scientific reports

Check for updates

OPEN

An in silico approach to study the role of epitope order in the multi-epitope-based peptide (MEBP) vaccine design

Muthu Raj Salaikumaran, Prasanna Sudharson Kasamuthu, Veeranarayanan Surya Aathmanathan & V. L. S. Prasad Burra✉

With different countries facing multiple waves, with some SARS-CoV-2 variants more deadly and virulent, the COVID-19 pandemic is becoming more dangerous by the day and the world is facing an even more dreadful extended pandemic with exponential positive cases and increasing death rates. There is an urgent need for more efficient and faster methods of vaccine development against SARS-CoV-2. Compared to experimental protocols, the opportunities to innovate are very high in immunoinformatics/in silico approaches, especially with the recent adoption of structural bioinformatics in peptide vaccine design. In recent times, multi-epitope-based peptide vaccine candidates (MEBPVCs) have shown extraordinarily high humoral and cellular responses to immunization. Most of the publications claim that respective reported MEBPVC(s) assembled using a set of in silico predicted epitopes, to be the computationally validated potent vaccine candidate(s) ready for experimental validation. However, in this article, for a given set of predicted epitopes, it is shown that the published MEBPVC is one among the many possible variants and there is high likelihood of finding more potent MEBPVCs than the published candidates. To test the same, a methodology is developed where novel MEBP variants are derived by changing the epitope order of the published MEBPVC. Further, to overcome the limitations of current qualitative methods of assessment of MEBPVC, to enable quantitative comparison and ranking for the discovery of more potent MEBPVCs, novel predictors, Percent Epitope Accessibility (PEA), Receptor specific MEBP vaccine potency (RMVP), MEBP vaccine potency (MVP) are introduced. The MEBP variants indeed showed varied MVP scores indicating varied immunogenicity. Further, the MEBP variants with IDs, SPVC_446 and SPVC_537, had the highest MVP scores indicating these variants to be more potent MEBPVCs than the published MEBPVC and hence should be preferred candidates for immediate experimental testing and validation. The method enables quicker selection and high throughput experimental validation of vaccine candidates. This study also opens the opportunity to develop new software tools for designing more potent MEBPVCs in less time.

The World Health Organization (WHO) announced SARS-CoV-2 as a pandemic in January 2020. Globally, as of date, there have been over 159 million confirmed cases of COVID-19 with the second COVID-19 wave kicking in around the 1st week of March 2021 and registering over 3.3 million deaths. According to a report, preventive and treatment options is one of the top two scenarios, the other being digitalization drive, which needs global adoption by the post-COVID-19 world to bounce back and get on to the revival path[1]. Since December 2019 the global COVID-19 pandemic has reached unprecedented deaths and death rates, currently continuing its more dreaded second wave. The world has called for preparedness for COVID-19 third wave as well. As of date, 475 COVID19 vaccine candidates are currently undergoing trials belonging to at least one of the fourteen different vaccine platforms[2–4]. Among them, 166 (34.9%) candidates are undergoing clinical trials, 251 (52.8%) candidates are undergoing the preclinical trials, 42 (8.8%) are in the Discovery phase and the remaining 16 (3.4%) have been discontinued. Currently, 34.9% (166) vaccine candidates belong to the protein subunit platform which emphasizes the importance and future potential of peptide-based vaccine candidates. There are 94 protein subunit-based vaccine candidates (~56.6%) that are under preclinical trials indicating there is a huge opportunity for peptide-based

Centre for Advanced Research and Innovation in Structural Biology of Diseases, K L E F University, Vaddeswaram, Andhra Pradesh 522 502, India. ✉email: dr.prasad.bvls@gmail.com

S.no	Molecular weight (kDa)	Sequence length	VCs	Epitope size (fixed(F)/varying (V))	Adjuvants	Linkers	Special motifs	TLRs	Targets	References
1	75	700	1	V	β-defensin 2	GPGPG, EAAAK, AAY, GGGS, KK	PADRE, TAT,	TLR3,4,8	S, M, N, E, Nsp8, Nsp3	24
2	20	183	1	F	β-defensin 2	EAAAK,AAY, GPGPG	–	TLR8	S	25
3	50	460	5	V	HSP70, TR-433, RS09	GPGPG, EAAAK, AAY	–	TLR4	S	27
4	17	146	1	V	HBD-2	EAAAK, AAY	–	TLR-2,3,4	S	28
5	30, 37, 50	254, 330, 475	3	V	–	KK, AA, GPGPG	–	TLR-2,3,4	S, M, N	26
6	50	485	1	F	HP-91, HBD-3	–	–	TLR3,5,8	S, N, M, E	29

Table 1. The common components, i.e., adjuvants, linkers and predicted epitopes specific to various targets of SARS-CoV-2 are used in a MEBPV construction. The data provides the components which are commonly used in the design of MEBPVCs with relevant references.

vaccine candidates[4]. As can be observed from the current vaccine approval statistics, protein subunit vaccine candidates have the highest approvals. Currently, out of 36 vaccines approved by at least one national regulatory authority(NRA) in the world, 12 belong to protein subunit platform as compared to the approval status eight months ago, where only one protein subunit vaccine: EpiVacCorona (Russia NRA), had the approval which is a peptide-antigen-based vaccine[5]. There are ten vaccines currently approved and listed under the WHO Emergency Use Listing (EUL) and among them two of the vaccines (COVOVAX™ and NUVAXOVID™) listed belong to the protein subunit platform where the same list did not have any Protein subunit platform based vaccines before the last revised list[6,7]. One of the first efforts of in silico vaccine design was published in 1999, formally launching immunoinformatics[8–10]. One of the first attempts to use multi-epitope-based polypeptides (MEBPs) for HIV-1 vaccine development dates back to 1999[11]. In recent times MEBPVCs have shown extraordinarily high humoral and cellular responses to immunization. One of the first computational methods for designing MEBP vaccine constructs was published in 2010[12,13]. Since then numerous articles related to in silico design of MEBP vaccine development have been published addressing vaccines against cancer, Chagas disease, filarial diseases, multi-drug resistance, Malaria, TB, and others[14–21]. MEBPVCs are gaining preference as they provide better control over the immunogenic components of the pathogen responsible for causing diseases, better reproducibility, and experimental control[22]. A recent article compiles all the interacting regions of the spike proteins from different pathogenic coronaviruses with human immune sensors like Toll-Like Receptors (TLRs), derives the epitopes specific to the interacting regions and designs MEBP. This methodology is claimed to provide a more comprehensive way to fight COVID-19 infection[23].

Recently, many novel multi-epitope-based peptide (MEBP) vaccine constructs (MEBPVCs) against Spike Protein and or multi-targets of SARS-CoV-2 that causes COVID-19 disease have been designed using in silico approaches. Rahmani et. al., 2021, proposed a trivalent (multi-target) MEBPVC that contains new components such as an intracellular delivery agent (TAT) and synthetic epitope (PADRE) in addition to conventional components such as adjuvants (β-defensin 2), the predicted epitopes, and linkers to boost the immune response[24]. The novelty in Saha et. al., 2021 publication is dual-purpose epitopes i.e., each epitope predicted is a B-Cell derived T-cell epitope with a fixed size of ten amino acids. In other words, each predicted epitope triggers a response from both B-Cells and T-Cells simultaneously. This approach keeps the size of MEBPVC small yet efficient, keeping the titers high from both B-cells and T-cells[25]. Similarly, Khairkhah et. al., 2020, in their design, have not used any adjuvants[26]. Table 1 gives a detailed summary of the most recent in silico approaches in MEBPVCs against spike protein or multi-targets of SARS-CoV-2.

It is a fundamental and proven fact in molecular biology that a change (modification/mutation) in the composition and or order of an individual or a group of amino/nucleic acids in the protein/DNA/RNA sequence could bring substantial change to the fold/3D structure and hence alter the function. In the context of peptide subunit vaccine platforms, this would mean that any change in the order or composition of amino acids of the peptide-based immunogen (especially large sized immunogens), may undergo changes in the fold and 3D structure. This in turn is expected to alter the biophysical and immunological properties of the MEBPVC such as surface accessibility of the epitopes in the MEBP. Any change in the epitope accessibility to the host immune system influences the immunogenicity/antigenicity of the MEBPVCs.

With increasing adoption, strong interdependence of structural biology and immunoinformatics, and with a trend for designing bigger and larger vaccines with sizes ranging from 10 to 100 kDa, the emphasis should be to utilize the protein structural biology knowledge for effective vaccine design. The opportunity for structure-based vaccine design leads to next-generation immunogen development[30,31]. In this article, we present a novel methodology to computationally design and validate MEBPVCs. This study indeed provided interesting insights that will help in developing novel MEBP specific design tools and also speedup and improve the current vaccine design methodologies and protocols which is not only the need of the hour but also a need for global preparedness for the future pandemics supporting novel platforms[32].

	SPVC_REF	SPVC_206	SPVC_214	SPVC_32	SPVC_357	SPVC_537	SPVC_383	SPVC_565	SPVC_446	SPVC_387
SPVC_REF	100.00	13.83	17.30	13.30	6.01	14.48	16.20	13.34	7.14	12.60
SPVC_206	46.24	100.00	16.32	10.27	14.01	8.75	12.88	8.82	13.15	14.79
SPVC_214	64.32	56.28	100.00	11.77	14.87	13.07	16.60	11.79	7.90	4.92
SPVC_32	54.10	49.47	57.30	100.00	7.19	13.24	15.21	12.75	3.38	15.08
SPVC_357	59.89	49.47	50.27	49.20	100.00	16.06	6.67	13.95	6.38	15.52
SPVC_537	55.68	43.09	51.34	70.27	50.81	100.00	11.60	14.23	3.72	7.18
SPVC_383	57.30	60.64	73.51	53.55	48.13	52.97	100.00	13.77	4.04	15.76
SPVC_565	50.54	49.47	53.19	71.89	51.35	65.95	52.97	100.00	13.63	13.66
SPVC_446	46.77	54.64	46.24	46.28	41.08	48.68	47.59	47.85	100.00	15.51
SPVC_387	58.12	54.64	39.13	47.06	56.15	42.47	50.27	50.27	55.14	100.00

Table 2. For ready comparison between the MEBP variants, the data is presented as percent sequence identity (lower triangle: below 100.00 diagonal) and root mean square deviation (RMSD) (upper triangle: above 100.00 diagonal). Lighter shades of a color indicate lower RMSD (higher structural similarity) and lower sequence identity whereas darker shades of the colors indicate higher RMSD (lower structural similarity) and higher sequence identity.

Results

The purpose of a vaccine (MEBPVC in the current context) is to elicit a strong response from the host immune system against a disease (COVID-19) releasing various neutralizing antibodies which continue to stay in the body to protect the host from any repeat infection thenceforth. Toll-like receptors (TLRs) are the common receptors that interact with the immunogen (MEBPVC) and trigger the downstream response and release of neutralizing antibodies. From the informatics point of view, to correlate the properties, to derive relationship with immunogenicity and changed epitope positions of the MEBPVC, the necessary data was generated at four levels: (a) sequence level, (b) 3D structure level, (c) receptor-ligand interaction level and (d) dosage versus immune response level. The data generated thus, are analysed to understand if the changed epitope positions influenced the various properties and eventually the immunogenicity. Prior to the analysis, to establish the fact that the change in the order/positions of the epitopes in a MEBPVC changes the immunogenicity, one has to first assess the diversity among the MEBP variants. To assess the same, pairwise alignment, multiple sequence alignment (MSA), and structure alignment were performed on the ten variants. The results are discussed below. Table 2 provides the percent sequence identity (Lower triangle) and root mean square deviation (RMSD) (upper triangle) as a ready reference.

