Received: 8 September 2019 | Accepted: 3 November 2019
DOI: 10.1002/prot.25845

RESEARCH ARTICLE

PROTEINS WILEY

Template-based modeling of diverse protein interactions in CAPRI rounds 38-45

Justas Dapkūnas | Visvaldas Kairys | Kliment Olechnovič | Česlovas Venclovas

Institute of Biotechnology, Life Sciences Center, Vilnius University, Vilnius, Lithuania

Correspondence
Česlovas Venclovas, Institute of Biotechnology, Life Sciences Center, Vilnius University, Saulėtekio 7, LT-10257 Vilnius, Lithuania.
Email: ceslovas.venclovas@bti.vu.lt

Funding information
Research Council of Lithuania, Grant/Award Number: S-MIP-17-60

Peer Review
The peer review history for this article is available at https://publons.com/publon/10.1002/prot.25845.

Abstract

Structures of proteins complexed with other proteins, peptides, or ligands are essential for investigation of molecular mechanisms. However, the experimental structures of protein complexes of interest are often not available. Therefore, computational methods are widely used to predict these structures, and, of those methods, template-based modeling is the most successful. In the rounds 38-45 of the Critical Assessment of PRediction of Interactions (CAPRI), we applied template-based modeling for 9 of 11 protein-protein and protein-peptide interaction targets, resulting in medium and high-quality models for six targets. For the protein-oligosaccharide docking targets, we used constraints derived from template structures, and generated models of at least acceptable quality for most of the targets. Apparently, high flexibility of oligosaccharide molecules was the main cause preventing us from obtaining models of higher quality. We also participated in the CAPRI scoring challenge, the goal of which was to identify the highest quality models from a large pool of decoys. In this experiment, we tested VoroMQA, a scoring method based on interatomic contact areas. The results showed VoroMQA to be quite effective in scoring strongly binding and obligatory protein complexes, but less successful in the case of transient interactions. We extensively used manual intervention in both CAPRI modeling and scoring experiments. This oftentimes allowed us to select the correct templates from available alternatives and to limit the search space during the model scoring.

KEYWORDS
docking, homology modeling, model quality assessment, protein-protein interactions, scoring, template-based modeling

1 | INTRODUCTION

Proteins often interact with each other and with a variety of ligands to accomplish their functions. Therefore, comprehensive understanding of protein function in many cases requires knowledge of protein interactions at the atomic resolution, which can be derived from the three-dimensional (3D) structures. However, determination of 3D structures is slow, expensive, and not always successful. Computational modeling is often used as an alternative, and structural models are becoming more and more helpful in the studies of protein function.[1]

Computational protein structure prediction is assessed in the Critical Assessment of protein Structure Prediction (CASP) experiments,[2] and structural modeling of protein interactions is evaluated during the Critical Assessment of PRediction of Interactions (CAPRI) experiments.[3-5] The organizers of both experiments provide modelers with the target sequences of proteins whose structure is already solved but not yet publicly available. The modelers are given a limited period of time to predict these structures, and the resulting models are then compared to the experimental structures. Such approach allows blind evaluation of structure modeling methods. Biennial CASP experiments focus on multiple aspects of protein structure prediction. CAPRI

experiments focus on modeling and scoring protein-protein complexes as well as protein complexes with other large molecules such as nucleic acids, peptides, and oligosaccharides.[3,4]

Currently, the most successful method to predict the structures of proteins is template-based modeling, also called homology or comparative modeling.[1] It is based on the observation that structures of homologous proteins are similar. Therefore, a known protein structure can serve as a modeling template for homologous proteins. Since the protein-protein interactions are also largely conserved,[6] homology modeling can be applied to model protein-protein interactions as well. Indeed, recent CAPRI and CASP experiments have confirmed utility of the template-based approach for structural modeling of protein complexes.[5,7-9]

We have successfully utilized template-based modeling during CAPRI round 37, held together with CASP12.[10] During the consecutive CAPRI rounds 38 to 45, we further extended template-based methods for modeling protein-peptide interactions and to some extent for protein-oligosaccharide docking. Additionally, our modeling involved multiple cases where manual intervention proved to be highly beneficial. This underscores the significance of expert knowledge in protein interaction modeling[11] and gives examples of what human experts are actually doing to obtain better structural models.

2 | METHODS

2.1 | Modeling outline

The outline of our workflow is given in Figure 1. In CAPRI rounds 38 to 45, we relied exclusively on template-based modeling methods, utilizing the pipeline developed during CAPRI round 37 (the CASP12-CAPRI experiment)[10] and expanding it to protein-peptide interaction modeling. For modeling protein-oligosaccharide interactions, we used protein-ligand docking methods and constraints from the known protein structures. In addition to template search by PPI3D[12] and HHpred[13] servers, we also searched published data for each CAPRI target using PubServer.[14] We then used all the collected information for both template selection and model evaluation.

