Electrostatic potential is the dominant force in antigenic selection of naïve T-cells and B-cells for activation and maturation.

Kripa N. Nand, Christopher Bystroff*
Department of Biological Sciences, Rensselaer Polytechnic Institute, Troy NY

Abstract

Peptide antigenicity can be predicted from sequence using a simple method invented by Hopp and Woods in the early 1980's. Since then, a much clearer understanding of T-cell/B-cell signaling and maturation calls for a new understanding of the amino acid determinants of antigenicity. We show that short peptides with more charged side chains generate significantly higher titers of peptide specific antibodies in co-immunized mice. Peptide docking simulations using linearized Poisson-Boltzmann calculations of electrostatic potential show that immunoglobulins distinguish the cognate peptide sequence from randomly selected sequences at "arms length" (10 - 20Å) with >70% of alternative sequences having higher energy at this distance, consistent with the weak specificity observed for naive T-cell and B-cell receptor interactions with MHC-bound antigen. Orders of magnitude lower complexity of the state space of charged surfaces as compared to the state space of surface shapes suggests a dominant role of electrostatics in selecting naive immune cells from the population of circulating cell lines. We propose a two-stage antigen recognition process, first electrostatic and then shape-based, that explains the dominant contribution of charged residues to peptide immunogenicity.

Introduction

Peptide-based vaccines offer the potential to target antibodies to specific functional sites or broadly neutralizing epitopes (1). Although weak immunogens by themselves, peptides that are arrayed on a carrier particle raise long-lasting sequence-specific immunity. But not all peptide sequences raise high immunity, even when all other factors are held constant. The mechanisms for selective recognition of surface-displayed peptides by T-cells remains a mystery (2)

The ties between peptide sequence and peptide immunogenicity have long been part of the immunology curricula. Hopp and Woods (3) assigned strong positive values to charged amino acids (D,E,K,R) and negative values to non-polar residues. Glycine and uncharged polar side chains were assigned near-zero scores. By averaging the score over a short window (typically 6 residues), Hopp-Woods (HW) scores reproduced the pattern of known B-cell epitopes over well studied proteins (4). We show here that an even simpler formula works just as well. If you simply count the number of charged residues (NC), the number correlates positively and significantly with serum antibody concentrations in mice vaccinated with 18 different peptide conjugates.

Despite the widespread acceptance of the hydrophilicy/immunogenicity connection, there is little published on its mechanistic underpinnings. In 1981, when Hopp and Woods observed the connection, knowledge of the germinal centers (GC) consisted of broad generalities about lymphatic immunoglobin production, high mitotic activity, and the mysterious light and dark zones (5). Vastly more is known now (6). Clonal expansion of hypermutated naîve B-cells, followed by selection against undigested antigen presented on follicular dendritic cells (FDC), lead to B-cells displaying high-affinity receptors (BCR). Avidity of BCR for FDC decides the B-

cell's fate, either apoptosis, further clonal expansion, or differentiation into plasma B and memory B-cells.

T-cell epitope recognition must play by the same biophysical rules as that of B-cells. The major histocompatibility complex 1 (MHC1) -antigen•naïve-TCR tripartite interaction is weak and degenerate, consistent with weak and degenerate surface charge patterns. Patterns of charge complementarity are inherently less complex than patterns of shape complementarity, thus more likely to exist by chance within a population of naïve T-cell lines. The naïve T-cells chosen for trafficking to, and maturation in, the secondary lymphatic organs, may already be charge complementary but not necessarily shape complimentary to a surface epitope on the antigen.

Here we propose a biophysical model for recognition of antigen by BCR/TCR that might shine light on the prevalence of charged groups in highly antigenic epitopes. A computational model for peptide/antibody docking is used here to demonstrate charge-based peptide sequence recognition at "arms length" (10 - 20Å from the bound position) while the shape complementarity forces of hydrogen bonding and desolvation exert forces at very short distances (7,8). Coulombic forces decay slowly with distance so that electrostatic forces persist and sometimes even *increase* with distance in the case of charge-dense protein surfaces.