Sequence & structural similarities between the MEBP variants. The average sequence identity between the variants is around 49% which is understandable considering the common adjuvants (45AA), linkers(AAY:3 × 5(copies) = 15AA, GPGPG:5 × 5(copies) = 25AA), and HIS tags(6AA) which make up to 49% (91AA) of the 183 AA long MEBPVC. Interestingly, the MEBP variants with IDs, SPVC_214 and SPVC_383, have 73.51% (highest) sequence identity (Fig. 1a).

When the epitope positions between SPVC_214 and SPVC_383 were studied, the same epitopes were present at the 2nd, 6th, 8th, 9th and 10th positions (Table 3).

Further, at the 7th position, almost identical epitopes, ***VLSFELL***HA, VV***VLSFELL*** were present in SPVC_214 and SPVC_383 respectively which clearly justifies the 73.51% sequence identity (Table 2). However, their RMSD is one among the highest, i.e. 16.60Å (Table 2) indicating very dissimilar structures which is contrary to the notion in homology modeling that is, high sequence identity (> 30%) indicates high structural similarity and same function.

The pair with the least sequence identity (39.13%) is SPVC_214 and SPVC_387 (Fig. 1b). Their RMSD is 4.92 Å indicating reasonable structural similarity though sequences are not very similar. The pair with the highest RMSD i.e.17.3 Å are REF_SEQ and SPVC_214. With such high RMSD, it is commonly expected to have very

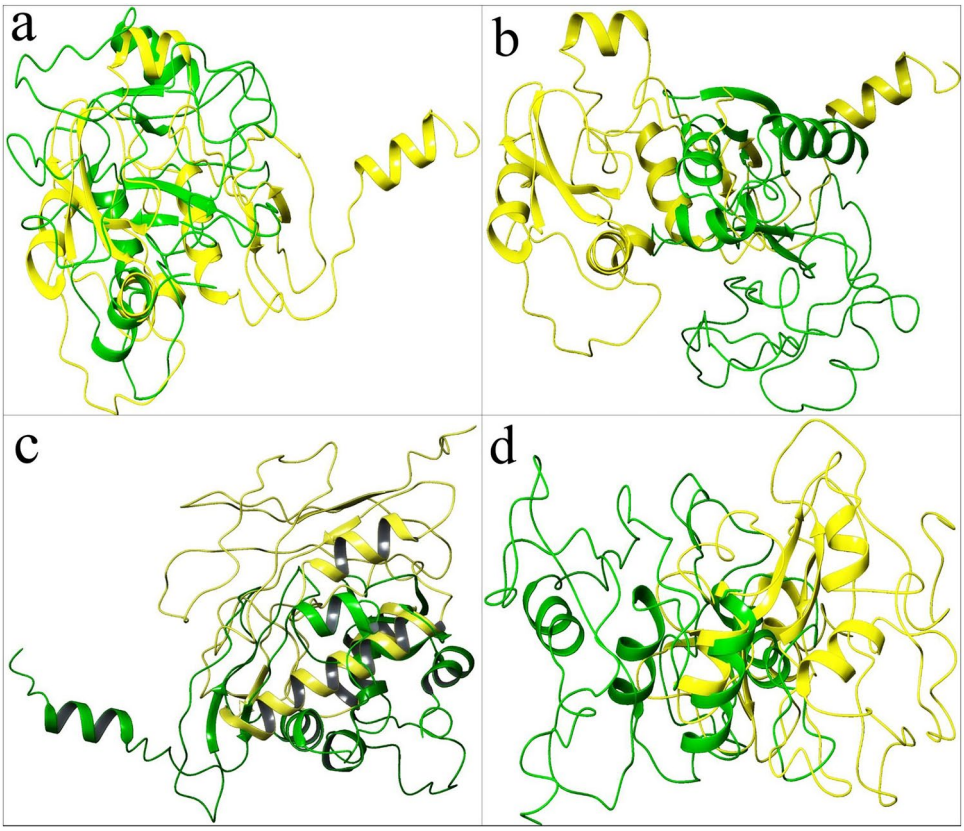

Figure 1. Structure alignment: (**a**) SPVC_214 (yellow) and SPVC_383 (green), 73.51% (SeqId), (**b**) SPVC_214 (yellow) and SPVC_387 (green), 39.13%, (**c**) REF_SEQ (yellow) and SPVC_214 (green), 17.3Å, (**d**) SPVC_32 (yellow) and SPVC_446 (green) 3.38 Å.

little sequence similarity between the two sequences. However, they have a sequence identity of around 65%. Further analysis revealed that the same epitopes were seen at 2nd, 3rd, and 4th positions in the two MEBP constructs (REF_SEQ and SPVC_214) (Table 2) justifying the above-average sequence identity. Similarly, the pair with the lowest RMSD i.e. 3.38 Å are SPVC_32 and SPVC_446 with their sequence identity of only 46.24% only. Figure 1c,d show structure alignments of the MEBP variant-pair with lowest (3.38Å, SPVC_32 and SPVC_446) and highest (17.3Å, REF_SEQ and SPVC_214) RMSDs. These typical cases are clearly indicating that there are changes in the 3D structures of the MEBP variants on changing the positions/order of the epitopes in the MEBPVCs.

Immunological and biophysical properties of the MEBP dataset. *Immunological properties.* Antigenicity, allergenicity. The antigenicity scores from Vaxijen 2.0 server indicate that all the MEBPVCs are probable antigens with a range between 0.62 to 0.78 (Table 4). The sequence, SPVC_446 has the highest antigenicity(0.78) and SPVC_214 (0.62) has the lowest antigenicity. All the allergenicity scores predicted using AllerTop v2.0 and AllergenFP v1.0 servers indicate that all the variants are non-allergens hence the allergen column was omitted from (Table 4).

Biophysical properties. Stability. The stability scores of MEBPVC variants fall between 73.17 to 76.97 which is the first parameter used as a filter. As a rule, all the MEBPVC variants must be predicted as stable which is found to be true for all the variants in the dataset. SPVC_214 is predicted to have the lowest and SPVC_537, to have the highest stability scores. The variations in the stability scores, though not very dispersed, indicate that change in the order of epitopes influenced the stability of the vaccine construct.

Solubility. The next biophysical property considered is solubility. Less soluble proteins are a major concern since the proteins synthesized may not fold to the right structure and hence lose the activity and function and are observed to precipitate out or form inclusion bodies leading to various disease states[33]. The solubility scores range from 0 to 1.0 where > 0.5 score indicates soluble and < 0.5 indicates insoluble peptide. In our case, all the MEBP variants had solubility scores ranging from 0.73 to 0.83 indicating all are soluble. The SPVC_446 variant has the lowest solubility and the SPVC_32 variant has the highest solubility. These solubility scores also indicate that the order of epitopes in the MEBPVC is important and crucial in the design of a good vaccine candidate.

MEBP Variant ID	% identity	RMSD	Epitope Position Number									
			1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th
REF_SEQ	100	0	TLDSKTQSL-AAY	**GKQGNFKNL-AAY**	**CYGVSPTKL-AAY**	**KIADYNYKL-AAY**	VVVLSFELL-GPGPG	IGINITRFQ-GPGPG	YGFQPTNGV-GPGPG	VLSFELLHA-GPGPG	LQIPFAMQM-GPGPG	IAIVMVTIM
SPVC_357	59.89	6.01	TLDSKTQSL-AAY	YGFQPTNGV-GPGPG	CYGVSPTKL-AAY	VVVLSFELL-GPGPG	IAIVMVTIM-GPGPG	IGINITRFQ-GPGPG	KIADYNYKL-AAY	GKQGNFKNL-AAY	VLSFELLHA-GPGPG	LQIPFAMQM
SPVC_446	46.77	7.14	YGFQPTNGV-GPGPG	VVVLSFELL-GPGPG	GKQGNFKNL-AAY	IAIVMVTIM-GPGPG	CYGVSPTKL-AAY	TLDSKTQSL-AAY	LQIPFAMQM-GPGPG	VLSFELLHA-GPGPG	KIADYNYKL-AAY	IGINITRFQ
SPVC_387	58.12	12.60	GKQGNFKNL-AAY	VLSFELLHA-GPGPG	IGINITRFQ-GPGPG	VVVLSFELL-GPGPG	CYGVSPTKL-AAY	TLDSKTQSL-AAY	KIADYNYKL-AAY	YGFQPTNGV-GPGPG	LQIPFAMQM-GPGPG	IAIVMVTIM
SPVC_32	54.10	13.3	CYGVSPTKL-AAY	KIADYNYKL-AAY	TLDSKTQSL-AAY	YGFQPTNGV-GPGPG	GKQGNFKNL-AAY	IGINITRFQ-GPGPG	VVVLSFELL-GPGPG	LQIPFAMQM-GPGPG	IAIVMVTIM-GPGPG	VLSFELLHA
SPVC_565	50.54	13.34	CYGVSPTKL-AAY	KIADYNYKL-AAY	YGFQPTNGV-GPGPG	VVVLSFELL-GPGPG	TLDSKTQSL-AAY	IGINITRFQ-GPGPG	IAIVMVTIM-GPGPG	LQIPFAMQM-GPGPG	GKQGNFKNL-AAY	VLSFELLHA
SPVC_206	46.24	13.83	YGFQPTNGV-GPGPG	GKQGNFKNL-AAY	VLSFELLHA-GPGPG	IGINITRFQ-GPGPG	KIADYNYKL-AAY	TLDSKTQSL-AAY	CYGVSPTKL-AAY	IAIVMVTIM-GPGPG	VVVLSFELL-GPGPG	LQIPFAMQM
SPVC_537	55.68	14.48	CYGVSPTKL-AAY	KIADYNYKL-AAY	GKQGNFKNL-AAY	YGFQPTNGV-GPGPG	VVVLSFELL-GPGPG	LQIPFAMQM-GPGPG	IAIVMVTIM-GPGPG	TLDSKTQSL-AAY	IGINITRFQ-GPGPG	VLSFELLHA
SPVC_383	57.30	16.20	KIADYNYKL-AAY	**GKQGNFKNL-AAY**	VVVLSFELL-GPGPG	TLDSKTQSL-AAY	CYGVSPTKL-AAY	**YGFQPTNGV-GPGPG**	**VLSFELLHA-GPGPG**	**IAIVMVTIM-GPGPG**	**IGINITRFQ-GPGPG**	**LQIPFAMQM**
SPVC_214	64.32	17.3	VLSFELLHA-GPGPG	**GKQGNFKNL-AAY**	**CYGVSPTKL-AAY**	**KIADYNYKL-AAY**	TLDSKTQSL-AAY	**YGFQPTNGV-GPGPG**	VV**VLSFELL**-GPGPG	**IAIVMVTIM-GPGPG**	**IGINITRFQ-GPGPG**	**LQIPFAMQM**

Table 3. Comparative epitope positions in the ten MEBP variants with % identity and RMSD with reference to REF_SEQ MEBPVC. Three comparisons are highlighted in this table (1) REF_SEQ vs SPVC_383, (2) REF_SEQ vs SPVC_214 and (3) SPVC_383 vs SPVC_214. Epitopes that are found at same position across the variants are highlighted (same color indicates same epitopes found at same position).