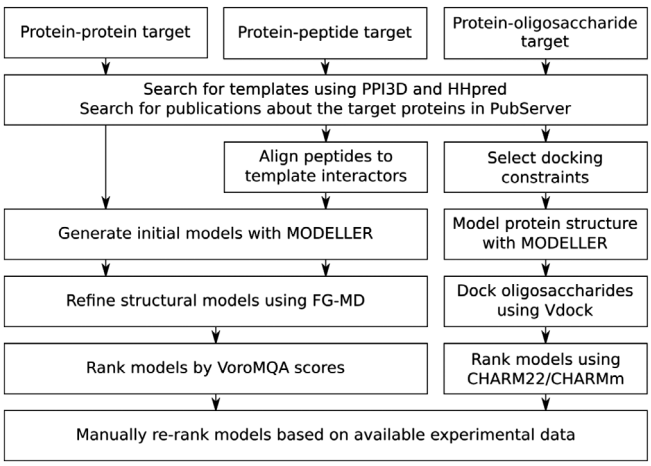

FIGURE 1 Modeling workflow in CAPRI rounds 38 to 45

2.2 | Modeling of protein-protein interactions

The templates for protein-protein interactions were identified with PPI3D,[12] and initial homology models were generated using MODELLER.[15] The PPI3D server lists only binary protein interactions. Therefore, if necessary, models were rebuilt to represent higher oligomeric states. Also, in some cases, the interfaces from several template structures were combined into a single model. When alternative interaction modes were identified among the templates, initial models were constructed using both interfaces.

Models were refined by improving sequence-structure alignment using HHpred[13,16] and by structural optimization utilizing fragment-guided molecular dynamics.[17] All resulting models were ranked by combining VoroMQA scores for the entire structure and for the interaction interface.[10,18] Finally, the ranking was adjusted according to the published data (if available) on target proteins.

2.3 | Modeling protein-peptide interactions

To model protein-peptide interactions, single-sequence queries were used in PPI3D server,[12] aiming to find all known interactions for the homologs of a target protein. After that, the target peptide sequences were aligned to the interaction partners of the identified homologs, and protein-peptide complexes were produced by MODELLER.[15] The models were subsequently refined in the same way as models for protein-protein interactions.

In the case of T134, the task was to predict which region (peptide) of the 57-residue long MAG protein binds to DLC8 (receptor). We used several methods to accomplish this task. First, we downloaded all the sequences of peptides in the identified templates and clustered them at 95% sequence identity.[19] We then generated multiple sequence alignment comprising clustered peptides and the target sequence by MAFFT[20] to reveal a similar sequence region in MAG. Additionally, we analyzed the conservation of the MAG protein, aiming to identify conserved sequence fragments that potentially could interact with the receptor. To this end, we searched for MAG homologs in the UniProt database using jackhmmer,[21,22] and then visually analyzed the resulting WebLogo. We also used structure prediction methods: all possible 12-residue peptides were derived from the MAG protein, and 30 models were generated for each peptide based on three templates identified by PPI3D and HHpred. The resulting structural models were then ranked using VoroMQA (see below), and ranks were summed up for each possible peptide. The final set of 10 models was selected based on a combination of sequence-based predictions and the VoroMQA scores for structural models.

In the case of T135, we already knew the sequence of the peptide and generated models semi-automatically. We used the same three templates as for T134, and shifted the peptide sequence systematically from −2 to +2 position compared to the peptide in the template

structure, generating five models for each position. This procedure resulted in a pool of 75 models, plus some additional models after combining two templates and structure refinement. All these models were then ranked automatically using VoroMQA and submitted according to the ranked order.

2.4 | Modeling protein-oligosaccharide interactions

For the protein-oligosaccharide interaction targets, the protein structures were modeled based on sugar-bound templates, and the constraints for ligand docking were also extracted from these structures. The templates were identified by HHpred,[13,16] and structural models of the target proteins were constructed by MODELLER.[15] To dock oligosaccharide molecules, we used Vdock (Verachem LLC, Germantown, Maryland), also known as the Mining Minima optimizer.[23,24] For T126-T129, the docking was based on constraints from templates, and final 10 models were selected by automated scoring using CHARM22[25] force field for proteins and CHARMm[26] for sugars, including solvent effects by 4*r* distance-dependent dielectric approximation.[27] The structure of T130 protein with a bound trisaccharide was already known. Therefore, the docking procedure aimed to extend oligosaccharide in the template structure to both sides, making it a pentamer. Final models for T130 were picked by hand, and only one model based on automated scoring was included in the submitted 10-model set.

2.5 | Ranking and selection of structural models of protein complexes

As in previous CAPRI and CASP rounds,[10,11] we employed VoroMQA[18] for model ranking and selection. VoroMQA is based on contact areas and produces scores at atomic, residue, and global levels. Thus, it can directly evaluate a protein-protein interaction interface, defined as contact areas of atoms coming from different protein subunits. We used three VoroMQA scores: (a) the global score, (b) the score of the interface atoms, and (c) the raw pseudo-energy of the interchain interaction. The models were ranked by a tournament procedure utilizing the three scores as described previously.[10,11]

During the CAPRI scoring experiment, for some targets highly diverse models were selected from the given scoring sets by the standard VoroMQA scoring. To avoid this, if there was information regarding residues important for protein-protein interaction in a particular target, we constrained the selection space to include only models that have these residues in the interaction interface.

2.6 | Post-CAPRI analysis

Model quality was evaluated using the standard CAPRI criteria,[28] and the evaluation data were taken from the CAPRI website. When analyzing the scoring results, we additionally evaluated the scoring sets and our selected models using CAD-score.[29] We used two CAD-score variants, the interface CAD-score and the binding site CAD-score, that were also employed as similarity measures for clustering structural data in the PPI3D server.[12] The interface CAD-score is similar to the fraction of native contacts used by CAPRI organizers and Interface Contacts Similarity (ICS) used by CASP assessors.[8] A more permissive variant, the binding site CAD-score, does not take into account the contacts across the interface and evaluates only residue contact areas, thus the precision of the binding site, similarly to Interface Patch Similarity (IPS), used in CASP.[8] CAD-score values were calculated using the Voronota package.[30]

3 | RESULTS AND DISCUSSION

3.1 | Results overview

The overview of our results is given in Table 1. We identified templates for six of eight protein-protein interaction targets (T122, T125, T131, T132, T133, and T136) using the PPI3D web server. Two targets, T123 and T124, had no templates even for one of the monomers; as a result, no template-based models could be generated for the corresponding protein complexes. The templates were detected for all three protein-peptide interaction targets (T121, T134, and T135). Modeling was only partially successful for T121, but high quality models were generated for T134 and T135. We also managed to generate models of at least acceptable quality for most of the protein-oligosaccharide docking targets.