Here we provide experimental verification of the correlation between Hopp-Woods score and immunogenicity under controlled circumstances in mice. The statistical correlation between ELISA and hydrophilicity is significant, even more so if we use a simple count of the charged residues, ignoring the Hopp-Woods scores for non-charged residues. We offer the results of a peptide docking simulations based on the known structures of five peptide•immunoglobulin complexes, showing that the cognate peptide sequence is at least partially distinguished from random sequences by electrostatic potential alone.

Results

Eight mice were injected with a mixture of 5 µg each of 18 peptide-KLH conjugates adsorbed on alum. Antibody titer trajectories were measured over time (Figure 1). Day 45 titers were compared across peptide sequences. As expected, the Hopp-Woods score correlated with titer strongly, with r=0.73 and p-value=0.001. The NC scoring function correlated even more strongly, with r=0.814 and p-value=0.0002 (Figure 2).

Five crystal structures of peptide•immunoglobulin complexes were used to calculate the minimum distance-dependent, orientation dependent, intermolecular electrostatic potential E_P. The sequence of each peptide was reassigned to a sequence randomly drawn from a globular protein to assess the specificity of E_P, only, for recognition of the cognate peptide, absent any consideration of short-range forces, desolvation, hydrogen bonding, or steric interactions.

For each of the five cases, the cognate peptide sequence electrostatic potential occupies among the lowest three of all 11 sequences plotted (Figure 3). For example, at a displacement of 15Å, the five cognate peptides ranked 1st, 2nd (twice) or 3rd (twice) out of 11 in each set. The chance likelihood of these rankings, based on random scrambling of the rank order, is p=0.0015. At distances just out of reach of direct contacts, the electrostatic potential is the sole driving force of binding and specificity. Therefore, sequence-specific attractive forces effectively increase the local concentration of electrically complementary peptides in the near

vicinity of the binding site, increasing the on-rate of binding. The off-rate from the fully bound state depends, of course, on shape-dependent forces including desolvation, hydrogen-bonding, and van de Waals contacts.

Discussion

Previous studies of correlations between peptide sequence score (Hopp-Woods and others) and immune response found weak correlations (r=0.20±0.05) (9) compared to those we found in this study (r=0.814). The reason for the low correlations may have been that Rockberg & Uhlén used much longer peptides. Although the length is not specified in that paper, prior work by the author Uhlen using the same protein epitope sequence tag (PrEST) library specified length 50 to 150 amino acids (10), much larger than MHC1 binding site, which can accommodate only 8 to 10 residues. Furthermore, we immunized mice with all 18 antigens together, factoring out any possible variations across individual animals.

Biophysics has a voice in explaining the observed correlation between charge content and immunogenicity. BCR feel the attractive force of antigen through charge complementarity first, before sampling shape complementarity. This happens at "arms length", or a distance just beyond one layer of solvation. Water molecules in this layer extend and diffuse the electrostatic field by orienting their dipoles against the underlying charged side chains, moving the attractive and repulsive interactions to the surface of the first hydration shell (11). At distances beyond shape recognition, BCR can freely reorient with respect to the epitope without significant steric hindrance, and without incurring an entropic penalty. Figure 3(f) shows space between peptide and immunoglobulin binding site at 15Å of displacement, the distance at which peptide sequence recognition is most strongly observed in the simulations (Figure 3(a-e)).

In the germinal centers (GC), weak interactions arrayed across an antigen particle combine to signal clonal expansion of naïve B-cells through clustered BCR interactions. Peptides with more charged residues interact more strongly with naïve BCR, whose binding sites likely lack shape complementarity but which may have charge complementarity because of the smaller complexity of the spatial arrangement of charges as compared to shapes (see Methods). Thus peptides with more charges will move more efficiently into the dark phase of the GC for somatic hypermutation and clonal expansion.