MEBP variant ID	Stability	Accessibility	solubility	Disorder	Aggregation	Hydrophobicity	Antigenicity	Sequence identity (%)	RMSD (str. similarity score)
REF_SEQ	74.52	36.36	0.82	0.17	3.30	0.20	0.62	0.00	0.00
SPVC_206	73.38	35.87	0.82	0.17	2.80	0.22	0.65	55.60	13.83 (0.31)
SPVC_214	73.17	36.56	0.82	0.15	2.80	0.19	0.62	64.70	17.3 (0.14)
SPVC_32	75.83	35.87	0.83	0.16	2.70	0.21	0.66	61.10	13.3 (0.34)
SPVC_357	73.84	35.79	0.82	0.17	2.80	0.22	0.63	64.10	6.01 (0.70)
SPVC_537	76.97	36.84	0.82	0.16	2.70	0.21	0.68	58.20	14.48 (0.28)
SPVC_383	73.84	37.23	0.82	0.16	2.80	0.22	0.65	60.00	16.20 (0.19)
SPVC_565	75.83	35.79	0.82	0.17	2.70	0.21	0.66	57.90	13.34 (0.33)
SPVC_446	76.05	36.56	0.73	0.16	3.10	0.24	0.78	50.70	7.14 (0.64)
SPVC_387	73.84	35.63	0.81	0.15	3.30	0.22	0.67	55.70	12.60 (0.37)

Table 4. The raw data of the Immuno- and Biophysical properties of the ten MEBP variants for ready reference.

Accessibility. Solvent accessibility is an important feature which, in the current context, has direct implications in eliciting the immune response in the host. The higher the epitope accessibility the more immunogenic the vaccine candidate. For the MEBP variants, the percent epitope accessibility (PEA) ranged between 35 to 37. The variant SPVC_387 has the lowest accessibility and SPVC_383 has the highest accessibility.

Disorder. In our MEBP variants disorder ranges between 0.15 to 0.16 and hence all variants are considered ordered. The MEBPVC sequence SPVC_357 has high disorder and SPVC_387 has low disorder among the variants. Low disorder is considered favorable for better vaccine design.

Aggregation. The predicted aggregation propensities ranged from 2.7 to 3.3 with lower values considered favorable. The sequences REF_SEQ and SPVC_387 have the highest aggregation propensity and SPVC_32, SPVC_565, SPVC_537 have the lowest propensity. Table 4 furnishes further details.

Hydrophobicity. Higher hydrophobicity shows better globularity, better accessible surface residues, and rigid 3D structure. The predicted hydrophobicity values ranged from 0.194 to 0.239, where higher hydrophobicity values are considered more favorable. The sequence SPVC_446 has the highest hydrophobicity and SPVC_214 has the lowest hydrophobicity.

Positive/negative influencers. Protein Stability Index, Surface Accessibility, Solubility, Sequence Identity & Sequence Alignments, Hydrophobicity, Docking (Z Rank) Score and Binding Affinity were categorized as positive influencers. Intrinsic Disorder, Protein Aggregation Propensity, RMSD and root mean square fluctuations (RMSF) were categorized as negative influencers.

Comparative docking analysis of MEBP variants. The properties compared in the previous sections were sequence-based. The analysis proved that the order of the epitopes indeed influenced the stability, solubility, accessibility, disorder, and aggregation properties. To make the analysis more complete and comprehensive, the following sections explore the docking and MD simulation studies using the ab initio modeled 3D structures of the MEBP variants. In the previous section, the MEBP models (variants) and their RMS deviations were discussed.

Among the family of Toll-Like Receptors (TLRs), the innate immune sensors, TLR8 and TLR4 are the most common receptors interacting with antigens/immunogens triggering an immune response from the host system to fight the immunogen. TLR8 plays an important role in the generation of effective immune responses in humans. TLR8 also senses the single-stranded RNA of viruses in the endosome and is predominantly expressed in the lungs. TLR4, plays an important role in the regulation of myocardial function, fibroblast activation, and acute inflammation by immune cells. Both the receptors are implicated in COVID-19. TLR4, is one of the 'fate-deciding' regulators of immunity and COVID-19 immunopathogenesis[34]. Table 5 shows the docking scores, binding affinities, minimization energies for both the receptors (TLR8 and TLR4) in complex with the MEBP variants.

ZRank Score is used to assess the quality of protein-protein docking. A more negative ZRank score indicates better docking. As can be seen from the tables, the ZRank scores are varying from (−100) to (−150) for the available MEBP variants. With TLR8, SPVC_537 has a better ZRank Score but an unfavorable binding affinity (11.92). With TLR4, SPVC_565 is showing favorable ZRank Score and acceptable binding affinity. It is interesting to note that there are at least four different MEBP variants having better ZRank scores when compared to REF_SEQ within a dataset size of ten variants. This definitely proves that changing the order of epitopes influences the 3D structure, which in turn influences the binding with immune machinery (TLRs) indicating the effectiveness of the immunogen (MEBPVC). Fig. 2 shows the docked MEBP variants with TLR4.

Simulation analysis of MEBPVC variants complexed with TLRs. The molecular dynamic simulation production run for 100 ns yielded a center projected trajectory in which the MEBP vaccine complexes were centered along the system in order to calculate the relative RMSD and RMSF for the MEBP vaccine complexes. The calculated RMSD for all the MEBP-TLR complexes were interpreted as maximum deviation data points and the average deviation data points to arrive at a holistic conclusion. Normalizing the data points and considering average deviation data points will provide us with much more stable MEBP-TLR complexes. Considering only the maximum deviation data points and neglecting the rest of the stable data points of the MEBP vaccine complex simulation over a 100 ns span will not be feasible for a simulation of this larger time span. The RMSF calculations were also performed similarly (RMSD calculation). The obtained total RMSD/RMSF calculations of ten MEBP vaccine candidates are represented in Table 5. The maximum and average RMSD of the REF_SEQ vaccine candidate is 0.53 nm and 0.45 nm for the TLR4 complex and 0.62 nm and 0.48 nm for the TLR8 complex.

The maximum and average RMSF of the REF_SEQ is 0.54 nm and 0.19 nm for the TLR4 complex and 0.71 nm and 0.18 nm for the TLR8 complex respectively. Considering REF_SEQ complex as the reference, all the other MEBP vaccine complexes were screened accordingly. Combinatorial approach of considering both the vaccine potency score and stability will help us arrive at the most potent of the MEBP vaccine candidates. The combined MVP is calculated with our scoring algorithm based on various physicochemical parameters. The MVP score is observed to be relatively higher for SPVC_446 and SPVC_537 with a score of 8.858 and 8.899 respectively, when compared to the REF_SEQ with an MVP score of 8.595. Both SPVC_446 and SPVC_537 prove to be promising vaccine candidates with high potency scores. The maximum/average RMSD of SPVC_446 is 0.4 nm/0.39 nm and SPVC_537 is 0.43 nm/0.35 nm for TLR4. The maximum/average RMSF of SPVC_446 is 0.36 nm/0.32 nm and SPVC_537 is 0.49nm/0.4 nm for TLR8. Figure 3 represents all the RMSD and RMSF calculations and respective

Receptor: TLR4						
MEBP variant ID	Ligand (MEBP variant) minimization (kcal/mol)	Complex minimization (kcal/mol)	Z rank score	Binding affinity (MMGBSA) ΔG(kcal/mol)	Max. RMSD of vaccine complex (nm) TLR4	Max. RMSF of vaccine complex (nm) TLR4
REF_SEQ	−8897.360	−49,810.762	−117.004	−76.85	0.53	0.54
SPVC_206	−9050.086	−49,328.161	−107.88	−87.25	0.41	0.53
SPVC_214	−8860.793	−49,637.407	−127.058	−54.61	5.13	2.75
SPVC_32	−8762.506	−49,244.125	−106.228	−97.5	0.45	0.58
SPVC_357	−8768.899	−48,041.522	−109.683	−78.16	0.59	0.75
SPVC_383	−9081.057	−49,148.629	−130.189	−74.5	0.42	0.47
SPVC_387	−9302.765	−49,276.327	−112.027	−81.84	0.48	0.75
SPVC_446	−8784.142	−48,974.441	−142.458	−88.09	0.48	0.45
SPVC_537	−8901.443	−48,695.629	−115.899	−91.22	0.43	0.72
SPVC_565	−9366.776	−49,240.549	−125.234	−77.85	4.45	2.67
Receptor: TLR8						
MEBP variant ID	Ligand (MEBP variant) minimization (kcal/mol)	Complex minimization (kcal/mol)	Z rank score:	Binding affinity (MMGBSA) ΔG (kcal/mol)	Max. RMSD of vaccine complex(nm) TLR8	Max. RMSF of vaccine complex (nm) TLR8
REF_SEQ	−8897.360	−59,185.545	−133.564	−117.14	0.62	0.71
SPVC_206	−9050.086	−59,388.441	−128.738	−53.78	0.39	0.67
SPVC_214	−8860.793	−59,004.133	−129.108	−72.06	0.55	0.91
SPVC_32	−8762.506	−58,969.654	−121.058	−38.62	0.78	0.93
SPVC_357	−8768.899	−59,208.700	−120.129	−36.01	0.84	1.28
SPVC_383	−9081.057	−59,568.962	−150.21	−62.54	3.98	2.34
SPVC_387	−9302.765	−59,425.876	−114.922	−80.93	4.21	1.25
SPVC_446	−8784.142	−59,213.308	−115.699	−56.48	0.36	0.67
SPVC_537	−8901.443	−59,614.015	−119.375	−56.37	0.49	0.685
SPVC_565	−9366.776	−59,603.462	−133.233	−66.48	0.45	1.03

Table 5. Minimization energies, Docking scores, binding affinities (MMGBSA), RMSD and RMSF of MEBP variants with both the receptors (TLR4 and TLR8).

plots. The RMSD plots show that the SPVCs are more stable when in complex with both TLR4 and TLR8. It can be observed that REF_SEQ is relatively less stable when compared to SPVC_446 and SPVC_537. RMSF plots of SPVC_446 and SPVC_537 indicate better interactions and higher stability, when in complex with TLR4 (Fig.3e). However, all the three SPVCs show similar high stability when in complex with TLR8 without a clear distinction as seen in TLR4 complexes (Fig. 3f).

Taking the stability of the complexes of MEBP vaccine candidates into consideration to derive a conclusion to select the best of the vaccine candidates, the SPVC_446 proves to be the best among the vaccine candidates. In spite of SPVC_537 having higher MVP score, the complex falls short in the stability parameter which is an important property to be looked into for biological activity.

Dosage versus immune response simulation analysis. As the last of the in silico tasks, we performed dosage vs immune response simulation for the ten vaccine constructs (MEBP variants) using the C-IMMSIM server with the same objective, to see if the variants trigger different responses than REF_SEQ, if so will the response indicate more potency or less. Two simulation experiments were done: (a) with adjuvant and HIS-tag and (b) without adjuvant and HIS-tag. The MEBPVC variants had all the parameters within the optimal and recommended ranges for them to be considered as a potent vaccine candidate individually with the exception of SPVC_387 (Supp. 2 Fig. S19) (without adjuvants + HIS-tag). A common observation has been that a repeated exposure led to an overall increase in the immune response and a decrease in the antigenic load (Supp. 2 Fig. S1–20).