Our structural modeling relied not only on the structures of templates from the Protein Data Bank (PDB),[31] but also on the additional published information available for most of the targets. This information was highly useful in ranking and selecting models both in structure prediction and in scoring experiments. In further sections, we describe our modeling results in detail.

3.2 | Modeling of protein-protein interactions

We generated models of medium or high quality for four of the six template-based protein-protein interaction targets (Table 1), and failed to produce at least acceptable models for T131 (HopQ type I complex with CEACAM1) and T132 (HopQ type II complex with CEACAM1). The monomer structures for the HopQ and CEACAM1 proteins were either known or could be easily obtained by homology modeling.[32,33] However, there were no templates for the hetero-interaction. The N-terminal domain of the CEACAM1 protein has immunoglobulin fold,[33] therefore, we generated models based on HopQ homolog BabA interactions with nanobodies[34] used for stabilization of the protein during X-ray crystallography experiments. These models turned out to be incorrect. Perhaps, better quality models for T131 and T132 could be obtained by docking, as some of the interface residues were already known from the literature[32] and could be used as docking constraints.

Modeling of the other four targets was more successful. Some of them were straightforward homology modeling cases. For example, heterodimeric T133, a re-designed colicin E2 DNase interaction with immunity protein Im2, had multiple templates in PDB. Models were generated by PPI3D and ranked by VoroMQA using a semi-

TABLE 1 Overview of the results

Target	Annotation	Templates	Target PDB ID	Our best model	Best CAPRI model
Protein-protein interactions					
T122 (ABC)	IL23R, IL12B, IL23A	5MJ3, 2D9Q, 1I1R, 1P9M	5MZV	Medium	Medium
T123 (AB)	PorM N-terminus, nb01		6EY0		Incorrect
T124 (A_2B)	PorM C-terminus, nb130		6EY6		Incorrect
T125 (A_2B_4)	LLT1, NKR-P1	4IOP, 4QKH, 5J6G, 3G8K, …	5MGT	High	High
T131 (AB)	HopQ type I, CEACAM1	5F7K, 5LP2, 4QXW, 2GK2, …	6GBG	Incorrect	Medium
T132 (AB)	HopQ type II, CEACAM1	5F7K, 5LP2, 4QXW, 2GK2, …	6GBH	Incorrect	Medium
T133 (AB)	Edes3, Imdes3	7CEI, 3U43, …	6ERE	Medium	Medium
T136 (A_{10})	LdcA	2VYC, 3N75		Medium	Medium
Protein-peptide interactions					
T121	Tol-Pal system protein			Incorrect	Medium
T134	DLC8, MAG	4D07, 1CMI, 5E0L	6GZL	High	High
T135	DLC8, MAG	4D07, 1CMI, 5E0L	6GZL	High	High
Protein-oligosaccharide interactions					
T126	AbnE + A6	5F7V, 3K00		Incorrect	Medium
T127	AbnE + A5	5F7V, 3K00		Acceptable	Medium
T128	AbnE + A4	5F7V, 3K00		Acceptable	Medium
T129	AbnE + A3	5F7V, 3K00		Acceptable	Medium
T130	AbnB + A5	3CU9, 3D5Z	6F1G	High	High

automated pipeline. Although there were conformational changes in the designed protein structure,[35] the overall interaction interface was predicted correctly using available templates (f_{nat} > 0.6), resulting in medium quality models according to CAPRI criteria. Similarly, simple homology modeling lead to medium quality models of the decameric target T136.

Modeling of T122 was a bit more complicated. This heterotrimeric target consisted of interleukin-23 receptor (IL-23R), interleukin-12 subunit beta (IR-12B), and interleukin-23 subunit alpha (IL-23A). IL-23A together with IL-12B form a complex named interleukin-23 (IL-23), and several structures of this protein complex were already known.[36-38] Of these, the one with best resolution (PDB: 5MJ3) was selected for further modeling. Moreover, based on the analysis of IL-23 structure and its interactions with antagonists, a tryptophan residue in the IL-23A subunit (Trp-156 in the CAPRI target sequence) was suggested as a hot-spot, mediating IL-23 interaction with the receptor.[36-38]

Using the PPI3D web server, we identified several templates for the IL-23 interaction with IL-23R (PDB: 1I1R, 1P9M, 2D9Q).[39-41] Interestingly, two binding modes were observed in these structures, usually referred to as site II and site III (site I is the interaction interface between IL-12B and IL-23A; Figure 2). We manually ranked the models that had Trp-156 in the interface to be the first ones in our submission to CAPRI. These four models were of acceptable and medium quality, as Trp-156, occurring at the site III, is indeed important for IL-23R binding.[42] As expected, all other models, corresponding to the alternative interaction mode, were incorrect. Thus, T122 is a nice example of beneficial human input in choosing the correct template from available alternatives by integrating the information from multiple sources. Similar situations are quite common, and we encountered them also in the CASP12-CAPRI and CASP13 experiments.[10,11]