Naïve B-cells acting as professional antigen presenting cells (APC) activate and are activated by CD4+ T-helper cells (Th2) that recognize the MHC2-presented foreign peptides. Selection of the peptides to be presented on the MHC2 is done within the cell by the transporter associated with antigen processing (TAP) and by the MHC2 itself, providing two additional sources of sequence-dependency in epitope selection. But TAP exerts little specificity on the sequence of the peptides presented to MHC1 (12,13). The constitutively expressed proteosome, which cleaves antigen into peptides in the phagosomes, does have a preference for arginine and lysine side chains in the N-terminal end the peptide, but the immunoproteosome expressed in hemopoietic cells has less of this preference (14). Moreover, no correlation between overall charge and immunogenicity was observed in our studies.

Interactions between Th2 and MHC2-bound peptides on naïve B-cells acting as APCs are of course governed by the principles of biophysics. Long-range electrostatic interactions, not short-range shape-based interactions, dominate in the selection of B-cell lines for maturation as much as they do in the maturation process itself. Indeed, the presence of complementary shape in the naïve B-cell is even less likely than in the maturing B-cell, the latter having passed rounds of FDC screening for shape complementarity. Thus electrostatic interactions are probably more important at the naïve BCR stage than they are during maturation.

CD8+ cells have a lesser role in humoral immunity, and therefore MHC1 and components of the corresponding peptide processing pathway are not likely part of the mechanism for the sequence/immunogenicity correlation as measured by antibody titer. MHC1 peptide loading involves peptide sequence preferences (15,16), and these may influence the cellular immunogenicity as a correlate of amino acid content, but MHC1 is also extremely polymorphic across individual animals (17), lessening its likely role in establishing the broad correlation that we observe and are trying to explain.

Although this state space calculations are imprecise, there is clearly at least an order of magnitude difference between the complexity of shape space and that of the space of side chain charge arrangements. In our estimations roughly one in 200 naïve BCR/TCR can recognize the arrangement of five charged side chains on an 8-residue epitope. Peptides with fewer charges would be recognized by more naïve receptors, but binding would be weaker, while peptides with more charges would be recognized by fewer naïve receptors but binding would be stronger. This number contrasts with the roughly one in ten-thousand naïve BCR/TCR that can recognize an 8-residue epitope by its 3D shape. The contrast in the size of the two state spaces may be even greater, since the energy of close contacts is inherently sensitive to small differences in shape, while coulombic interactions beyond contacting distances have no hard boundaries. Thus the likelihood of a chance encounter with a naïve BCR/TCR that recognizes and binds an 8-residue peptide by its charge is at least an order of magnitude higher than the chance of recognizing the peptide by its shape. It makes sense that it is the electrostatic interactions that carry cells forward into the maturation process and thus define the more immunogenic epitopes from the less immunogenic ones.

Further consideration should be made to the state of the antigen when interacting with naïve cells (MHC1/MHC2-bound) versus the state of the antigen during B-cell maturation (undigested, FDC-presented). The conformation of the MHC-bound peptide is largely defined by its binding to MRC, and is surrounded by MHC on three sides (18). As such, MHC-bound peptide is both non-native and inaccessible. However, since the peptide retains its arrangement of charged side chains, its recognition by naïve BCR is meaningful.

But during maturation in the GC, the antigen is presented in its native, undigested state, allowing shape to play a meaningful role in selecting B-cells from a library of clones. Certainly, stronger shape-based forces take over the selection process. Indeed the epitope-paratope interactions seen in immunoglobulin complex crystal structures always exclude significant solvent (19). But at that point in the process, the selection of the naïve B-cell line to carry forward into maturation in the GC has already been made. Refinement of shape-dependent interactions by clonal expansion only finishes the job.

Our hypothesis, that naïve B and T-cell clonal selections made by APC-presented antigen are made at an arms length distance, explains the dominance of charged residues in the correlation of sequence with antibody titer. The simplification of the HW score to the NC score points to the unimportance of non-charged side chains at the decision points of immunogenicity.