Few observations are presented here. When compared to REF_SEQ, all other variants trigger strong antibody (especially IgM or IgM + IgG) responses with their 1st dose (exposure). Of all the variants, SPVC_214 is seen to trigger the highest titers of IgG + IgM. Of all the constructs, REF_SEQ triggers the weakest. The titers reach ~650,000 counts per ml for SPVC_214 and others but only ~580,000 counts per ml for REF_SEQ.

It is interesting to note that variants without adjuvants + HIS-tag seem to trigger more strongly than with adjuvants + HIS-tag. The antibody titers reach 90,000 counts per ml without adjuvants + HIS-tag as compared to only 20,000 counts per ml with adjuvants + HIS-tag on exposure to 1st dose of SPVC_214. The IgM + IgG titers reach ~760,000 counts per ml on the last (third) exposure. A similar trend is seen for all other variants as well, where without adjuvants + HIS-tag are triggering a better immune response. Figs. 4a–c and 5a–c show the level of immunoglobulins (with and without adjuvant + HIS-tag) at two different doses (1st, 2nd, and 3rd).

It is also observed that some variants such as SPVC_383 trigger high TH cell populations per state with counts reaching around 8200 per mm^3 with an average antibody (IgG + IgM) response of around 570,000 counts. When

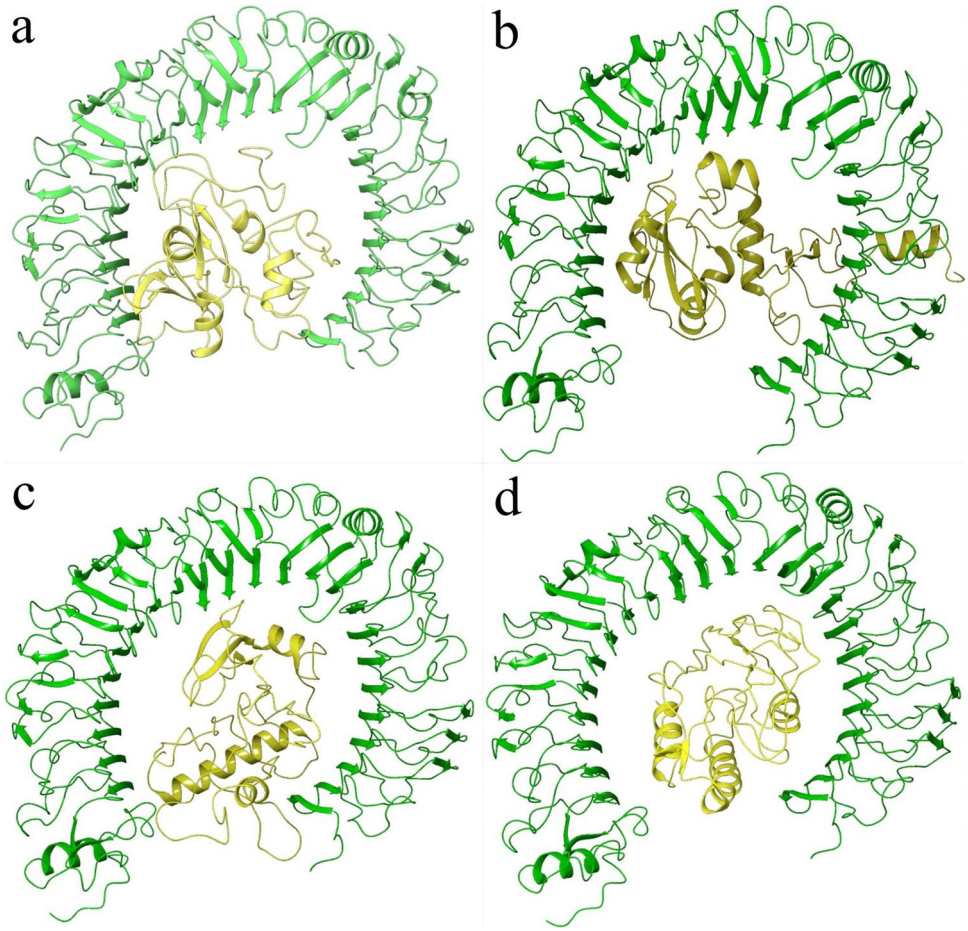

Figure 2. Docked ligand of (**a**) SPVC_32, (**b**) SPVC_214, (**c**) SPVC_565, (**d**) REF_SEQ (yellow) with TLR4 (green).

compared to others, the same variant also shows the best B cell population counts (800 per mm^3). According to the MEBP dose versus IFN-γ response simulation, it is interesting to note that variants with adjuvant + HIS-tag and without adjuvant + HIS-tag have totally different trends as seen in Fig. 6a,b. For example, SPVC_446 (with adjuvant+HIS-tag) triggers the highest concentration of IFN-γ and REF_SEQ has the lowest concentration of IFN-γ in the first dose. These above observations clearly demonstrate that change in the epitope order in a MEBP vaccine candidate influences immunogenicity.

Ranking the ten variants and identifying the most potent MEBPVC. Table 6 summarizes the receptor specific scores (RMVPs) and final MVP score for each variant. From the MVP score, SPVC_446 is predicted to be the most potent MEBPVC followed by SPVC_537. As can be seen, the least potent is REF_SEQ clearly proving that better and more potent MEBPVCs are possible by changing the epitope order and that epitope order influences immunogenicity (details about the normalized data are provided in Supp. 3).

Discussion

Vaccine development typically takes 10 years. In the pre-COVID-19 world, the fastest vaccine development time recorded was four years against mumps. It is no small feat to develop a vaccine against COVID-19 in a span of 9–10 months and vaccinate nearly 1.5 billion people. This shall be the new benchmark and reference for future vaccine development strategies and preparedness for future pandemics. This has become possible because of global cooperation for vaccine research and distribution[35].

The current COVID-19 vaccines listed under EUL have respective advantages and disadvantages[36]. The major disadvantage of Pfizer/BioNtech Comirnaty vaccine is its stringent cold chain requirement though it has shown very good titers[37]. The adenovector-based vaccines show relatively less effective neutralizing antibody response[38]. Inactivated vaccines seem to show inferior immunogenicity and low T Cell response, though have shown lower adverse reactions[39]. Similar to inactivated vaccines, the protein subunit vaccines show low immunogenicity. However, the possible advantages and potential benefits attracted the pharma companies to invest in protein subunit platforms. More than 30% of the total COVID-19 vaccine candidates undergoing trials are protein subunit vaccines with 65% under preclinical trials. Peptide-based vaccines have many unique benefits such as

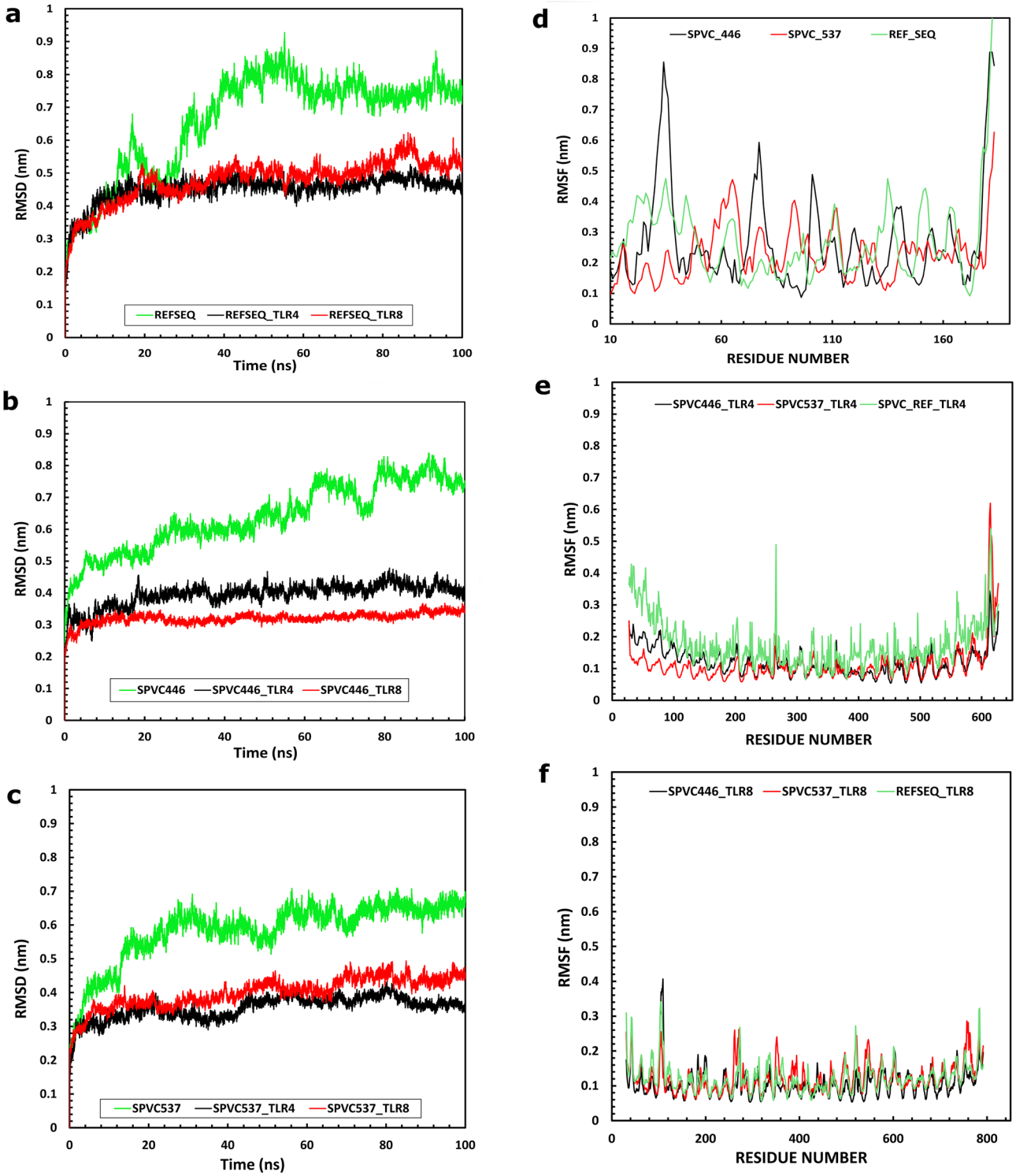

Figure 3. Molecular dynamic simulation studies of each SPVC interactions with TLR4 and TLR8. (**a**) RMSD plot of REF_SEQ, REF_SEQ-TLR4 and REF_SEQ-TLR8 complexes (**b**) RMSD of SPVC_446, (**c**) RMSD of SPVC_537, (**d**) RMSF plot of REF_SEQ, SPVC_446 and SPVC_537, (**e**) RMSF plot of TLR4 in complex with REF_SEQ, SPVC_446 and SPVC_537, (**f**) RMSF plot of TLR8 in complex with REF_SEQ, SPVC_446 and SPVC_537.

