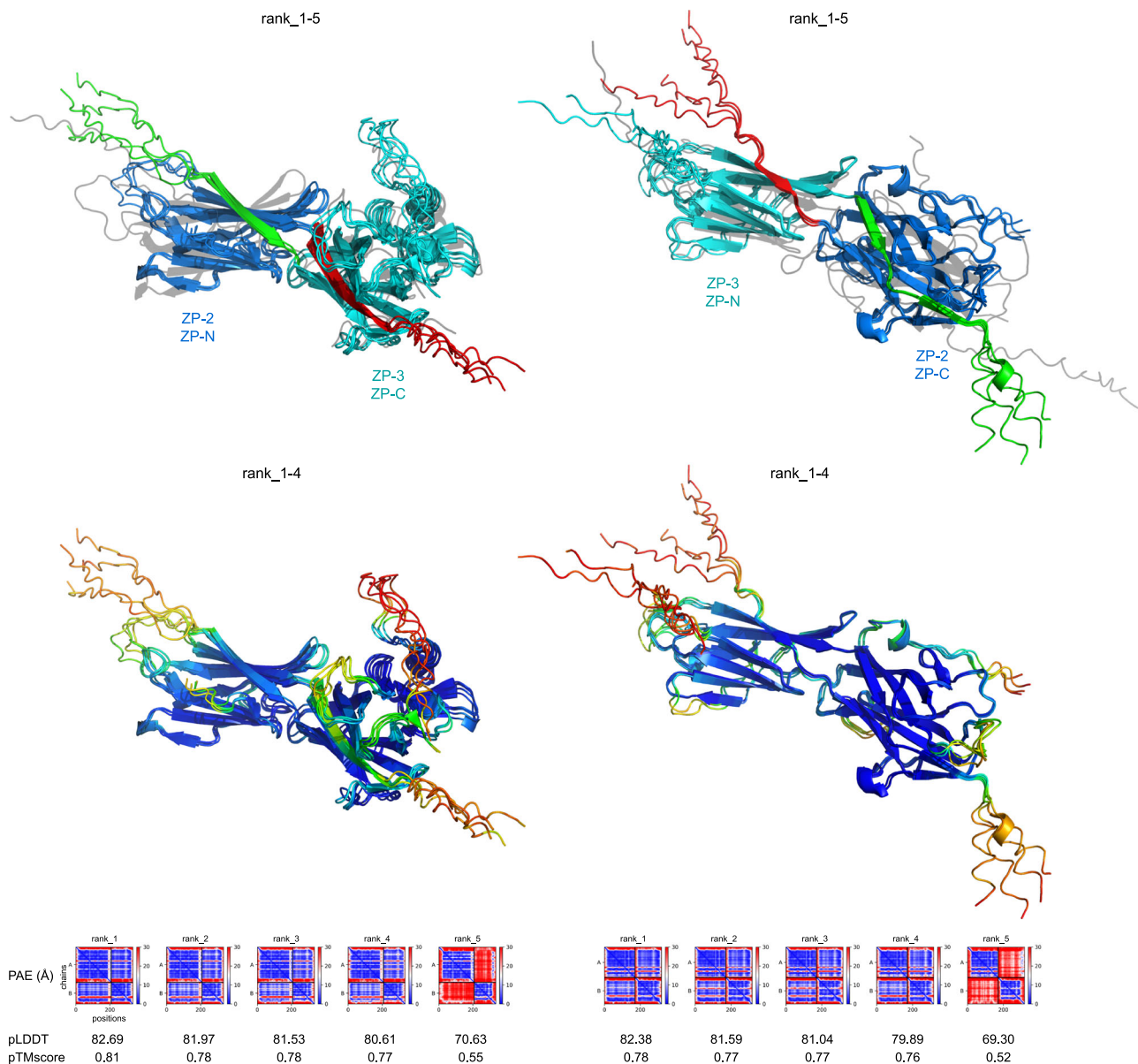


FIGURE 4 ColabFold modeling of ZP2/ZP3 protein complexes consisting of the ZP-N domain plus interdomain linker of one subunit and the linker plus ZP-C domain of the other. The top and middle panels show superimpositions of the top four ranked models, colored by protein (ZP3: blue, with the interdomain linker in red; ZP2: cyan, with the interdomain linker in green) or by prediction confidence (from blue to red according to a 100-(per-residue confidence (pLDDT; Jumper et al., 2021)) scale that ranges from 0 (blue; maximum confidence) to 100 (red; minimum confidence)), respectively. The members of both sets of models are highly similar to each other and show interdomain linker interactions equivalent to those of uromodulin (UMOD). Each set of models 1–4 is also clearly separated from the respective fifth model (rank_5; shown in the top panel as a semi-transparent gray cartoon), which in both cases adopts a significantly different ZP-N/ZP-C conformation that is clearly scored as inferior by all the metrics reported in the bottom panel (Jumper et al., 2021; Mirdita et al., 2021). Note that a low inter predicted aligned error (PAE) between chains indicates a confident prediction; also, only approximately half of each interdomain linker interacts with the adjacent ZP-N/ZPC domain from the other subunit, thus explaining the high structural variability/low pLDDT of the remaining half (which, in the simplified two-domain systems modeled by this prediction, lacks a binding partner)

raises the possibility that EHP dissociation could also lead to a non-productive state, where the interdomain linker undergoes the $\alpha 1$ to $\alpha 1\beta$ conversion but partially interacts with its own molecule rather than other subunits (Figure 2b). Considering that plasma membrane anchoring is required for UMOD polymerization (Brunati et al., 2015), the formation of such a state may be counteracted by the high local

concentration of UMOD precursors at the plasma membrane, and/or the homodimeric ZP-N/ZP-N interactions observed in crystals of the UMOD polymerization region (Bokhove et al., 2016). Alternatively, a state similar to the one modeled in Figure 2a may exist as a transient polymerization intermediate that is eventually replaced by the much more extensive interactions found in the UMOD filament

(Stsiapanava et al., 2020). Notably, such a state would be very difficult to detect and study experimentally, because UMOD expression constructs truncated at the hepsin cleavage site are retained in the endoplasmic reticulum and not secreted (Schaeffer et al., 2009). Regardless of whether this particular form of UMOD truly exists, it is clear that machine-based structure modeling has the potential to unveil protein conformations that may be hard or even impossible to capture experimentally.