T125 was the most challenging of the protein-protein interaction targets. This was a heterohexamer (A_2B_4) in which lectin-like transcript (LLT-1, protein A in further text) interacts with natural killer cell surface protein P1A (NKR-P1, protein B). Several structures of LLT-1 dimer (A_2) were available in PDB,[43] and the highest resolution structure was chosen for modeling of the protein complex (PDB: 4QKH). Not surprisingly, the A_2 homodimer interface was of high quality in all of our models. A heteromeric template was also available for the LLT-1 and NKR-P1 (AB) interface (PDB: 4IOP).[44] The residues responsible for the interaction between LLT-1 and NKR-P1 were suggested after a mutagenesis study,[45] and based on the conservation of these residues, PDB structure 4IOP has been proposed as a suitable template for modeling the AB interaction.[44] Therefore, it is not surprising that we generated high quality models for the hetero-interface. Based on the A_2 and AB interfaces, the assembly of a heter-tetramer A_2B_2 was straightforward. However, addition of two NKR-P1 (B) subunits turned out to be a non-trivial task, as we had to combine interfaces from A_2B_2 templates with B_2 templates to build a complete hexamer. After all, despite the availability of several dimeric B_2 templates, all the modeled interactions between NKR-P1 subunits were evaluated as incorrect according to CAPRI criteria.

In an attempt to understand why only part of the interaction interfaces of T125 were predicted successfully, we decided to analyze the

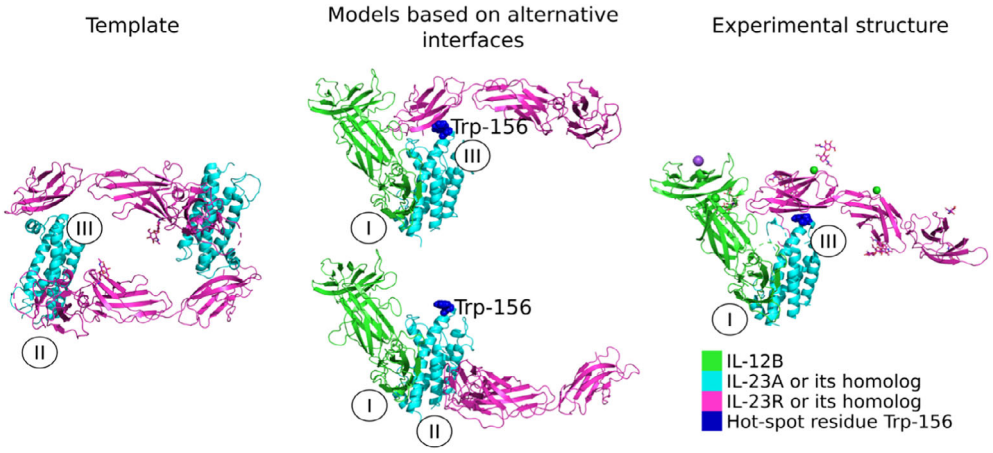

FIGURE 2 Two alternative protein-protein interaction interfaces are available among the templates (e.g. PDB: 2D9Q) for CAPRI target T122, and the site III, containing the hot-spot residue Trp-156, corresponds to the experimental target structure (PDB: 5MZV)

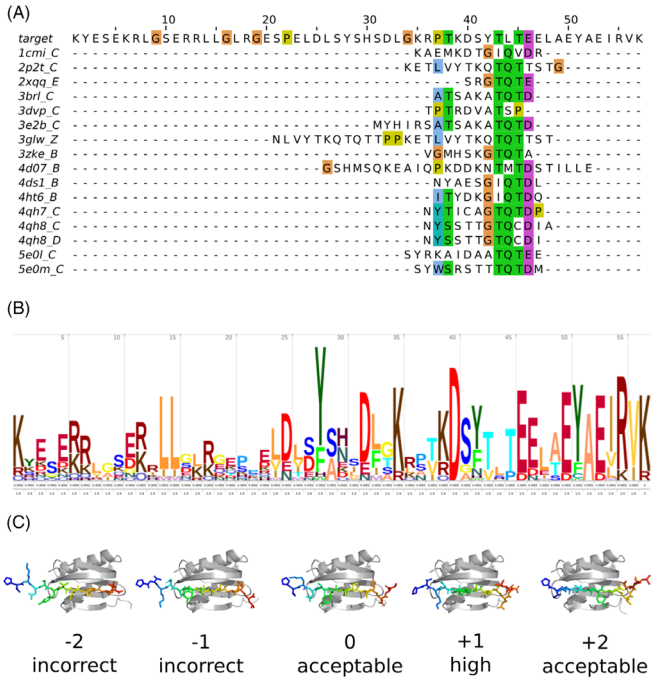

FIGURE 3 Modeling of protein-peptide interactions for T134 and T135. A, peptides from putative templates aligned to the sequence of target protein MAG; B, analysis of the MAG sequence conservation, C, shifting the peptide register in the [−2; +2] range with respect to the template structure 5E0L resulted in models of different quality

experimental structure of T125 (PDB: 5MGT) in detail. No article has been published regarding this structure, and in PDB it is split by the authors into three homodimeric biological assemblies (A_2, B_2, and B_2). PDBePISA[46] suggests two dimeric complexes: a homodimer A_2 and a heterodimer AB. EPPIC-web[47] suggests a tetrameric assembly (A_2B_2). In addition, we checked the interaction interfaces of the asymmetric unit of the PDB structure 5MGT using the VoroMQA server.[48] Some of these interfaces had highly unfavorable energy, suggesting that they are unlikely biologically relevant. Given these inconsistencies, it cannot be excluded that the original definition of the T125 target as a hetero-hexamer having the stoichiometry A_2B_4 may not be biologically relevant. Sometimes it is not possible to make confident assignments of the true biological assemblies from crystal structures. This is a well-known problem that hinders the assessment of performance in modeling of protein assemblies.[9] In our view, T125 is a good illustration of this problem and, perhaps in the absence of additional experimental data, the questionable interaction interfaces should be excluded from the assessment altogether.