(1) fully defined composition, (2) affordable large scale production, (3) stable in storage and freeze-dryable, (4) absence of biological contamination, (5) minimum allergic and autoimmune responses, (6) customizable multipurpose therapeutic, and (7) standard rDNA technology-based production and manufacturing protocols

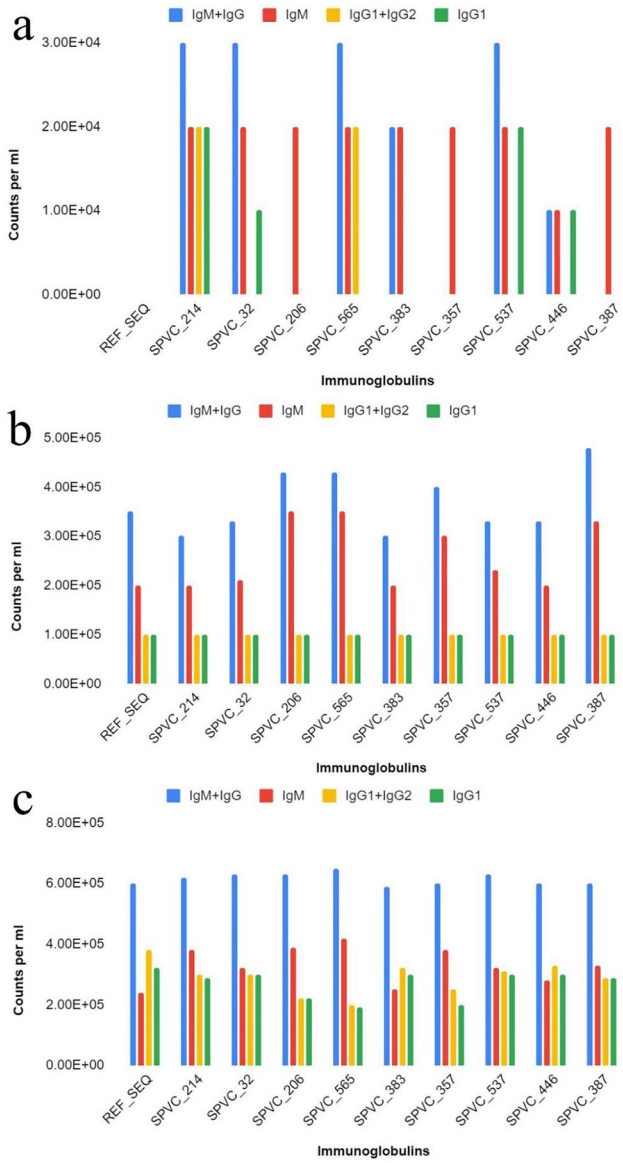

Figure 4. Immunoglobulin counts of MEBP variants (with adjuvants + HIS-tag) after (**a**) 1st dose, (**b**) 2nd dose, (**c**) 3rd dose (X axis—vaccine constructs, Y axis—counts per ml).

in place. MEBPVC has further advantages in designing multi-epitope, multi-target, multi-copy, multi-disease, and custom-size (molecular weight) vaccine constructs. The MEBP subunit vaccine platforms are in the initial phases of development.

Applications of in silico approach to design a MEBP vaccine is one of the pragmatic opportunities that can reduce the time in developing vaccines and reach the market in shorter time duration. The in silico methodology presented in this article shall further reduce the time to identify potential new vaccine candidates under the protein subunit vaccine platform.

It is known that peptide vaccines are weakly immunogenic. Considering the advantages offered by peptide-based vaccines that include MEBP vaccines, it is worth addressing the peptide vaccine-specific issues, where the major issue seems to be lower immunogenicity. This limitation is being effectively addressed through (a) combining with adjuvants such as β-defensin 2, HSP70, HBD-2, Matrix-M1, nanoparticles, (b) altering the size (molecular weight), and others. Adjuvants have shown to significantly boost immunogenicity but have not matched the current platforms such as RNA, adenovirus vector, and inactivated virus-based platforms[40–42].

It is the fundamental phenomenon that changes in the amino acid order change the structure and function, giving the clue that the earlier reported MEBPVC (REF_SEQ) could have variants if the epitope order changed. A set of ten variants were generated manually to explore if the variants thus generated have altered immunogenicity.

The variants were analyzed at the sequence, structure, interaction and dosage levels. In homology modeling, it is a common rule of thumb that for any two sequences, if the sequence identity is > 30%, it is assumed that their 3D structures shall be similar and likely to have identical function[43]. Further, it is also believed that with

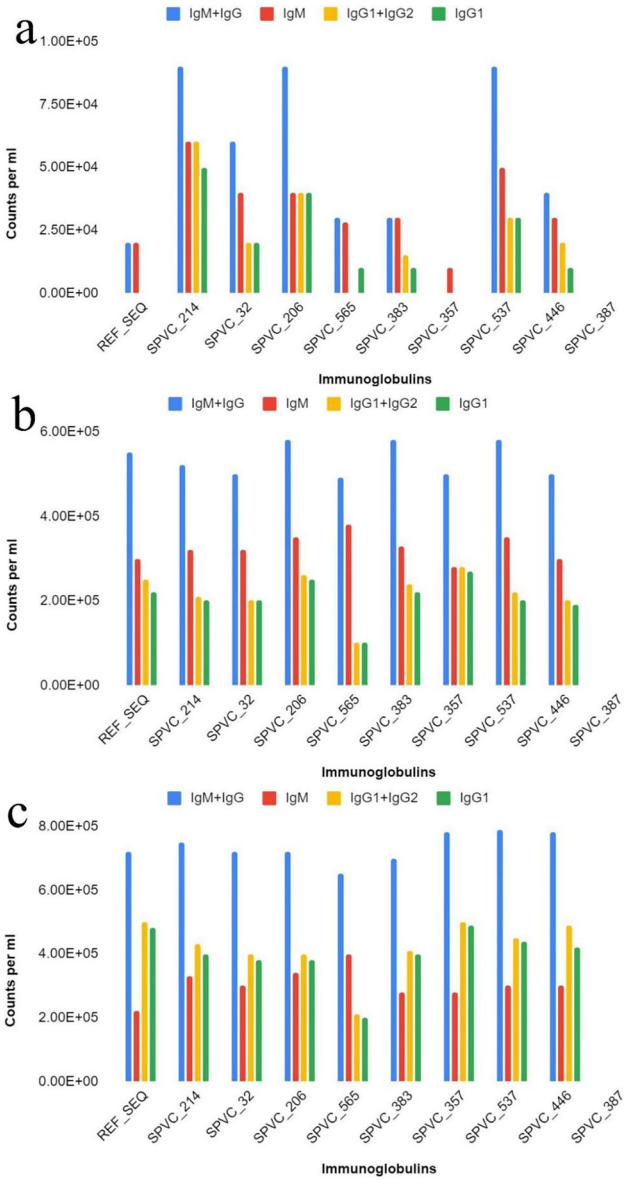

Figure 5. Immunoglobulin counts of MEBP variants (without adjuvants + HIS-tag) after (**a**) 1st dose, (**b**) 2nd dose, (**c**) 3rd dose (X axis—vaccine constructs, Y axis—counts per ml).

the increase in the sequence identity, the structural similarity also increases, i.e., RMSD decreases. However, it is interesting to note that the variants show deviations from the rule of thumb. As can be seen from Table 2 and Table 3, there are many pairs that show deviations. There have been studies that proved that 3D structures of 100% identical sequences were having natural conformations that have RMSDs as large as 24Å[44]. There have been studies where an all-α helix protein (Protein G) was engineered and transformed into an all-β protein(Protein Rop) by changing only 50% of the amino acid composition[45–47]. There is a need to experimentally verify the MEBP variant 3D structures through experimental structure determination techniques such as X-Ray Crystallography, NMR and or Electron microscopy.

The analysis, indeed, strongly suggests that changing the epitope order in MEBPVC changes the structure and hence the various associated properties resulting in the alteration of immunogenicity of the variant. Hence more potent MEBPVCs can be identified if the protocol described in this article is followed. The step to generate a dataset of shuffled variants is key in the analysis as this step enables the comparative study which otherwise has not been reported till now in MEBP vaccine design protocols. MVP score has also been developed which provides an opportunity to rank and identify the most potent MEBPVC from the dataset. Further, the data generated becomes the necessary input for developing better scoring schemes and algorithms.

The MVP scores were calculated through summing the individual RMVP scores of TLR4/8. The RMVP scores had two influencers: positive and negative. The positive influencers majorly indicate the physicochemical parameters, stability parameters and binding affinity of the protein complex. It also gives an overall insight about the functional ability of the MEBP-TLR complex. Likewise negative influencers were also taken into consideration:

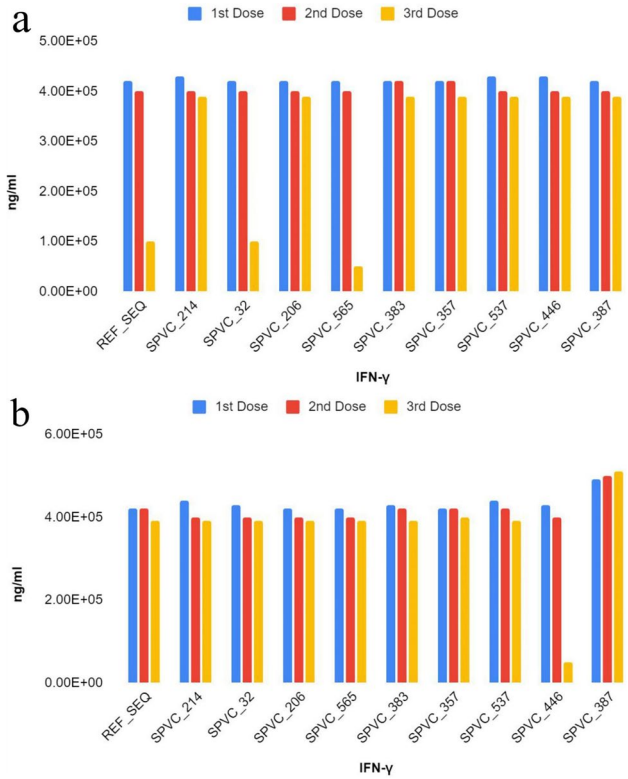

Figure 6. Concentration of IFN-γ for all MEBP variants '(**a**) with adjuvant + HIS-tag, (**b**) without adjuvant + HIS-tag) after the 1st, 2nd, and 3rd doses (X axis—vaccine constructs, Y axis—ng/ml).

