IN SILICO-BASED PREDICTION OF PROTEIN SOLUBILITY: THE CASE OF A MULTI-EPITOPE VACCINE CANDIDATE

V. S. Alves[1,2], I. G. Leme[2], A. C. M. Pássaro[2] & V. M. Gonçalves[2]*

[1] *Graduation Program in Biotechnology, University of São Paulo, São Paulo, Brazil.*
[2] *Laboratory of Vaccine Development, Instituto Butantan, São Paulo, Brazil.*
* *viviane.goncalves@butantan.gov.br.*

ABSTRACT

Immunoinformatics facilitate the screening of immunogenic regions to design versatile antigens based on epitopes with diverse immunogenic profiles that can be genetically fused to form multi-epitope vaccines (MEV), which are potential alternatives for serotype-dependent diseases, like pneumococcal ones. After designing a MEV candidate, physicochemical parameters, like protein solubility, remain critical for subsequent steps for vaccine development, which involves gene expression and protein purification. In this sense, we have evaluated *in-silico* solubility parameters determined for a pneumococcal MEV candidate and subsequently produced the protein to evaluate its effective solubility. We have observed that, although parameters indicated an appropriate solubility for the molecule, MEV severely aggregated due to exposed hydrophobic patches and the formation of salt bridges with phosphate ions. Therefore, solubility-only analyses are not enough to determine the overall solubility of proteins, requiring aggregation and deep formulation analyses to ensure protein solubility in the presence of chemicals employed in the downstream processing of vaccine production.

Keywords: Bioinformatics. Multi-epitope vaccine. Pneumococcal. Solubility. Aggregation.

1 INTRODUCTION

Immunoinformatics has been able to generate significant immunological information by combining immunology concepts and bioinformatic tools for epitope prediction, enabling the discovery of vaccine candidates by scanning protein sequences of a given pathogen. The application of such *in-silico* techniques has contributed to the emergence of new vaccine designs, such as multi-epitope vaccines (MEV)[1]. The MEV approach presents many advantages over classical vaccines, including the assembly of epitopes obtained from different antigens and the ability to activate both antibody-mediated and cell-mediated immunological responses. In this sense, such vaccines present great potential for fighting serotype-dependent infections, such as the ones caused by *Streptococcus pneumoniae*[1]. Pneumonia cases alone were responsible for 15% of deaths worldwide in children under 5 years of age in 2019, and serotype replacement caused by non-vaccine serotypes and high vaccine prices pose some drawbacks to the available pneumococcal vaccines. Consequently, the design of a serotype-independent pneumococcal MEV becomes highly interesting in such context[2].

Nevertheless, effective MEV design depends on the prediction of several properties and information about the predicted molecule. Hence, bioinformatics can also be applied to the prediction of allergenic, antigenic and physicochemical properties of the designed MEV. Among such properties, protein solubility and aggregation stand out as fundamental concepts, which are related to protein structure and have led to many implications in the expression of the gene and purification of therapeutic proteins[3]. Many factors affect solubility and aggregation proneness of a protein, ranging from intrinsic properties (primary sequence and hierarchical arrangements) to external factors (solvent composition, additives and physical conditions)[3,4]. However, widely used solubility predictors might not be enough to predict the solubility of a protein, especially considering the dependance of more complex factors, such as second and tertiary structures and their interactions with environmental factors.

In this sense, this paper presents an *in-silico* analysis performed for a pneumococcal MEV candidate previously designed in our laboratory with immunoinformatic tools, and compares the *in-silico* data with the wet lab data. Furthermore, discrepancies observed were investigated for reasonable explanations and reported solutions were tested to increase MEV solubility.

2 MATERIAL & METHODS

MEV design: Sequences from 8 pathogen relevant proteins were selected from the literature and retrieved from GenBank for screening of B-cell, MHC II and MHC I epitopes. Then, overlapping of all epitopes using IEDB clustarization and R-Studio scripts and analyses of antigenicity (Vaxijen) and allergenicity (AllergenFP and AllerTOP) were employed to select epitopes with potent immunological responses. To compose the MEV, linkers (KK and GPGPG) that stimulate CD4+ or CD8+ response were included according to the expected immunological response.