3.3 | Modeling of protein-peptide interactions

There were three protein-peptide interaction targets, and we had varying results for them.

For T121, TolA-TolB interaction from *Pseudomonas aeruginosa*, we generated homology models based on *Escherichia coli* TolA bound to different viral peptides (PDB: 1TOL, 2X9A), and TolB is suggested to bind to the same site.[49] In addition, several binding site residues were also known.[50] However, all our models had only a small part of contacts predicted, and were evaluated as incorrect according to CAPRI criteria. Unfortunately, the experimental structure of T121 was not available at the time of writing, only a structural cartoon has been published in a PhD thesis.[51] According to it, we correctly identified the binding site in the TolA protein, but incorrectly predicted the inter-chain β-sheet: the TolB peptide forms a parallel β-sheet, while we predicted antiparallel orientation based on the viral templates.

In the cases of T134 and T135, our modeling was more successful. The task for T134 was to predict which peptide of a 57-residue long MAG protein binds to DLC8 (receptor). The structures were already known for the receptor and its homologs as well as for their complexes with different peptides.[52-55] When we aligned all the peptides from templates identified by PPI3D to the target sequence, a possible binding site emerged near the C-terminus of the MAG protein (Figure 3A). A similar region was also identified based on the residue conservation analysis (Figure 3B). Therefore, we constrained the selection of structural models to the C-terminus of the MAG protein, and included only two models where the binding peptide was in the N-terminal side, both of which were incorrect. Seven of eight submitted C-terminal peptide models were of acceptable or higher quality, indicating that our approach of combining sequence and structure analysis worked reasonably well.

In the case of T135, the peptide sequence was supplied. We varied the position of the peptide in the models of protein complexes (Figure 3C), ranked models using VoroMQA scores, and submitted the models according to this ranking. Only a fraction of the 10 top ranked models were of acceptable or better quality, indicating possible issues with model ranking.

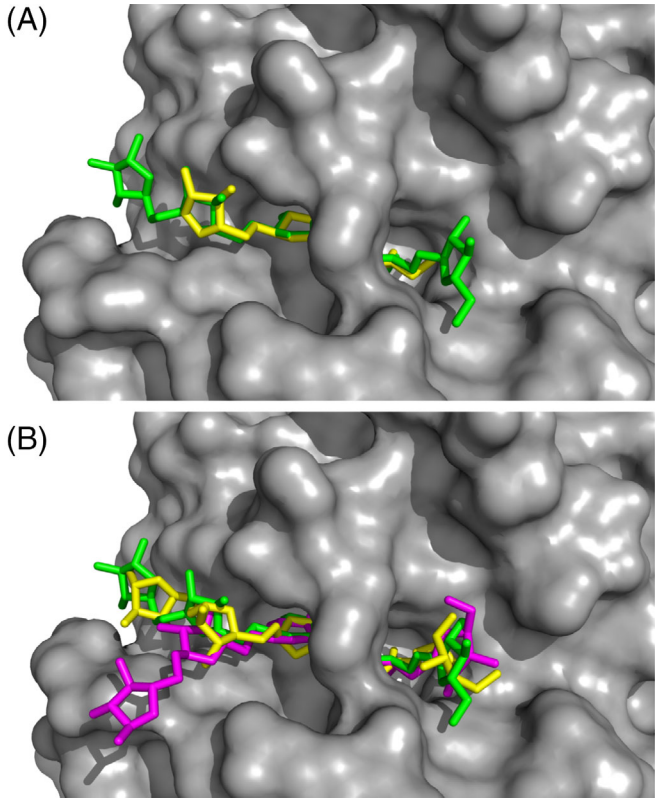

FIGURE 4 Template-based modeling of protein-oligosaccharide interactions in T130. A, extending the template sugar molecule (template: yellow, model: green); B, comparison of the experimental sugar conformation (yellow, PDB: 6F1G) to manually chosen docking pose (high-quality model, green) and automatically selected one (medium-quality model, magenta)

3.4 | Modeling of protein-oligosaccharide interactions

We also utilized template-based approach for modeling protein-oligosaccharide interactions. For targets T126-T129, we derived constraints from sugar-bound templates. Application of template-based modeling with these constraints resulted in models of varying quality. For T130, the structure of a mutant protein bound to a trisaccharide was already known, and we simply extended the oligosaccharide in this molecule to get a pentamer, obtaining a set of high-quality models.

In the case of T126-T129, the target protein AbnE was modeled according to sugar-bound templates (PDB: 5F7V, 3K00), chosen from HHpred results. Not much information was available about the AbnE protein; therefore, in model selection, we had to rely on automated scoring. The sugar molecules were even rotated by 180° in some of the top-ranked models, thus it is not surprising that only part of our models were acceptable according to CAPRI criteria. Predictably, the shorter was the sugar molecule, the higher number of acceptable models was generated. Yet, the overall results for T126-T129 are not impressive.

In case of T130, the structure of the target protein AbnB bound to a trisaccharide was known (PDB: 3CU9, 3D5Z).[56] The docking procedure aimed to produce a pentasaccharide by extending the sugar molecule in this template structure to each side (Figure 4A). Then, we manually selected nine models, in which the sugar was bound in the groove forming the active site of the enzyme,[56] and one model was selected based on automatic scoring. According to CAPRI criteria, the hand-picked models were of high quality, while the automatically chosen one was only of medium quality because the oligosaccharide was only partially bound to the groove in the protein molecule (Figure 4B).