MEBP vaccine potency (MVP)			
MEBP variant ID	**RMVP(TLR4)**	**RMVP(TLR8)**	**MVP (TLR4 + TLR8)**
REF_SEQ	3.933	4.662	8.595
SPVC_206	4.313	4.044	8.357
SPVC_214	0.366	3.995	4.361
SPVC_32	4.479	3.457	7.936
SPVC_357	3.914	2.916	6.830
SPVC_383	4.353	1.197	5.550
SPVC_387	4.119	2.444	6.563
SPVC_446	4.642	4.216	**8.858**
SPVC_537	4.529	4.371	**8.899**
SPVC_565	1.153	3.806	4.958

Table 6. MVP of each variant is calculated by adding receptor associated MVP (RMVP)s. RMVPs were calculated for the two receptors, TLR4 and TLR8 using the normalized values as described in the "Methods" section. Significant values are in bold.

disorderness, aggregation, RMSD and RMSF. The negative influencers directly incorporate the instability aspect of the MEBP-TLR complex to the MVP score. The RMSD and RMSF in depth analysis through molecular dynamic simulation provided data about the interaction stability and residue fluctuation (OPLS all atom forces fields and SPCE water model were used). SPVC_446 and SPVC_537 showed the highest MVP score and hence, are the most potential MEBPVCs from among the variant dataset. The normalized positive influencer scores of these two MEBPVCs were higher as it had positive influence over the stability and binding interactions. The normalized negative influencer scores of these two MEBPVCs were lower as it inversely influenced the disorderness and binding affinity of the MEBPVCs. The RMSD and RMSF tend to be higher for highly disordered or misfolded proteins[48]. Hence the RMSD and RMSF were considered as the negative influencers while calculating the MVP score. Lower the RMSD and RMSF of the MEBPVCs, higher the stability. The lower RMSD and RMSF indicates higher stability and lower misfolding candidates i.e. SPVC_446 and SPVC_537.

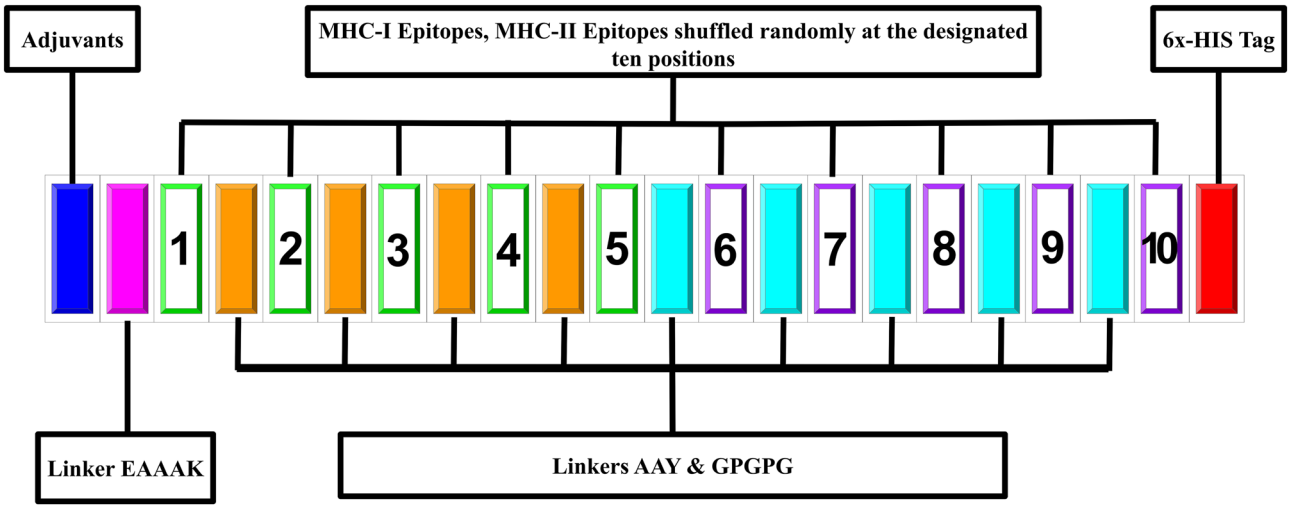

Figure 7. Schematic template of a typical MEBPVC to generate variants from REF_SEQ. The MEBPVC contains five major components: (1) N-terminal adjuvant (blue color), (2) predicted MHC I epitopes (green color border, position number 1–5, TLDSKTQSL, GKQGNFKNL, CYGVSPTKL, KIADYNYKL, VVVLSFELL), (3) predicted MHC II epitopes (magenta color border, position number 6–10, IGINITRFQ, YGFQPTNGV, VLSFELLHA, LQIPFAMQM, IAIVMVTIM), (4) linkers AAY (orange color), GPGPG (cyan color), (5) C-terminal HIS-tag (red color). The MEBP variants were generated by shuffling the epitopes at the ten positions.

Compared to REF_SEQ, all other variants trigger strong antibody responses. SPVC_537 is seen to trigger the higher titers of IgG + IgM. Of all the constructs, REF_SEQ triggers the weakest. IFN-γ is secreted by natural killer (NK) cells, which is activated through cell damage signals from infected or damaged cells through Cytotoxic T cells (CTLs). IFN-γ plays a major role in activation of macrophages, dendritic cells and T cells to curb the further downstream infection[49]. According to the MEBP dose versus IFN-γ response simulation, SPVC_446 triggers the highest concentration of IFN-γ and REF_SEQ has the lowest concentration of IFN-γ in the first dose. All the other higher responsive SPVCs were rejected based on the lower MVP score and stability. Similarly, the SPVCs that triggered higher antibody responses were also ignored, and more stable complexes were chosen to avoid possibility of protein aggregation and complications caused through misfolded proteins[50].

It is pertinent to mention that the number of variants was restricted to ten out of a possible 3,628,800 (10! possibilities) unique variants of similar length. The peptide length was restricted to 183 amino acids. The scope of the work was to test if the immunogenicity changes with the epitope order in the MEBPVC. For some parameters, it was also observed that the changes in the biophysical and immune parameters were not significant. This can be attributed to the small sample size, restriction on the length of the construct, and manual shuffling of epitopes for such small changes. The next challenges are to work with a bigger dataset of variants, optimize the parameters influencing the MEBP vaccine design, gain deeper insights into the mechanism(s) behind the mutations and their virulence and improvise the MVP score incorporating the knowledge and the decision making systems such as AI and ML[51].

Materials and methods

In general, a MEBP sequence is constructed with three broad components (peptide sequence patterns), namely, Linkers (for example AAY, GPGPG, EAAAK), Adjuvants for example β-defensin 2 and HSP70, and predicted target specific MHCI and MHC II Epitopes (oligopeptides with size ranging from 8AA to 20AA). The MEBP sequence (REF_SEQ) published earlier is taken as a reference for this study[25]. Table 4 lists the properties of REF_SEQ MEBPVC.

Construction of MEBP and its variants. The construction of the REF_SEQ is described elsewhere[25]. Using the REF_SEQ another nine MEBPVCs were generated by shuffling the epitope-linker set at the predesignated positions of REF_SEQ (Fig. 7).

The generated MEBP variants are given unique IDs following the format as: SPVC_NNN, where SPVC stands for shuffled peptide vaccine candidate and NNN stands for a unique number. The ten sequences are provided as supplementary data (Supp. 1 file).

The following restraints were applied while generating the MEBP variants: (a) The position and order of N-terminal adjuvant (β-defensin 2) followed by EAAAK (linker) and the C-terminal HIS-tag were kept unchanged. (b) The length of the construct was kept unchanged at 183AA only. (c) The B-cell-derived T-cell epitope (9AA long) plus the linker (AAY/GPGPG), together, were rearranged manually to create nine MEBPVC variants. (d) No distinction was made between MHC-I (green) and MHC-II (magenta) epitopes. The molecular weight, the sequence length, the isoelectric point (pI), the aliphatic index, the number of atoms, half-life, and the chemical formula remained identical for all the variants.

Sequence alignment. The formatting code, all-to-all sequence alignment among the members of the MEBP dataset was done using BioInt[52], Biobhasha (https://www.biobhasha.org), and BOSv2.0 (Biological Object-Based Software (BOS): An Integrative Biological Programming Environment). Multiple sequence alignment and sequence alignment renderer were performed using clustalw (https://www.genome.jp/tools-bin/clustalw) and ESpript 3.0 (https://espript.ibcp.fr/ESPript/cgi-bin/ESPript.cgi) [53,54].

Prediction of immunological and biophysical properties of MEBPVCs. The following relevant immuno- (*antigenicity and allergenicity*) and biophysical- (*protein stability index, surface/solvent accessibility, solubility, inherent intrinsic disorder, aggrescan, hydrophobicity*) properties were predicted using the online/offline web servers/tools to compare and develop a rationale for identifying the more potent ones.

Antigenicity. Antigenicity is the extent to which the host immune system responds to the antigen (foreign body) triggering both humoral and cellular responses. VaxiJen2.0 server (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html) was used to predict the antigenicity of the MEBP variants[55]. The result from VaxiJen 2.0 categorizes the peptide input into either Probable ANTIGEN or Probable NON-ANTIGEN. Only those variants which were categorized as Probable ANTIGEN were selected for further processing.

Allergenicity. Allergenicity is the extent to which an immunogen or antigen induces allergic reactions in the host system resulting in discomfort and or inconvenience or any allergic conditions such as asthma, diarrhea, skin rashes, and others. AllerTop v2.0 server (https://www.ddg-pharmfac.net/AllerTOP/) was used to predict the allergenicity of the MEBP variants[56]. The output from AllerTop v2.0 indicates if the input sequence is an allergen or not using the following restricted text values i.e. PROBABLE ALLERGEN or PROBABLE NON-ALLERGEN. Only those variants were selected for further processing which were categorized PROBABLE NON-ALLERGEN.

Peptide/protein stability index. Protein stability is an important property especially to understand the structure-function and activity relationships of a protein. The EXPASY ProtParam server (https://web.expasy.org/protparam/) was used for predicting the Instability index of the nine MEBPVCs[57]. The instability index was modified to the stability index by subtracting the score from 100. A score of more than 60 henceforth indicates the input protein to have better stability.

Solvent accessibility. Solvent accessibility is an important feature for understanding and interpreting the structure–activity relationship[58]. The solvent accessibility or the surface exposed residues provide data that helps in various predictions such as protein–protein interactions, receptor-ligand interactions, drug designing, protein folding, and others. Scratch Protein Predictor (http://scratch.proteomics.ics.uci.edu) was used to predict the solvent accessibility of the variants[59]. The output from Scratch Protein Predictor contained residue level accessibility with higher values indicating higher accessibility and vice versa. In the context of peptide vaccine design, higher accessibility and especially the residues with high accessibility in the epitope regions of MEBP is preferred since to elicit an immune response, the epitopes in the vaccine construct should be accessible, be exposed, and be on the surface. The accessibility predictor gives a string output, equal to the length of the sequence, with 'e' (for each exposed/accessible residue) and 'b'(for each buried residue). A more meaningful accessibility score in the context of MEBP vaccine design is percent epitope accessibility (PEA) defined and calculated as per the formula given below:

$$PEA = \frac{Total\ count\ of\ 'e's\ in\ the\ Epitope\ Regions\ of\ the\ MEBPVC\ sequence}{Total\ count\ of\ 'e's\ in\ the\ MEBPVC\ sequence} \times 100$$

The higher PEA value is considered favorable.

Solubility. Solubility of protein is an important biophysical property that depends on the amino acid composition and the 3D structure. Solubility influences the production of a protein and its half-life in the cell. Less soluble proteins are a major concern since the proteins synthesized may precipitate out or form inclusion bodies which lead to various disease states. SoluProt v1.0 server (https://loschmidt.chemi.muni.cz/soluprot/) was used to predict the solubility of all the MEBP variants[60]. For each input sequence, i.e. an MEBP variant in the current context, the SoluProt v1.0 server gives a score in the range of 0–1.0 where > 0.5 score indicates soluble and < 0.5 indicates insoluble peptide. The vaccine constructs with higher solubility (> 0.5) were selected for further processing.