Conclusions

All signaling in the immune system is mediated via physical interactions between biomolecules. These interactions can be dissected into contacting and non-contacting, where the latter comprises electrostatic forces. Here we argue that electrostatic interactions dominate in the selection process of naïve cells for maturation, citing three reasons. First, electrostatic

interactions are lower in complexity than shape-based interactions, increasing the odds of a random attractive encounter with a naïve B or T cell. Second, electrostatic forces are longer-range than shape-based forces such as hydrogen bonding, desolvation and van der waals contacts. Thus a separation distance range exists (10-20Å) where electrostatic forces are both selective and dominant. Third, cell-cell interactions involving antigen happen in the context of the MHC complex, which retains the arrangement of charges on the peptide but not its three-dimensional conformation. Taken together, it is unlikely that shape plays much of a role in signaling cells for maturation, leaving that task to electrostatic interactions. This view is consistent with the experimental observation of a strong correlation between charged side chain content and specific antibody production.

Materials and Methods

Antibody titer

Peptide immunogenic potential is quantified using ELISA. Immunizations containing 5 µg each of 18 different peptides (Table 1), each individually conjugated to KLH (keyhole limpet hemocyanin) and adsorbed onto alum were injected intramuscularly into eight, 6-week old, FVB/J mice thrice at 15 day intervals. Sera were collected at Day 45 and peptide specific antibody concentrations were determined by ELISA (A490) at a 1:500 dilution. Day 45 sera were used rather than 60 or 90 day sera because the latter measurements were at or near the instrument saturation limit for some mice, and the Day 45 choice was justified by the strong correlation (r=0.83-0.96) between Day 45 titer and Day 60 titer when the latter was not at saturation. Day 0 sera were used as the control. Peptides ranged in length from 14 to 22 residues. The median A490 value of each group of 8 was plotted versus the peptide Hopp-Woods score or the number of charges (NC) score. The correlation (*r*) was determined by Pearson's method (20). To test *r* for significance, the data were randomly scrambled and *r* was recalculated for each scramble. Scrambling was carried out 10,000 times. The *p*-value for the correlation was equal to the number of times that the scrambled *r* value was equal to or greater than the unscrambled *r* value, over 10,000. Multiple such tests showed that the *p*-value estimate was highly reproducible.

Electrostatic potential calculation

The electrostatic potential E_P was calculated using the linearized Poisson-Boltzmann (P-B) equation (21), $E_P = \sum \psi_{ij}$, where, $\psi_{ij} = \psi_0 exp \frac{-r_{ij}}{\lambda_D}$, and $\psi_0 = \sum \frac{q_i q_j}{\varepsilon r_{ij}^2}$, where q_i is the charge on point i. The sum is over all atoms i in ligand (peptide) and j in receptor (antibody). The linearized P-B is salt-dependent in two terms, the Debye length (λ_D) and the permittivity ε= 8.854e-12 (22). Charges were ignored for amino and carboxy termini. Approximate partial charges were assigned to side chain atoms for Arg, Lys, Asp, Glu and His as follows: Arg NH1 +0.5, Arg NH2 0.5, Lys NZ +1.0,Asp OD1 -0.5, Asp OD2 -0.5, Glu OE1 -0.5, Glu OE2 -0.5, His ND1 and NE2 +0.5x/(1+x), where $x=10^{7\text{-pH}}$. Energies due to desolvation, steric interactions, hydrogen bonds, stereochemistry and all intramolecular interactions are ignored for this experiment since we are looking at medium to long-range intermolecular interactions only.

Peptide docking

Five unrelated immunoglobulin•peptide crystal structures (PDBids: 2HRP, 1IFH, 1SVZ, 1HI6, and 1U8K) were separated into receptor and peptide coordinates. Ten new peptide sequences were selected for each of the five peptides and coordinates for these new sequences were generated by rotamer search and energy minimization using MOE's Protein Modeler tool.