TABLE 2 Comparison of our best selected models and the best available models in CAPRI scoring sets for protein-protein and protein-peptide interactions

	Best selected models			Best models in the scoring set	
Target	CAPRI category	Interface CAD-score	Binding site CAD-score	Interface CAD-score	Binding site CAD-score
T121[a]	Incorrect				
T122[b]	Acceptable	0.40	0.55	0.45	0.63
T123	Incorrect	0.03	0.17	0.06	0.33
T124	Incorrect	0.02	0.32	0.03	0.36
T125[b,c]	High/medium	0.23	0.35	0.30	0.41
T131[b]	Acceptable	0.11	0.48	0.11	0.49
T132	Incorrect	0.01	0.16	0.31	0.61
T133	Medium	0.51	0.69	0.59	0.73
T134	High	0.74	0.81	0.75	0.82
T135	High	0.73	0.84	0.77	0.87
T136[a]	Medium				

[a]Target structure was not available at the time of writing, therefore, CAD-scores were not calculated.
[b]Constraints were used when selecting models from CAPRI scoring sets.
[c]CAPRI evaluation is given for the two best predicted interfaces.

Our results for protein-oligosaccharide targets indicate several docking issues. Carbohydrates are highly flexible molecules, and their interactions with proteins involve not only electrostatic interactions and complex H-bonding patterns, but also hydrophobic interactions.[57,58] This makes protein-oligosaccharide docking more challenging compared to other small molecules. Our scoring methods also did not correctly recognize the bound oligosaccharide conformation. Specialized force fields have been suggested for carbohydrates,[58] which improve protein-sugar interaction modeling.[59] Including these force fields probably might have improved our modeling results as well. In case of T130, we at least partially solved docking and scoring problems by picking models by hand, but it was not possible for T126-T129, where information about the target protein was sparse. As a result, more advanced docking methods and scoring functions might have been beneficial for these targets.

3.5 | Scoring of models of protein complexes using VoroMQA

The results of the CAPRI scoring experiment are summarized in Table 2. We selected medium quality models for homomeric T136, as well as medium and high quality models for strongly interacting heteromers T133, T134, and T135.[35,60] CAD-score values in the table also show that the selected models were close to the best possible models in the CAPRI scoring sets. On the other hand, we had problems in automated selection of models of transient interactions, such as T122, T125, T131, and T132. When we knew the interacting residues from previously published experiments, we constrained the selection using VoroMQA to the set of models where these known residues were in the interaction interface. This helped us to choose at least some models of better quality for these targets.

Taken together, the results of the CAPRI scoring experiment illustrate that the currently used VoroMQA procedure works quite well on strong and obligatory interactions, but encounters issues with weak and transient interactions. This should not be surprising if we consider that VoroMQA was developed for scoring models for protein folding and has not been tailored for scoring of protein-protein interfaces. Of course, the set of CAPRI targets is too small and too diverse to draw broad conclusions. In a larger-scale CASP13 experiment, we found out that a combination of only two VoroMQA scores, the global score and the interface energy, works better than the combination of the three scores used here for CAPRI targets.[11] The selection of models for protein complexes might be further improved by including other types of interface scores, such as those using evolutionary information.[61-63]

4 | CONCLUSIONS

Our results show that the template-based methods are applicable for modeling diverse protein interactions. The PPI3D web server proved to be highly useful in finding and analyzing the templates for protein-protein and protein-peptide interactions. The ability of PPI3D to identify alternative protein interaction interfaces present in some templates facilitated selection of the correct binding modes based on the prior knowledge of the interface residues. The human expert evaluation was also beneficial in integrating the data from multiple sources during the modeling of protein-peptide and protein-oligosaccharide interactions. Scoring of models of protein complexes using VoroMQA facilitated the selection of higher quality models, but for targets involving transient interactions, the selection had to be constrained by taking into account known interface residues. Altogether, our results show that the human influence was highly valuable for most of the CAPRI targets. Thus, significant advancements of the current modeling and scoring methods are needed to fully automate the template-based modeling of protein complexes.

ACKNOWLEDGMENTS

We wish to thank structural biologists who provided structures of protein complexes as prediction targets. We also wish to thank the CAPRI organizers for providing the platform for the assessment of modeling and scoring methods.

ORCID

Justas Dapkūnas https://orcid.org/0000-0002-0496-6107
Visvaldas Kairys https://orcid.org/0000-0002-5427-0175
Kliment Olechnovič https://orcid.org/0000-0003-4918-9505
Česlovas Venclovas https://orcid.org/0000-0002-4215-0213

REFERENCES

1. Schmidt T, Bergner A, Schwede T. Modelling three-dimensional protein structures for applications in drug design. *Drug Discov Today*. 2014;19(7):890-897.
2. Moult J, Fidelis K, Kryshtafovych A, Schwede T, Tramontano A. Critical assessment of methods of protein structure prediction (CASP)-round XII. *Proteins*. 2018;86(suppl 1):7-15.
3. Lensink MF, Wodak SJ. Docking and scoring protein interactions: CAPRI 2009. *Proteins*. 2010;78(15):3073-3084.
4. Lensink MF, Wodak SJ. Docking, scoring, and affinity prediction in CAPRI. *Proteins*. 2013;81(12):2082-2095.
5. Lensink MF, Velankar S, Wodak SJ. Modeling protein-protein and protein-peptide complexes: CAPRI 6th edition. *Proteins*. 2017;85(3):359-377.
6. Aloy P, Ceulemans H, Stark A, Russell RB. The relationship between sequence and interaction divergence in proteins. *J Mol Biol*. 2003;332(5):989-998.
7. Lensink MF, Velankar S, Baek M, Heo L, Seok C, Wodak SJ. The challenge of modeling protein assemblies: the CASP12-CAPRI experiment. *Proteins*. 2018;86(suppl 1):257-273.
8. Lafita A, Bliven S, Kryshtafovych A, et al. Assessment of protein assembly prediction in CASP12. *Proteins*. 2018;86(suppl 1):247-256.
9. Guzenko D, Lafita A, Monastyrskyy B, Kryshtafovych A, Duarte JM. Assessment of protein assembly prediction in CASP13. *Proteins*. 2019;87:1190-1199.
10. Dapkūnas J, Olechnovič K, Venclovas Č. Modeling of protein complexes in CAPRI round 37 using template-based approach combined with model selection. *Proteins*. 2018;86(suppl 1):292-301.