Inherent intrinsic disorder. Intrinsically disordered proteins (IDPs) [a.k.a intrinsically unstructured proteins (IUPs)] are proteins that deviate from the dogma that every protein has a rigid 3D structure. IDPs or intrinsically disordered regions (IDRs) regulate many important biological functions such as transcription regulation, tissue-specific expression, and other signal transduction pathways. IDRs show high conformational changes, influence the stability of the protein, and affect the binding modes during ligand-receptor interactions. These regions play an important role in protein-protein interactions. IUPred2A server (https://iupred2a.elte.hu/) was used to predict the disorder of the MEBP variants[61]. The output has residue-wise disorder values and contiguous IDRs. A value >0.5 is considered disordered and the value <0.5 is considered ordered. An average disorder for

each MEBP variant was calculated using the residue-wise disorder. The variants with the low average disorder are considered for further analysis.

Protein aggregation. Protein aggregation is a biological process in which protein/peptide subunits instead of forming regular and functional assemblages, misfold, aggregate (intra- or extracellularly), and precipitate. Protein aggregation is one of the important phenomena implicated in diseases such as Parkinson's, Alzheimer's, and prion diseases[62]. To predict the aggregation for the variants, we used the AGGRESCAN server (http://bioinf.uab.es/aggrescan/)[63,64]. Variants with lower predicted aggregation were considered for further analysis.

Hydrophobicity. Hydrophobicity is the 'water-hating/avoiding/repelling' property of molecules. Hydrophobic amino acids tend to fold and shrink together to minimize contact with the solvent water or hydrophilic surroundings. The hydrophobic effect is a well-known important property to understand the 3D structure of a protein[65]. Hydrophobic interactions are an important driving force in protein folding hence the overall 3D structure. The shape determines the function of the protein. In the context of MEBPs which are the potential immunogens, higher hydrophobicity indicates more globularity, rigid 3D structure, and associated accessible surface residues. Kyte and Doolittle's method was used to generate the Hydrophobicity values by using EMBOSS pepwindow server (https://www.ebi.ac.uk/Tools/seqstats/emboss_pepwindow/)[66]. The outputs were screened and filtered for higher hydrophobicity and used for further processing.

Dosage versus immune response simulation. C-IMMSIM (http://150.146.2.1/C-IMMSIM/index.php) is one of the most commonly used in silico immune system simulation servers. Given the vaccine candidate (MEBP amino acid sequence), the dosage volume, and dosage intervals, C-IMMSIM predicts the humoral and cellular immune responses in the host system over a period of time. The server uses machine learning techniques to predict the immune responses of immunogens. The time step of injection, 3 doses, was set at 1 h, 84 h (3.5 days), and 168 h (7 days). The simulation steps and simulation volume were kept at 1100 and 10.

Structural studies. *Ab initio 3D structure prediction.* A local installation of the I-TASSER-suite was used to predict the tertiary structures of the MEBP variants. The suite uses the Local Meta-Threading server (LOMET) to find out the suitable template structure from the input sequence. In the second step, the server performs the template-based fragment assembly simulation to create a full structure and the final step is that it will produce the top five predicted models with TM-score (Quantitative assessment of similarity between protein structures) and C-score (confidence score for estimating the quality of predicted models).

Molecular docking and binding affinity calculations. The protein-protein docking was performed in the Discovery studio (Z_DOCK module)[67]. The receptors (TLR4 and TLR8) were downloaded from RCSB with PDBIDs:4G8A, 3W3M, and the MEBP variants were used as ligands. The docking resulted in ~2000 clusters with each cluster containing ~70 to 80 poses of ligands (MEBP variants) with the receptor (TLR4 and TLR8). The top poses with the highest Z_RANK-SCORE were selected to find R_DOCK-SCORES (Refined docked protein). The R_DOCK uses CHARMm energy to optimize docked poses produced by the Z_DOCK module. The top hits from the R_DOCK module were selected for further simulation studies. The binding affinities were calculated using the MMGBSA module from Discovery studio[68].

Molecular dynamics simulation. The MEBP variants were scored and assorted based on the favorability of each property it represented individually. Among hundreds of variants, we have selected the top 10 MEBP for finer investigation to produce a single potent MEBP vaccine candidate based on RMVP score (MEBP vaccine potency). The molecular dynamics simulation helps us study the stability and interaction of the MEBP-TLR complex in an ion based solvent system. All the top-scored MEBP vaccine candidates were subjected to molecular dynamics simulation using GROMACS v.2021 and topologies were generated using the OPLS force field. The system was solvated in a cubic box conformation using SPCE water model, energy minimized until the steepest descent energy, i.e. atoms are realigned to reduce the maximum net forces on them. The minimized atoms exert least forces on each atom and therefore serves as a favourable start point for molecular dynamics simulations. Pressurized and increasing temperature conditions were implemented for 100 ps. The molecular dynamics simulation was produced for a span of 100 ns over the centered projected trajectory. RMSD and RMSF were calculated using the gmx_rms option with reference to the MEBP Vaccine complex to TLR4 and TLR8 for all the ten MEBP TLR4/8 complexes.

Implementation of a scoring scheme for ranking and discovering the best MEBPVC from the library of variants.. The values obtained for the ten MEBPVC variants for the above discussed properties were grouped into (a) positive influencers (Stability, Accessibility, Solubility, Hydrophobicity, Antigenicity, ZrankScore, Binding Affinity (MMGBSA) b) negative influencers (Disorder, Aggregation, RMSD and RMSF). All the values of a property were normalized using normalization by averaging i.e.

$$x_{n,i} = \frac{x_i}{\underline{x}}$$

where x is the original value of the property, n is the normalized value, i is the index of the variant, $\underline{x}$ is the average, all the variants.

$$\text{Receptor specific MEBP Vaccine Potency (RMVP)}_{\text{i,m}} = \sum_{k=1}^{8} x_{n,k} - \sum_{l=1}^{4} x_{n,l}$$

where k, l are the normalized values of each positive and negative influencers (property of the variant) respectively. Where m is the receptor i.e. TLR4(1) or TLR8(2).

Final MVP for each variant, i, was calculated by summing individual RMVPs.

Conclusions

In pursuit to design and or discover more potent MEBPVCs, given the list of predicted epitopes, a methodology to generate variants and three predictors, percent epitope accessibility (PEA), receptor-specific vaccine potency (RMVP) and MEBP vaccine potency (MVP) scores have been introduced to quantify and enable an opportunity for the development of efficient MEBP vaccine design platform. This enabled not only ranking but also identifying the best MEBPVC. Thus, the results prove that the reported MEBPVC (REF_SEQ) is not the most potent candidate after all. In this article, only the *order* of epitopes has been used for generating variants. However, other parameters such as length of epitopes, length of the sequence, copy numbers, and or other parameters should also be explored in the future to design and discover more potent MEBPVCs. In conclusion, our in silico analysis and results indeed prove that changes in the position or order of the epitopes change the properties of the MEBPVC. Henceforth, the dataset generation method, PEA, RMVP, and MVP score should enable generating novel MEBPVCs and may be adopted in all MEBPVC design pursuits. The method enables the design and discovery of the computationally validated most immunogenic MEBPVC, each having a unique epitope order. Experimental validation and verification has no substitute; hence, the computationally validated vaccine constructs, with IDs, SPVC_446 and SPVC_537, need to be validated through in vitro, and in vivo experimental studies. The experimental validation provides important insights and inputs for improvising and developing a more efficient and more reliable scoring function and the improved versions of the software.

Data availability

All data generated or analysed during the study are included in the submitted manuscript.

Received: 23 October 2021; Accepted: 11 July 2022
Published online: 22 July 2022

References

1. Planning for the Post-COVID-19 Workforce: Four Scenarios. https://knowledge.wharton.upenn.edu/article/planning-post-covid-19-workforce-four-scenarios/.
2. Nagy, A. & Alhatlani, B. An overview of current COVID-19 vaccine platforms. *Comput. Struct. Biotechnol. J.* **19**, 2508–2517 (2021).
3. COVID-19 vaccine tracker and landscape. https://www.who.int/publications/m/item/draft-landscape-of-covid-19-candidate-vaccines.
4. COVID-19 Vaccine Market Dashboard. https://www.unicef.org/supply/covid-19-vaccine-market-dashboard.
5. Status of COVID-19 vaccines within WHO EUL/PQ evaluation process updated. *WHO—Prequalification of Medical Products (IVDs, Medicines, Vaccines and Immunization Devices, Vector Control)*. https://extranet.who.int/pqweb/news/status-covid-19-vaccines-within-who-eulpq-evaluation-process-updated (2021).
6. COVID-19 vaccines WHO EUL issued. *WHO—Prequalification of Medical Products (IVDs, Medicines, Vaccines and Immunization Devices, Vector Control)*. https://extranet.who.int/pqweb/vaccines/vaccinescovid-19-vaccine-eul-issued (2021).
7. The Novavax vaccine against COVID-19: What you need to know. https://www.who.int/news-room/feature-stories/detail/the-novavax-vaccine-against-covid-19-what-you-need-to-know.
8. DeWitt, N. In silico vaccine design?. *Nat. Biotechnol.* **17**, 523–523 (1999).
9. Groot, A. S. D., De Groot, A. S. & Rothman, F. G. In silico predictions; in vivo veritas. *Nat. Biotechnol.* **17**, 533–534 (1999).
10. Sturniolo, T. *et al.* Generation of tissue-specific and promiscuous HLA ligand databases using DNA microarrays and virtual HLA class II matrices. *Nat. Biotechnol.* **17**, 555–561 (1999).
11. Cano, C. A. The multi-epitope polypeptide approach in HIV-1 vaccine development. *Genet. Anal.* **15**, 149–153 (1999).
12. Liao, Y.-C., Lin, H.-H., Lin, C.-H. & Chung, W.-B. Identification of cytotoxic T lymphocyte epitopes on swine viruses: Multi-epitope design for universal T cell vaccine. *PLoS ONE* **8**, e84443 (2013).
13. Chakraborty, S. *et al.* A computational approach for identification of epitopes in dengue virus envelope protein: A step towards designing a universal dengue vaccine targeting endemic regions. *In Silico Biol.* **10**, 235–246 (2010).
14. Pandey, R. K., Bhatt, T. K. & Prajapati, V. K. Novel immunoinformatics approaches to design multi-epitope subunit vaccine for malaria by investigating anopheles salivary protein. *Sci. Rep.* **8**, 1125 (2018).
15. Abbas, G., Zafar, I., Ahmad, S. & Azam, S. S. Immunoinformatics design of a novel multi-epitope peptide vaccine to combat multi-drug resistant infections caused by *Vibrio vulnificus*. *Eur. J. Pharm. Sci.* **142**, 105160 (2020).
16. Michel-Todó, L. *et al.* Design of an epitope-based vaccine ensemble for Chagas disease. *Front. Immunol.* **10**, 2698 (2019).
17. Nezafat, N., Ghasemi, Y., Javadi, G., Khoshnoud, M. J. & Omidinia, E. A novel multi-epitope peptide vaccine against cancer: An in silico approach. *J. Theor. Biol.* **349**, 121–134 (2014).
18. Shey, R. A. *et al.* In-silico design of a multi-epitope vaccine candidate against onchocerciasis and related filarial diseases. *Sci. Rep.* **9**, 4409 (2019).
19. Buteau, C., Markovic, S. N. & Celis, E. Challenges in the development of effective peptide vaccines for cancer. *Mayo Clin. Proc.* **77**, 339–349 (2002).
20. Purcell, A. W., McCluskey, J. & Rossjohn, J. More than one reason to rethink the use of peptides in vaccine design. *Nat. Rev. Drug Discov.* **6**, 404–414 (2007).
21. Bibi, S. *et al.* In silico analysis of epitope-based vaccine candidate against tuberculosis using reverse vaccinology. *Sci. Rep.* **11**, 1249 (2021).
22. Pourseif, M. M., Yousefpour, M., Aminianfar, M., Moghaddam, G. & Nematollahi, A. A multi-method and structure-based in silico vaccine designing against through investigating enolase protein. *Bioimpacts* **9**, 131–144 (2019).
23. Choudhury, A., Sen Gupta, P. S., Panda, S. K., Rana, M. K. & Mukherjee, S. Designing AbhiSCoVac - A single potential vaccine for all 'corona culprits': Immunoinformatics and immune simulation approaches. *J. Mol. Liq.* **351**, 118633 (2022).