During energy minimization the peptide was placed in its bound position relative to the receptor. All peptide atoms and nearest neighbor receptor atoms were unfixed, while all other atoms and all alpha carbons were fixed. Sequences were selected randomly from non-overlapping segments of mouse catsper beta (NP_766611.2) or *Phodopus roborovskii* dihydrofolate reductase (CAH6787515) without regard to structure, omitting any sequence that contained no charged side chains. Peptides were displaced from their bound positions along a vector connecting receptor and peptide centers of mass, from zero to 30Å in 1Å steps. At each displacement distance 1000 random orientations was calculated and energetically evaluated. The rotation matrix for each random orientation was calculated by selecting three random numbers between -1 and 1 to define the rotation axis, omitting cases where the corresponding vector length was greater than 1.0 in order to sample rotational space evenly, then selecting a random number between 0 and 360° to define the amount to rotate about that axis. A random displacement of ±0 to 2Å was applied in a random direction normal to the vector of displacement. Peptide orientations having close contacts with any atom of the receptor (<5.8Å) were ignored. For each displacement distance, E_P was calculated for all 1000 random orientations, keeping only the lowest E_P. For each peptide sequence, lowest E_P were plotted (Figure 2).

Complexity calculations

Complexity is defined as the number of states. There are many ways to calculate the shape complexity of a surface manifold (23,24), which in this case is the part of a protein's surface responsible for binding. The relevant measure of complexity must relate to the binding energy, therefore the calculation must be independent of reference frame, have the characteristic smoothness of a protein surface, and must be coarse-grained enough to exclude shape differences within natural protein flexibility. Complexity of shape is dependent on the interaction surface area (25), which for antibody/antigen interactions is a function of the length of an epitope. Here we consider a linear epitope of 8 amino acid residues. An empirical measure of shape complexity may be derived from peptide local structure complexity in proteins. Cluster analysis of peptide conformations in proteins (26,27) found that the vast majority of peptides of lengths 5 - 15 residues could be classified into 13 structural classes and subclustered into 262 sequence-structural classes (28). In that work, an average residue position within a cluster had 2 to 4 possible side chain choices, ignoring rotameric variability. Considering 2 to 4 options at 4 positions (one face of an 8-mer peptide is roughly 4 positions), the shape complexity for a typical peptide epitope is 262 times 2^4 to 4^4, or 4000 to 70000 states.

The complexity of electrostatic interactions at long distances is 2 (attractive or repulsive), since all pairwise interactions approach point charges at long distances. At nearer distances, the force depends on the spatial arrangement of the charges. For this complexity calculation, we take the number of positions in space to be the number of side chains, in this case 8, times the number of ways 5 positive or negative charges can be arranged on those 8 positions (Five was selected as the smallest number of charges that gives a maximum antibody response in our experiments, regardless of the sign of the charges. There was no observed correlation of antibody titer with total charge.), times the number of ways plus and minus formal charges can be assigned to *n* positions, or 2^n. On average, half of the five (2 or 3) side chains are facing the binding site and contribute more to the electrostatic potential. The resulting complexity ranges from 8-choose-2 times 2^2, or 112, to 8-choose-3 times 2^3, or 448. The average shortest distance between charged groups within a peptide, based on the 55 peptides used in the simulations, was 10Å. Therefore as surfaces approach 10Å separation, the directionality of the intermolecular charge-charge interaction vectors becomes increasingly non-collinear and thus increasingly spatial arrangement-dependent. Roughly 100 to 400 is the range of ways 5

charges can be arrayed in space and be energetically differentiable at 10 - 20Å separation distance.

Tables

Table 1. Peptide sequences.

Figures

Figure 1. Antibody titers versus time .

Figure 2. Antibody titers versus sequence.

Figure 3. Simulations of peptide docking.