11. Dapkūnas J, Olechnovič K, Venclovas Č. Structural modeling of protein complexes: current capabilities and challenges. *Proteins*. 2019;87:1222-1232.
12. Dapkūnas J, Timinskas A, Olechnovič K, Margelevičius M, Dičiūnas R, Venclovas Č. The PPI3D web server for searching, analyzing and modeling protein-protein interactions in the context of 3D structures. *Bioinformatics*. 2017;33(6):935-937.
13. Zimmermann L, Stephens A, Nam S-Z, et al. A completely reimplemented MPI bioinformatics toolkit with a New HHpred server at its Core. *J Mol Biol*. 2018;430(15):2237-2243.
14. Jaroszewski L, Koska L, Sedova M, Godzik A. PubServer: literature searches by homology. *Nucleic Acids Res*. 2014;42(Web Server issue):W430-W435.
15. Šali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol*. 1993;234(3):779-815.
16. Söding J. Protein homology detection by HMM-HMM comparison. *Bioinformatics*. 2005;21(7):951-960.
17. Zhang J, Liang Y, Zhang Y. Atomic-level protein structure refinement using fragment-guided molecular dynamics conformation sampling. *Structure*. 2011;19(12):1784-1795.
18. Olechnovič K, Venclovas Č. VoroMQA: assessment of protein structure quality using interatomic contact areas. *Proteins*. 2017;85(6):1131-1145.
19. Li W, Godzik A. CD-HIT: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*. 2006;22(13):1658-1659.
20. Katoh K, Kuma K, Toh H, Miyata T. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucl Acids Res*. 2005;33(2):511-518.
21. Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD. HMMER web server: 2018 update. *Nucleic Acids Res*. 2018;46(W1):W200-W204.
22. UniProt Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res*. 2019;47(D1):D506-D515.
23. David L, Luo R, Gilson MK. Ligand-receptor docking with the mining minima optimizer. *J Comput Aided Mol Des*. 2001;15(2):157-171.
24. Kairys V, Gilson MK. Enhanced docking with the mining minima optimizer: acceleration and side-chain flexibility. *J Comput Chem*. 2002;23(16):1656-1670.
25. MacKerell AD, Bashford D, Bellott M, et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B*. 1998;102(18):3586-3616.
26. Momany FA, Rone R. Validation of the general purpose QUANTA®3.2/CHARMm® force field. *J Comput Chem*. 1992;13(7):888-900.
27. Gelin BR, Karplus M. Side-chain torsional potentials: effect of dipeptide, protein, and solvent environment. *Biochemistry*. 1979;18(7):1256-1268.
28. Méndez R, Leplae R, De Maria L, Wodak SJ. Assessment of blind predictions of protein-protein interactions: current status of docking methods. *Proteins*. 2003;52(1):51-67.
29. Olechnovič K, Kulberkytė E, Venclovas Č. CAD-score: a new contact area difference-based function for evaluation of protein structural models. *Proteins*. 2013;81(1):149-162.
30. Olechnovič K, Venclovas Č. Voronota: a fast and reliable tool for computing the vertices of the Voronoi diagram of atomic balls. *J Comput Chem*. 2014;35(8):672-681.
31. Berman HM, Westbrook J, Feng Z, et al. The Protein Data Bank. *Nucleic Acids Res*. 2000;28(1):235-242.
32. Javaheri A, Kruse T, Moonens K, et al. Helicobacter pylori adhesin HopQ engages in a virulence-enhancing interaction with human CEACAMs. *Nat Microbiol*. 2016;2:16189.
33. Fedarovich A, Tomberg J, Nicholas RA, Davies C. Structure of the N-terminal domain of human CEACAM1: binding target of the opacity proteins during invasion of Neisseria meningitidis and N. gonorrhoeae. *Acta Crystallogr D Biol Crystallogr*. 2006;62(Pt 9):971-979.
34. Moonens K, Gideonsson P, Subedi S, et al. Structural insights into polymorphic ABO glycan binding by helicobacter pylori. *Cell Host Microbe*. 2016;19(1):55-66.
35. Netzer R, Listov D, Lipsh R, et al. Ultrahigh specificity in a network of computationally designed protein-interaction pairs. *Nat Commun*. 2018;9(1):5286.
36. Lupardus PJ, Garcia KC. The structure of interleukin-23 reveals the molecular basis of p40 subunit sharing with interleukin-12. *J Mol Biol*. 2008;382(4):931-941.
37. Beyer BM, Ingram R, Ramanathan L, et al. Crystal structures of the pro-inflammatory cytokine interleukin-23 and its complex with a high-affinity neutralizing antibody. *J Mol Biol*. 2008;382(4):942-955.
38. Desmet J, Verstraete K, Bloch Y, et al. Structural basis of IL-23 antagonism by an Alphabody protein scaffold. *Nat Commun*. 2014;5:5237.
39. Chow D, He X, Snow AL, Rose-John S, Garcia KC. Structure of an extracellular gp130 cytokine receptor signaling complex. *Science*. 2001;291(5511):2150-2155.
40. Boulanger MJ, Chow D, Brevnova EE, Garcia KC. Hexameric structure and assembly of the interleukin-6/IL-6 alpha-receptor/gp130 complex. *Science*. 2003;300(5628):2101-2104.
41. Tamada T, Honjo E, Maeda Y, et al. Homodimeric cross-over structure of the human granulocyte colony-stimulating factor (GCSF) receptor signaling complex. *Proc Natl Acad Sci USA*. 2006;103(9):3135-3140.
42. Bloch Y, Bouchareychas L, Merceron R, et al. Structural activation of pro-inflammatory human cytokine IL-23 by cognate IL-23 receptor enables recruitment of the shared receptor IL-12Rβ1. *Immunity*. 2018;48(1):45-58.e6.
43. Skálová T, Bláha J, Harlos K, et al. Four crystal structures of human LLT1, a ligand of human NKR-P1, in varied glycosylation and oligomerization states. *Acta Crystallogr D Biol Crystallogr*. 2015;71(Pt 3):578-591.
44. Li Y, Wang Q, Chen S, Brown PH, Mariuzza RA. Structure of NKp65 bound to its keratinocyte ligand reveals basis for genetically linked recognition in natural killer gene complex. *Proc Natl Acad Sci USA*. 2013;110(28):11505-11510.
45. Kamishikiryo J, Fukuhara H, Okabe Y, Kuroki K, Maenaka K. Molecular basis for LLT1 protein recognition by human CD161 protein (NKRP1A/KLRB1). *J Biol Chem*. 2011;286(27):23823-23830.
46. Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol*. 2007;372(3):774-797.
47. Bliven S, Lafita A, Parker A, Capitani G, Duarte JM. Automated evaluation of quaternary structures from protein crystals. *PLoS Comput Biol*. 2018;14(4):e1006104.
48. Olechnovič K, Venclovas Č. VoroMQA web server for assessing three-dimensional structures of proteins and protein complexes. *Nucleic Acids Res*. 2019;47(W1):W437-W442.
49. Ridley H, Lakey JH. Antibacterial toxin colicin N and phage protein G3p compete with TolB for a binding site on TolA. *Microbiology*. 2015;161(Pt 3):503-515.
50. Cassels E. *Characterisation of the E. coli and Pseudomonas aeruginosa TolA-TolB Interaction*. York: University of York; 2012.
51. Holmes P. *Structure and Mode of Action of the TolA-TolB Complex from Pseudomonas aeruginosa*. Oxford: University of Oxford; 2016.
52. Liang J, Jaffrey SR, Guo W, Snyder SH, Clardy J. Structure of the PIN/-LC8 dimer with a bound peptide. *Nat Struct Biol*. 1999;6(8):735-740.
53. Gallego P, Velazquez-Campoy A, Regué L, Roig J, Reverter D. Structural analysis of the regulation of the DYNLL/LC8 binding to Nek9 by phosphorylation. *J Biol Chem*. 2013;288(17):12283-12294.
54. Bodor A, Radnai L, Hetényi C, et al. DYNLL2 dynein light chain binds to an extended linear motif of myosin 5a tail that has structural plasticity. *Biochemistry*. 2014;53(45):7107-7122.
55. Clark S, Nyarko A, Löhr F, Karplus PA, Barbar E. The anchored flexibility model in LC8 motif recognition: insights from the Chica complex. *Biochemistry*. 2016;55(1):199-209.