24. Rahmani, A., Baee, M., Saleki, K., Moradi, S. & Nouri, H. R. Applying high throughput and comprehensive immunoinformatics approaches to design a trivalent subunit vaccine for induction of immune response against emerging human coronaviruses SARS-CoV, MERS-CoV and SARS-CoV-2. *J. Biomol. Struct. Dyn.* **21**, 1–17 (2021).
25. Saha, R., Ghosh, P. & Prasad, V. L. Designing a next generation multi-epitope based peptide vaccine candidate against SARS-CoV-2 using computational approaches. *3 Biotech* **11**, 1–14 (2021).
26. Khairkhah, N., Aghasadeghi, M. R., Namvar, A. & Bolhassani, A. Design of novel multiepitope constructs-based peptide vaccine against the structural S, N and M proteins of human COVID-19 using immunoinformatics analysis. *PLoS ONE* **15**, e0240577 (2020).
27. Mitra, D., Pandey, J., Jain, A. & Swaroop, S. Design of multi-epitope-based peptide vaccine against SARS-CoV-2 using its spike protein. *J. Biomol. Struct. Dyn.* **29**, 1–14 (2021).
28. Jyotisha, S. S. & Qureshi, I. A. Multi-epitope vaccine against SARS-CoV-2 applying immunoinformatics and molecular dynamics simulation approaches. *J. Biomol. Struct. Dyn.* **40**, 1–17 (2020).
29. Yazdani, Z., Rafiei, A., Yazdani, M. & Valadan, R. Design an efficient multi-epitope peptide vaccine candidate against SARS-CoV-2: An in silico analysis. *Infect. Drug Resist.* **13**, 3007–3022 (2020).
30. Malonis, R. J., Lai, J. R. & Vergnolle, O. Peptide-based vaccines: Current progress and future challenges. *Chem. Rev.* **120**, 3210–3229 (2020).
31. Bellini, C. & Horváti, K. Recent advances in the development of protein- and peptide-based subunit vaccines against tuberculosis. *Cells* **9**, 2673 (2020).
32. To, K.K.-W. *et al.* Lessons learned 1 year after SARS-CoV-2 emergence leading to COVID-19 pandemic. *Emerg. Microbes Infect.* **10**, 507–535 (2021).
33. Matsubara, T. *et al.* Neuropathy/intranuclear inclusion bodies in oculopharyngodistal myopathy: A case report. *eNeurologicalSci* **24**, 100348 (2021).
34. Mukherjee, S. Toll-like receptor 4 in COVID-19: Friend or foe?. *Future Virol.* https://doi.org/10.2217/fvl-2021-0249 (2022).
35. Piccirillo, J. & Ledger, J. *A COVID-19 Vaccine Is Coming—Will It Be Safe*? https://www.yalemedicine.org/news/covid-19-vaccine-safety (2020).
36. He, Q. *et al.* Heterologous prime-boost: Breaking the protective immune response bottleneck of COVID-19 vaccine candidates. *Emerg. Microbes Infect.* **10**, 629–637 (2021).
37. Polack, F. P. *et al.* Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine. *N. Engl. J. Med.* **383**, 2603–2615 (2020).
38. Zhu, F.-C. *et al.* Immunogenicity and safety of a recombinant adenovirus type-5-vectored COVID-19 vaccine in healthy adults aged 18 years or older: A randomised, double-blind, placebo-controlled, phase 2 trial. *Lancet* **396**, 479–488 (2020).
39. Zhang, Y. *et al.* Safety, tolerability, and immunogenicity of an inactivated SARS-CoV-2 vaccine in healthy adults aged 18–59 years: A randomised, double-blind, placebo-controlled, phase 1/2 clinical trial. *Lancet Infect. Dis.* **21**, 181–192 (2021).
40. Skwarczynski, M. & Toth, I. Peptide-based synthetic vaccines. *Chem. Sci.* **7**, 842–854 (2016).
41. Li, W., Joshi, M. D., Singhania, S., Ramsey, K. H. & Murthy, A. K. Peptide vaccine: Progress and challenges. *Vaccines (Basel)* **2**, 515–536 (2014).
42. Keech, C. *et al.* Phase 1–2 trial of a SARS-CoV-2 recombinant spike protein nanoparticle vaccine. *N. Engl. J. Med.* **383**, 2320–2332 (2020).
43. Pearson, W. R. An introduction to sequence similarity ('homology') searching. *Curr. Protoc. Bioinform.* **3**, 3.1 (2013).
44. Burra, P. V., Zhang, Y., Godzik, A. & Stec, B. Global distribution of conformational states derived from redundant models in the PDB points to non-uniqueness of the protein structure. *Proc. Natl. Acad. Sci. USA* **106**, 10505–10510 (2009).
45. Rose, G. D. Protein folding and the Paracelsus challenge. *Nat. Struct. Biol.* **4**, 512–514 (1997).
46. Rose, G. D. & Creamer, T. P. Protein folding: Predicting predicting. *Proteins* **19**, 1–3 (1994).
47. Dalal, S., Balasubramanian, S. & Regan, L. Protein alchemy: Changing beta-sheet into alpha-helix. *Nat. Struct. Biol.* **4**, 548–552 (1997).
48. Yun, S. & Guy, H. R. Stability tests on known and misfolded structures with discrete and all atom molecular dynamics simulations. *J. Mol. Graph. Model.* **29**, 663–675 (2011).
49. Castiglione, F. & Celada, F. *Immune System Modelling and Simulation*. https://doi.org/10.1201/b18274 (2015).
50. Bagashev, A. *et al.* CD19 alterations emerging after CD19-directed immunotherapy cause retention of the misfolded protein in the endoplasmic reticulum. *Mol. Cell. Biol.* **38**, 00383 (2018).
51. Cantón, R. *et al.* New variants of SARS-CoV-2. *Rev. Esp. Quimioter.* https://doi.org/10.37201/req/071.2021 (2021).
52. Desai, S. & Burra, P. BioInt: An integrative biological object-oriented application framework and interpreter. *Int. J. Bioinform. Res. Appl.* **11**, 247 (2015).
53. Jeanmougin, F., Thompson, J. D., Gouy, M., Higgins, D. G. & Gibson, T. J. Multiple sequence alignment with Clustal X. *Trends Biochem. Sci.* **23**, 403–405 (1998).
54. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320–W324 (2014).
55. Doytchinova, I. A. & Flower, D. R. VaxiJen: A server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinform.* **8**, 4 (2007).
56. Dimitrov, I., Flower, D. R. & Doytchinova, I. AllerTOP—A server for in silico prediction of allergens. *BMC Bioinform.* **14**(6), S4 (2013).
57. Gross, R. H. ExPASy. *Biotech. Softw. Internet Rep.* **2**, 161–165 (2001).
58. Magnan, C. N. & Baldi, P. SSpro/ACCpro 5: Almost perfect prediction of protein secondary structure and relative solvent accessibility using profiles, machine learning and structural similarity. *Bioinformatics* **30**, 2592–2597 (2014).
59. Cheng, J., Randall, A. Z., Sweredoski, M. J. & Baldi, P. SCRATCH: A protein structure and structural feature prediction server. *Nucleic Acids Res.* **33**, W72–W76 (2005).
60. Hon, J. *et al.* SoluProt: Prediction of soluble protein expression in *Escherichia coli*. *Chemrxiv* https://doi.org/10.26434/chemrxiv.13047818.v1 (2022).
61. Mészáros, B., Erdos, G. & Dosztányi, Z. IUPred2A: Context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res.* **46**, W329–W337 (2018).
62. Ramirez-Alvarado, M., Kelly, J. W. & Dobson, C. M. *Protein Misfolding Diseases: Current and Emerging Principles and Therapies* (Wiley, 2010).
63. de Groot, N. S., Castillo, V., Graña-Montes, R. & Ventura, S. AGGRESCAN: Method, application, and perspectives for drug design. in *Computational Drug Discovery and Design* (ed. Baron, R.). 199–220. (Springer, 2012).
64. Conchillo-Solé, O. *et al.* AGGRESCAN: A server for the prediction and evaluation of 'hot spots' of aggregation in polypeptides. *BMC Bioinform.* **8**, 65 (2007).
65. Baldwin, R. L. & Rose, G. D. How the hydrophobic factor drives protein folding. *Proc. Natl. Acad. Sci. USA* **113**, 12462–12466 (2016).
66. Madeira, F., Madhusoodanan, N., Lee, J., Tivey, A. R. N. & Lopez, R. Using EMBL-EBI services via web interface and programmatically via web services. *Curr. Protoc. Bioinform.* **66**, e74 (2019).
67. Chen, R., Li, L. & Weng, Z. ZDOCK: an initial-stage protein-docking algorithm. *Proteins* **52**, 80–87 (2003).

68. Haider, M. K., Bertrand, H.-O. & Hubbard, R. E. Predicting fragment binding poses using a combined MCSS MM-GBSA approach. *J. Chem. Inf. Model.* **51**, 1092–1105 (2011).

Acknowledgements
BVLS thanks Dr. Castiglione F for providing clarifications associated with the C-IMMSIM model and simulation parameters. DST-SERB-CRG and ICMR-ISRM are gratefully acknowledged. PARAM Yukti Facility under the National Supercomputing Mission(NSM), Government of India at the Jawaharlal Nehru Centre for Advanced Scientific Research (JNCASR); Bangalore is gratefully acknowledged.

Author contributions
Conceptualization: B.V.L.S. Investigation: B.V.L.S. Methodology: B.V.L.S., V.S.A. Software: B.V.L.S., S.M.R., V.S.A., K.P.S. Supervision: B.V.L.S. Validation: B.V.L.S. Writing-original draft: B.V.L.S. Writing-review & editing: B.V.L.S., S.M.R., V.S.A.

Funding
BVLS gratefully acknowledges DST-NSM (National SuperComputing Mission) for providing the financial support (DST/NSM/R&D_HPC_Applications/Sanction/2021/16).

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-16445-3.

Correspondence and requests for materials should be addressed to V.L.S.P.B.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.