When analyzing the models of complexes consisting of multiple copies of the activated ZP module of UMOD, it is important to consider that the current implementation of AlphaFold2/ColabFold does not allow to specify any kind of spatial restraint that may account for the aforementioned membrane-anchoring requirement for filament assembly. Despite this, ColabFold was able to generate models that not only include all the main interactions made by the ZP-N/ZP-C linker of polymeric UMOD (Figure 3a), but also recapitulate the basic subunit organization observed in the filament (Figure 3b). Clearly, these filament fragment models remain quite far from the corresponding experimental structure by missing additional contacts between subunits (in particular the one involving ZP-N β F' and ZPC α EF β ; Figure 3b) that also contribute to the unique overall conformation of the UMOD polymer (Stsiapanava et al., 2020). However, they do provide a valuable structural framework that—in the absence of an experimental structure—could either complement pre-existing biochemical and/or functional data, or be used to derive new hypotheses that could then be experimentally tested. Along these lines, the ZP2/ZP3 complex models shown in Figure 4 are, for example, entirely consistent with mass spectrometric and negative stain EM data suggesting that the basic architecture of VE/ZP filaments resembles that of UMOD (Stsiapanava et al., 2020). Together with further structural studies, these very different but complementary approaches should finally allow the development of a high-resolution model of the vertebrate egg coat, a prerequisite to fully understand the crucial functions played by this extracellular matrix in reproduction.

4 | MATERIALS AND METHODS

4.1 | Modeling of the UMOD ZP module with and without CTP

Models of monomeric UMOD were generated using a local copy of AlphaFold2 (Jumper et al., 2021), installed using the open-source code and instructions available at <https://github.com/deepmind/alphafold>. Runs were performed on an Ubuntu workstation with 128 GB RAM, 16 CPUs and 2 NVIDIA GeForce GTX 1080 Ti GPUs, using the full_dbs preset and excluding all PDB templates released after 2015-12-31; the latter was essential in order not to use as a template the crystallographic structure of the pronase-resistant fragment of UMOD (PDB ID 4WRN, released on 2016-01-27; Bokhove et al., 2016).

