# Supplementary Figures

A graph with different colored bars

Description automatically generated

**Figure S1.** **Peptide array reactivity and nonreactivity grouping.** Total number of peptides with raw median binding above cut-off values (arbitrary units) presented as peptide count. 3D10 reactive peptides were defined as peptides with a raw median binding value of 1024 or greater. Forty-three peptides were classified as 3D10 reactive, and 4335 peptides were classified as 3D10 nonreactive.

A graph of different numbers

AI-generated content may be incorrect.

**Figure S2. Frequency of sequence elements occurrence per peptide for DBPII, PvEBP2, VAR2CSA (FCR3 and NF54).** The frequency of lysine to occur is presented as its proportion relative to other amino acids; lysine represents ~10% of all amino acids for each protein. The total proportion of peptides that contain either a ‘KY’ or ‘YK’ sequence pattern is indicated for each protein.

**Table S1:** Feature descriptions

|  |  |  |
| --- | --- | --- |
| **Feature categories** | **Feature count** | **Feature category description** |
| Single amino acid counts | 20 | Counts of all 20 amino acids |
| Dipeptide counts | 400 | Counts of 2 amino acid combinations |
| Secondary structure | 3 | Odds for a peptide segment to be an alpha helix, beta sheet, or coilᵃ |
| Physicochemical values | 3 | Sums of physicochemical values for every residue based on polarityb, sidechain energyc, and hydropathyd |
| Charge | 3 | Counts of total positively-charged, negatively charged residues, and the sum of these two values (net charge) |
| Disulphide bond potential | 1 | A Boolean value denoting whether a peptide contains 2 cysteines 3 amino acids apart |

**Table S2**: Selected features

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Features** | **Variance feature selection** | **Feature utilization** | **Feature elimination (5) a** | **Feature elimination (10)a** | **Feature elimination (15) a** |
| Positive residue count |  |  |  |  |  |
| Side chain energy |  |  |  |  |  |
| VM count |  |  |  |  |  |
| KQ count |  |  |  |  |  |
| Asparagine count |  |  |  |  |  |
| Lysine count |  |  |  |  |  |
| KY count |  |  |  |  |  |
| Arginine count |  |  |  |  |  |
| Histidine count |  |  |  |  |  |
| Long sulphide bond potential |  |  |  |  |  |
| VT count |  |  |  |  |  |
| VN count |  |  |  |  |  |
| VQ count |  |  |  |  |  |
| Alanine count |  |  |  |  |  |
| Valine count |  |  |  |  |  |
| Net charge |  |  |  |  |  |
| Helix probability |  |  |  |  |  |
| Hydropathy |  |  |  |  |  |
| Coil probability |  |  |  |  |  |
| Negative residue count |  |  |  |  |  |
| Polarity |  |  |  |  |  |
| MT count |  |  |  |  |  |
| IQ count |  |  |  |  |  |
| Sheet probability |  |  |  |  |  |
| TW count |  |  |  |  |  |

a Feature elimination reduces features to a specified value indicated in parentheses

A close-up of a document

AI-generated content may be incorrect.

**Figure S3. Sequence similarity of DBPII and PvEBP2.** Sequence similarity and identity were determined by BLAST using a compositionally adjusted score matrix 1,2.

A close-up of a graph

AI-generated content may be incorrect.

**Figure S4. AlphaFold 3 generated VAR2CSA (FCR3) structure evaluation.** Structures generated by AlphaFold 3 were evaluated based on their similarity to previously resolved structures (A) and prediction confidence (B-C). (A) RMSD matchmaking overlay was performed using Chimera 1.8x internal software. AlphaFold 3 predicted structure derived from the reference cryo-EM resolved VAR2CSA (FCR3) structure (PDB ID 7JGE) 3. Pruned RMSD score and whole RMSD score calculated by Chimera 1.8x and presented unaltered. (B) Per-residue confidence values, predicted local distance difference test (plDDT), are colour-coded onto respective residues of VAR2CSA (FCR3) AlphaFold 3 predicted structure. (C) Predicted aligned error plot of AlphaFold 3 predicted VAR2CSA (FCR3). Plot denotes expected positional error in a pair-wise manner to other residues.

**Table S3:** Bootstrapped distance probabilities between sequence elements

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Boot-strapped odds of amino acid distances | | | | | | |
| Distances | Number of 'other' residues per peptide | | | | | |
| 1 | 2 | 3 | 4 | 5 | 19 |
| 0 | 10.00% | 19.47% | 28.42% | 36.84% | 44.74% | 0.53% |
| 1 | 9.47% | 16.43% | 21.22% | 24.15% | 25.52% | 0.00% |
| 2 | 8.95% | 13.68% | 15.47% | 15.30% | 13.93% | 0.00% |
| 3 | 8.42% | 11.23% | 11.00% | 9.38% | 7.33% | 0.00% |
| 4 | 7.89% | 9.06% | 7.64% | 5.61% | 3.81% | 0.00% |
| 5 | 7.37% | 7.19% | 5.20% | 3.34% | 2.03% | 0.00% |
| 6 | 6.84% | 5.61% | 3.52% | 2.04% | 1.15% | 0.00% |
| 7 | 6.32% | 4.33% | 2.41% | 1.30% | 0.68% | 0.00% |
| 8 | 5.79% | 3.33% | 1.71% | 0.85% | 0.40% | 0.00% |
| 9 | 5.26% | 2.63% | 1.24% | 0.54% | 0.22% | 0.00% |
| 10 | 4.74% | 2.11% | 0.87% | 0.33% | 0.11% | 0.00% |
| 11 | 4.21% | 1.64% | 0.58% | 0.18% | 0.05% | 0.00% |
| 12 | 3.68% | 1.23% | 0.36% | 0.09% | 0.02% | 0.00% |
| 13 | 3.16% | 0.88% | 0.21% | 0.04% | 0.01% | 0.00% |
| 14 | 2.63% | 0.58% | 0.10% | 0.01% | 0.00% | 0.00% |
| 15 | 2.11% | 0.35% | 0.04% | 0.00% | 0.00% | 0.00% |
| 16 | 1.58% | 0.18% | 0.01% | 0.00% | 0.00% | 0.00% |
| 17 | 1.05% | 0.06% | 0.00% | 0.00% | 0.00% | 0.00% |
| 18 | 0.53% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |

Citations

1 Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389-3402 (1997). <https://doi.org:10.1093/nar/25.17.3389>

2 Altschul, S. F. *et al.* Protein database searches using compositionally adjusted substitution matrices. *FEBS J* **272**, 5101-5109 (2005). <https://doi.org:10.1111/j.1742-4658.2005.04945.x>

3 Ma, R. *et al.* Structural basis for placental malaria mediated by Plasmodium falciparum VAR2CSA. *Nat Microbiol* **6**, 380-391 (2021). <https://doi.org:10.1038/s41564-020-00858-9>