3D structure and prediction of parameters: MEV 3D structure was predicted with AlphaFold2[5], refined with GalaxyWeb (https://galaxy.seoklab.org/) and validated with PROCHECK (https://saves.mbi.ucla.edu/) and Prosa-Web (https://prosa.services.came.sbg.ac.at/prosa.php). Physical and chemical parameters, including solubility, were predicted with ProtParam tool at Expasy (https://web.expasy.org/protparam/), Protein-Sol[6] and TISIGNER[7].

***E. coli* cultures and cell lysis:** The DNA sequence of the 3D model was codon and mRNA translation optimized and cloned in pET-28a(+). The protein was then produced in *E. coli* BL21(DE3), first at 100 mL scale and then at 1 L scale. Experiments (inoculum and production) at 100 mL were performed in 300 mL Tunair® flasks (IBI Scientific, USA) using an autoinduction medium[8], at 300 rpm and 30°C for 7 h, and cells were disrupted with BugBuster® Protein Extraction Reagent (Merck, USA). Experiments at 1 L scale were performed in 2.5 L Tunair® flasks in the same manner but for 8 h, and cells were disrupted in a PandaPLUS 2000 high-pressure homogenizer (GEA Group, Germany) using either phosphate or HEPES buffer. A third lysis buffer was evaluated adding 50 mM L-arginine (L-Arg) and 50 mM L-glutamic acid (L-Glu) to HEPES. MEV was quantified and analyzed by SDS-PAGE[9] in non-reducing conditions and densitometry.

3 RESULTS & DISCUSSION

The designed MEV presents a molecular mass of 57.11 kDa and a theoretical pI of 9.45, and its predicted 3D structure is represented by Figure 1A. The molecule also exhibits stability, as indicated by a predicted instability index of 29.74 (index < 40 means stable) and validation analyses (data not shown). Solubility parameters obtained by different methods are displayed in Table 1. Such parameters include the grand average of hydropathicity index (GRAVY), the predicted scaled solubility (in terms of the amount of soluble protein in clarified lysates compared to the total amount of protein) and the probability of solubility of the MEV. The designed molecule is predicted as soluble in water (GRAVY index < 0) and showed a 93% probability that it would be soluble. Moreover, according to experimental solubility datasets previously reported[10], the predicted scaled solubility for the MEV is greater than the one predicted for the average *E. coli* proteins (0.45). Therefore, all three methods predicted a soluble behavior for the designed MEV protein.

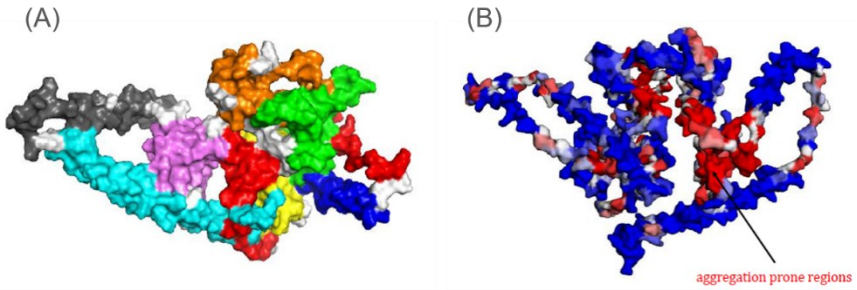

Figure 1 (A) Predicted 3D structure and (B) aggregation model for MEV. In (A), each color represents a different epitope, whereas in (B) blue patches represent soluble residues and red patches indicate aggregation-prone residues.

Table 1 Parameters obtained for MEV by bioinformatic tools.

GRAVY index (ProtParam tool)	Scaled solubility[5]	Solubility probability[7]
-0.93	0.67	0.93