56. Alhassid A, Ben-David A, Tabachnikov O, et al. Crystal structure of an inverting GH 43 1,5-alpha-L-arabinanase from *Geobacillus stearothermophilus* complexed with its substrate. *Biochem J*. 2009;422(1):73-82.
57. Laederach A, Reilly PJ. Modeling protein recognition of carbohydrates. *Proteins*. 2005;60(4):591-597.
58. Fadda E, Woods RJ. Molecular simulations of carbohydrates and protein-carbohydrate interactions: motivation, issues and prospects. *Drug Discov Today*. 2010;15(15–16):596-609.
59. Guvench O, Mallajosyula SS, Raman EP, et al. CHARMM additive all-atom force field for carbohydrate derivatives and its utility in polysaccharide and carbohydrate-protein modeling. *J Chem Theory Comput*. 2011;7(10):3162-3180.
60. Myllykoski M, Eichel MA, Jung RB, Kelm S, Werner HB, Kursula P. High-affinity heterotetramer formation between the large myelin-associated glycoprotein and the dynein light chain DYNLL1. *J Neurochem*. 2018;147(6):764-783.
61. Andreani J, Faure G, Guerois R. InterEvScore: a novel coarse-grained interface scoring function using a multi-body statistical potential coupled to evolution. *Bioinformatics*. 2013;29(14):1742-1749.
62. Ovchinnikov S, Kamisetty H, Baker D. Robust and accurate prediction of residue-residue interactions across protein interfaces using evolutionary information. *Elife*. 2014;3:e02030.
63. Hopf TA, Schärfe CPI, Rodrigues JPGLM, et al. Sequence co-evolution gives 3D contacts and structures of protein complexes. *Elife*. 2014;3:03430.

How to cite this article: Dapkūnas J, Kairys V, Olechnovič K, Venclovas Č. Template-based modeling of diverse protein interactions in CAPRI rounds 38-45. *Proteins*. 2020;88:939–947. https://doi.org/10.1002/prot.25845