4.2 | Modeling of UMOD and ZP subunit multimers

Modeling of multiple copies of UMOD and ZP subunit complexes was carried out using the homo- and hetero-oligomer options of ColabFold (Mirdita et al., 2021), respectively. This notebook (<https://github.com/sokrypton/ColabFold>) uses multiple sequence alignments produced by MMseqs2 (Mirdita et al., 2019) as input for template-free structure prediction by AlphaFold2.

4.3 | Structure analysis and comparison

Structures were visualized, inspected, and superimposed using PyMOL (Schrödinger, LLC), which was also used to make all figures. GDT_TS scores were calculated using the AS2TS server (Zemla, 2003).

ACKNOWLEDGMENTS

I thank Kathryn Tunyasuvunakool and John Jumper (DeepMind) for discussions about AlphaFold2, Sergey Ovchinnikov (Harvard University) for information on ColabFold, Benjamin Wiseman and Alena Stsiapanava (Karolinska Institutet) for comments. This study was supported by the Knut and Alice Wallenberg Foundation project grant 2018.0042 and the Swedish Research Council project grant 2020-04936.

CONFLICT OF INTERESTS

The authors declare that there are no conflict of interests.

ORCID

Luca Jovine  <http://orcid.org/0000-0002-2679-6946>

REFERENCES

- AlQuraishi, M. (2021). Machine learning in protein structure prediction. *Current Opinion in Chemical Biology*, 65, 1–8.
- Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., Wang, J., Cong, Q., Kinch, L. N., Schaeffer, R. D., Millán, C., Park, H., Adams, C., Glassman, C. R., DeGiovanni, A., Pereira, J. H., Rodrigues, A. V., van Dijk, A. A., Ebrecht, A. C., ... Baker, D. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. *Science*, 373, 871–876.
- Bokhove, M., & Jovine, L. (2018). Structure of zona pellucida module proteins. *Current Topics in Developmental Biology*, 130, 413–442.
- Bokhove, M., Nishimura, K., Brunati, M., Han, L., de Sanctis, D., Rampoldi, L., & Jovine, L. (2016). A structured interdomain linker directs self-polymerization of human uromodulin. *Proceedings of the National Academy of Sciences of the United States of America*, 113, 1552–1557.
- Bork, P., & Sander, C. (1992). A large domain common to sperm receptors (Zp2 and Zp3) and TGF- β type III receptor. *FEBS Letters*, 300, 237–240.
- Brunati, M., Perucca, S., Han, L., Cattaneo, A., Consolato, F., Andolfo, A., Schaeffer, C., Olinger, E., Peng, J., Santambrogio, S., Perrier, R., Li, S., Bokhove, M., Bachi, A., Hummler, E., Devuyt, O., Wu, Q., Jovine, L., & Rampoldi, L. (2015). The serine protease hepsin mediates urinary secretion and polymerisation of Zona Pellucida domain protein uromodulin. *eLife*, 4, e08887.