The designed MEV was cloned in pET-28a(+) vector and expressed in *E. coli* for protein production in flask cultures. Cells were disrupted and both soluble and insoluble fractions were analyzed through SDS-PAGE (Figure 2A). The red arrow indicates the presence of MEV around 60 kDa, as expected, after 5.5 h of culture (Figure 2A). However, the protein was majorly found in the insoluble fraction as inclusion bodies, which contradicts the data in Table 1. A further analysis of the molecule using Aggrescan4D[11] showed many aggregation-prone binding sites in the structure of the MEV (Figure 1B). Given that these sites are not buried in the core of the structure, as expected for hydrophobic amino acid residues, active conformation of the MEV aligned with hydrophobicity of these sites facilitate the approximation of other molecules and their aggregation[3]. Average and maximum score values greater than zero in Aggrescan4D indicate higher aggregation tendency[11], corroborating with the behavior observed experimentally (Figure 2). Moreover, data predicted by this tool indicate that, for a pH range from 4.0 to 9.0, both score values change, ranging from -1.01 to -1.73 (average score) and from 6.89 to 7.60 (maximum score). Aggregation scores predicted at pH 7.5 were -1.53 (average score) and 6.89 for the MEV, indicating that working at neutral, slightly alkaline pH ranges (7.0 to 8.0) could contribute to avoid aggregation.

Moreover, the buffer of choice has also been reported to play a role in modifying protein-protein interactions, synergistically interacting with intrinsic protein properties, such as hydrophobicity, electrostatics and charge distribution[12,3]. Due to the high charge density of phosphate ions, these ions can interact with positively charged amino acid residues in proteins, consequently screening repulsion between molecules and allowing them to aggregate[12,13]. Indeed, when culture was performed at 1 L scale and lysis was carried out with phosphate buffer in a high-pressure homogenizer (instead of using BugBuster® Reagent), MEV aggregated so tightly clustered that not even 8 M urea could solubilize the pellet (data not show). Other proteins such as interferon tau[14], lysozyme[12], hemoglobin[15], monoclonal[16] and humanized[17] antibodies have also been reported to form aggregates in the presence of phosphate buffer.

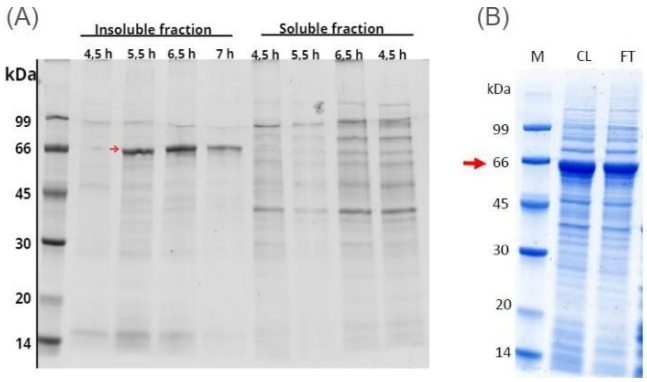

Figure 2 SDS-PAGE of samples during (A) 100 mL scale culture and of (B) 1 L scale culture samples after purification by CEXC. Lane 1 refers to the molecular marker (M), CL refers to clarified lysate and FT, to the insoluble fraction of chromatography flow-through.

After replacing phosphate buffer with HEPES buffer, the MEV remained soluble at first but precipitated later on or after purifying the clarified lysate through cation-exchange chromatography (CEXC) (Figure 2B). Therefore, buffer replacement was not enough to increase MEV solubility due to aggregation sites in the molecule, whose 3D conformation facilitates aggregation and, hence, precipitation. Therefore, we have tested the additives L-Arg and L-Glu amino acids in equimolar concentrations to decrease MEV aggregation, as it has been previously reported for other proteins[18,19]. The simultaneous addition of L-Arg and L-Glu has been reported to synergistically enhance the suppression of protein aggregation by crowding, as both can interact with oppositely charged groups on the surface of the protein, whereas aliphatic hydrophobic parts of side chains of these amino acids can cover adjacent exposed hydrophobic sections of the molecule[18,20]. When biomass lysis was carried out again in a high-pressure homogenizer, but this time with HEPES buffer and 50 mM L-Arg and 50 mM L-Glu, MEV barely precipitated and remained soluble even after longer times or CEX chromatography.

4 CONCLUSION

Solubility parameters alone predicted by bioinformatic tools are not enough to effectively predict the solubility of MEV in every environment. The exposure of hydrophobic patches by protein folding aligned with buffer composition were determinant for the solubility degree of the MEV, which naturally aggregated in the absence of competing agents (L-Arg and L-Glu) and presented a synergistic aggregation effect in the presence of phosphate buffer. Thus, when designing a new antigen, previous aggregation *in-silico* analyses, based on the 3D structure, are crucial for avoiding later issues in downstream processing.