1	**NSLSYSFYNHSLFR**
2	**DFQMDEREYAME**
3	**EIELMESTNTALWP**
4	**YIDNRAQGAWYII**
5	**LVVLLGVPAHSVWLQ**
6	**RDYTEEEIFRYNSPLDTTNSLI**
7	**WSTFESDIENEEPFLW**
8	**QTWDSMIEENPDIPLDDVWG**
9	**CDGTVYLRTEDEFTKLDES**
10	**CHIPPEDWISGVHKDSQGFNM**
11	**CVIDKELLRESLSDNLK**
12	**CHEDNKDNPLL**
13	**CNEPENVKKMKHYLEPLLKTPVYN**
14	**CKGYSYDYYENTWRKLEGISEP**
15	**CTKVERTTEDKKFYIMSHESPG**
16	**CFDSVIKDAEMPSF**
17	**CIDKKRASEQGMIGRNIKKT**
18	**GRQSIELAMEENEKRNIIC**

Table 1. Amino acid sequences of KLH-conjugated peptides used in this experiment. Some peptides were conjugated using a N- or C-terminal cysteine, others using the N-terminal amino group.

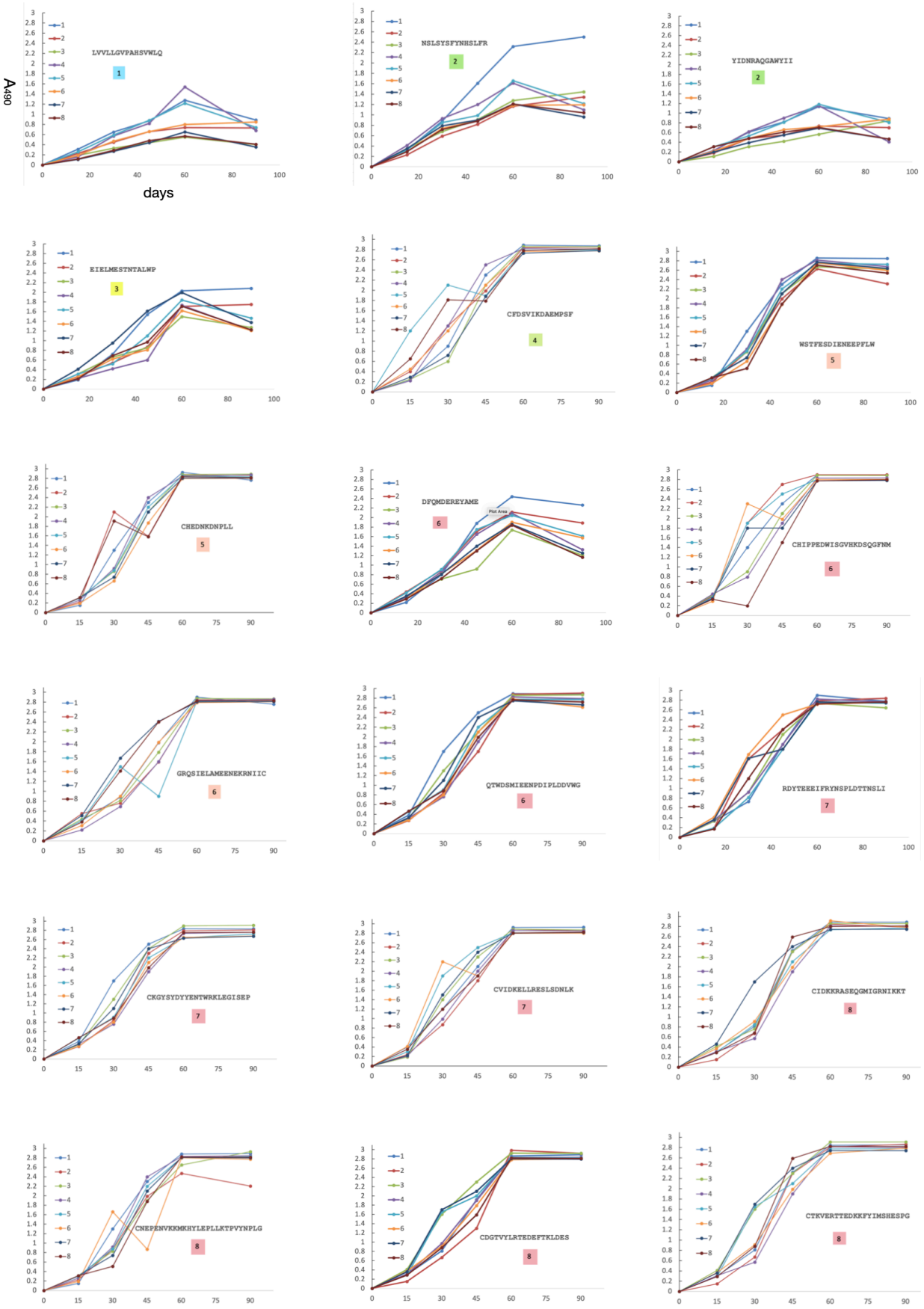
A490
days
LVVLLGVPAHSVWLQ
1
NSLSYSFYNHSLFR
2
YIDNRAQGAWYII
2
EIELMESTNTALWP
3
CFDSVIKDAEMPSF
4
WSTFESDIENEEPFLW
5
CHEDNKDNPLL
5
DFQMDEREYAME
6
CHIPPEDWISGVHKDSQGFNM
6
GRQSIELAMEENEKRNIIC
6
QTWDSMIEENPDIPLDDVWG
6
RDYTEEEIFRYNSPLDTTNSLI
7
CKGYSYDYYENTWRKLEGISEP
7
CVIDKELLRESLSDNLK
7
CIDKKRASEQGMIGRNIKKT
8
CNEPENVKKMKHYLEPLLKTPVYNPLG
8
CDGTVYLRTEDEPTKLDES
8
CTKVERTTEDKKFYIMSHESPG
8