- Flower, T. G., & Hurley, J. H. (2021). Crystallographic molecular replacement using an in silico-generated search model of SARS-CoV-2 ORF8. *Protein Science: A Publication of the Protein Society*, 30, 728–734.
- Gupta, M., Azumaya, C. M., Moritz, M., Pourmal, S., Diallo, A., Merz, G. E., & Verba, K. A. (2021). CryoEM and AI reveal a structure of SARS-CoV-2 Nsp2, a multifunctional protein involved in key host processes. *bioRxiv*. <https://doi.org/10.1101/2021.05.10.443524>
- Han L., Monné, M., Okumura, H., Schwend, T., Cherry, A. L., Flot, D., Matsuda, T., & Jovine, L. (2010). Insights into egg coat assembly and egg-sperm interaction from the X-ray structure of full-length ZP3. *Cell*, 143, 404–415. <https://pubmed.ncbi.nlm.nih.gov/20970175>
- Jovine, L., Darie, C. C., Litscher, E. S., & Wassarman, P. M. (2005). Zona pellucida domain proteins. *Annual Review of Biochemistry*, 74, 83–114.
- Jovine, L., Qi, H., Williams, Z., Litscher, E., & Wassarman, P. M. (2002). The ZP domain is a conserved module for polymerization of extracellular proteins. *Nature Cell Biology*, 4, 457–461.
- Jovine, L., Qi, H., Williams, Z., Litscher, E. S., & Wassarman, P. M. (2004). A duplicated motif controls assembly of zona pellucida domain proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 101, 5922–5927.
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., ... Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596, 583–589.
- Killingbeck, E. E., & Swanson, W. J. (2018). Egg coat proteins across metazoan evolution. In E. S. Litscher, & P. M. Wassarman (Eds.), *Current topics in developmental biology* (Vol. 130, pp. 443–488). Academic Press.
- Litscher, E. S., & Wassarman, P. M. (2020). Zona pellucida proteins, fibrils, and matrix. *Annual Review of Biochemistry*, 89, 695–715.
- Millán, C., Keegan, R. M., Pereira, J., Sammito, M. D., Simpkin, A. J., McCoy, A. J., Lupas, A. N., Hartmann, M. D., Rigden, D. J., & Read, R. J. (2021). Assessing the utility of CASP14 models for molecular replacement. *Proteins*. <https://doi.org/10.1002/prot.26214>
- Mirdita, M., Ovchinnikov, S., & Steinegger, M. (2021). ColabFold: Making protein folding accessible to all. *bioRxiv*, 8, 9939. <https://doi.org/10.1101/2021.08.15.456425>
- Mirdita, M., Steinegger, M., & Söding, J. (2019). MMseqs2 desktop and local web server app for fast, interactive sequence searches. *Bioinformatics*, 35, 2856–2858. <https://pubmed.ncbi.nlm.nih.gov/30615063/>
- Pereira, J., Simpkin, A. J., Hartmann, M. D., Rigden, D. J., Keegan, R. M., & Lupas, A. N. (2021). High-accuracy protein structure prediction in CASP14. *Proteins*. <https://doi.org/10.1002/prot.26171>
- Schaeffer, C., Devuyt, O., & Rampoldi, L. (2021). Uromodulin: Roles in health and disease. *Annual Review of Physiology*, 83, 477–501.
- Schaeffer, C., Santambrogio, S., Perucca, S., Casari, G., & Rampoldi, L. (2009). Analysis of uromodulin polymerization provides new insights into the mechanisms regulating ZP domain-mediated protein assembly. *Molecular Biology of the Cell*, 20, 589–599.
- Stanisich, J. J., Zyla, D. S., Afanasyev, P., Xu, J., Kipp, A., Olinger, E., Devuyt, O., Pilhofer, M., Boehringer, D., & Glockshuber, R. (2020). The cryo-EM structure of the human uromodulin filament core reveals a unique assembly mechanism. *eLife*, 9, e60265.
- Stsiapanava, A., Xu, C., Brunati, M., Zamora-Caballero, S., Schaeffer, C., Bokhove, M., Han, L., Hebert, H., Carroni, M., Yasumasu, S., Rampoldi, L., Wu, B., & Jovine, L. (2020). Cryo-EM structure of native human uromodulin, a zona pellucida module polymer. *The EMBO Journal*, 39, e106807.
- Tunyasuvunakool, K., Adler, J., Wu, Z., Green, T., Zielinski, M., Žídek, A., Bridgland, A., Cowie, A., Meyer, C., Laydon, A., Velankar, S., Kleywegt, G. J., Bateman, A., Evans, R., Pritzel, A., Figurnov, M., Ronneberger, O., Bates, R., Kohl, S., ... Hassabis, D. (2021). Highly accurate protein structure prediction for the human proteome. *Nature*, 596, 590–596.
- Zemla, A. (2003). LGA: A method for finding 3D similarities in protein structures. *Nucleic Acids Research*, 31, 3370–3374.

How to cite this article: Jovine, L. (2021). Using machine learning to study protein–protein interactions: From the uromodulin polymer to egg zona pellucida filaments. *Mol Reprod Dev*, 88, 686–693. <https://doi.org/10.1002/mrd.23538>