REFERENCES

1 OLI, A. N., OBIALOR, W. O., IFEANYICHUKWU, M. O., ODIMEGWU, D. C., OKOYEH, J. N., EMECHEBE, G. O., ADEJUMO, S. A., IBEANU, G. C. 2020. Immunotargets Ther. 9. 13-30.
2 OLIVEIRA, G. S., OLIVEIRA, M. L. S., MIYAJI, E. N., RODRIGUES, T. C. 2021. Vaccines. 9 (11). 1338.
3 QING, R., HAO, S., SMORODINA, E., JIN, D., ZALEVSKY, A., ZHANG S. 2022. Chem. Rev. 122 (18). 14085-14179.
4 VIHINEN, M. 2020. ADMET & DMPK. 8 (4). 391-399.
5 JUMPER, J., EVANS, T., PRITZEL, A. *et al.* 2021. Nature. 296. 583-589.
6 HEBDITCH, M., CARBALLO-AMADOR, M. A., CHARONIS, S., CURTIS, R., WARWICKER, J. 2017. Bioinform. 33 (19). 3098-3100.
7 BHANDARI, B. K., LIM, S. L., GARDNER, P. P. 2021. Nucleic Acids Research. 49 (W1). W654-W661.
8 FUSCO, F., PIRES, M. C., LOPES, A. P. Y., ALVES, V. S., GONÇALVES, V. M. 2024. Front. Bioeng. Biotechnol. 11. 130496
9 LAEMMLI, U. K. 1970. Nature. 227 (5259). 680-685.
10 NIWA, T., YING, B., SAITO, K., JIN, W., TAKADA, S., UEDA, T., TAGUCHI, H. 2009. PNAS. 106 (11). 4201-4206.
11 BÁRCENAS, O., KURIATA, A., ZALEWSKI, M., IGLESIAS, V., PINTADO-GRIMA, C., FIRLIK, G., BURDUKIEWICZ, M., KMIECIK, S., VENTURA, S. 2024. Nucleic Acids Research. gkae382. 1-6
12 BRUDAR S., HRIBAR-LEE, B. 2021. J. Phys. Chem. B. 125 (10). 2504−2512
13 HARUTYUNYAN, E. H., KURANOVA, I. P., VAINSHTEIN, B, K., HÖHNE, W. E., LAMZIN, V, S., DAUTER, Z., TEPLYAKOV, A. V., WILSON, K. S. 1996. Eur. J. Biochem. 239 (1). 220-228.
14 KATAYAMA, D. S., NAYAR, R., CHOU, D. K., VALENTE, J. J., COOPER, J., HENRY, C. S., VELDE, D. G. V., VILLARETE, L. LIU, C. P., MANNING, M. C. 2006. J. Pharm. Sci. 95 (6). 1212-1226.
15 CHEN, K., BALLER, S. K., HANTGAN, R. R., KIM-SHAPIRO, D. B. 2004. Biophys. J. 87 (6). 4113-4121.
16 ROBERTS, D., KEELING, R., TRACKA, M., VAN DER WALLE, C. F., UDDIN, S., WARWICKER J., CURTIS R. 2015. Mol. Pharmaceutics. 12 (1). 179-193.
17 KAMEOKA, D., MASUKAZI, E., UEDA, T. IMOTO, T. 2007. J. Biochem. 142 (3). 383-391.
18 GOLOVANOV, A. P., HAUTBERGUEM G. M., WILSON, S. A., LIAN, L. 2004. J. Am. Chem. Soc. 126 (29). 8933-8939.
19 KHEDDO, P., TRACKA, M., ARMER, J., DEARMAN, R. J., UDDIN, S., VAN DER WALLE, C. F., GOLOVANOV, A. P. 2014. Int. J. of Pharm. 473 (1-2). 126-133.
20 SHUKLA, D., TROUT, B. L. 2011. J. Phys. Chem. B. 115 (41). 11831-11839.

ACKNOWLEDGEMENTS

The authors acknowledge the financial support from FAPESP (grants 2017/24832-6 and 2021/02930-1) and CNPq (grants 310973/2022-8 and 105844/2024-1).