Figure 1. Antibody titers (A490) were determined by ELISA at 1:500 dilution for sera from eight vaccinated FVB/J mice at 15-day intervals. Legend: mouse number. Insets: sequence of immobilized, KLH-conjugated peptide. Number of charged side chains, with heat-map background color. Correlations were calculated using Day 45 data.

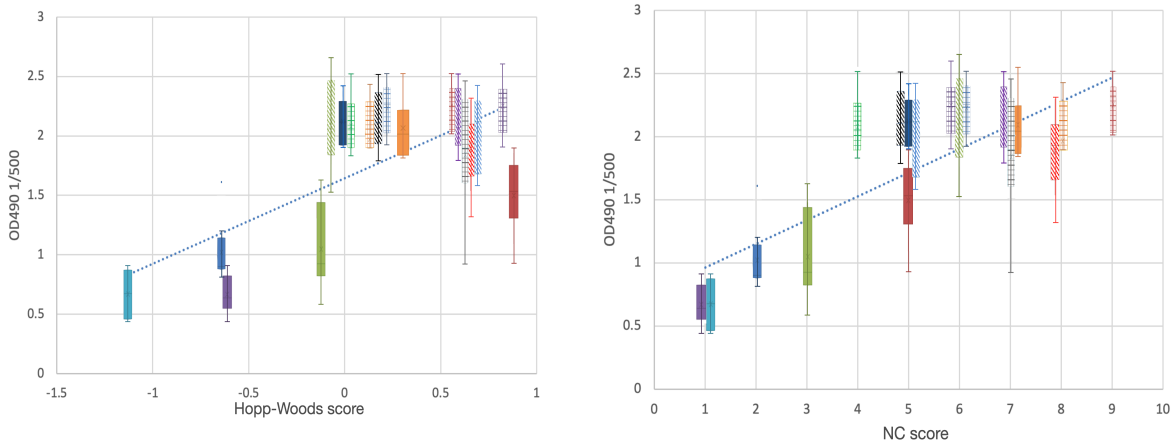

Figure 2. Ranges are shown for measurements from the sera of eight mice, vaccinated with 18 peptide-KLH conjugates, ranging in length from 14 to 22 residues. Day 45 ELISA

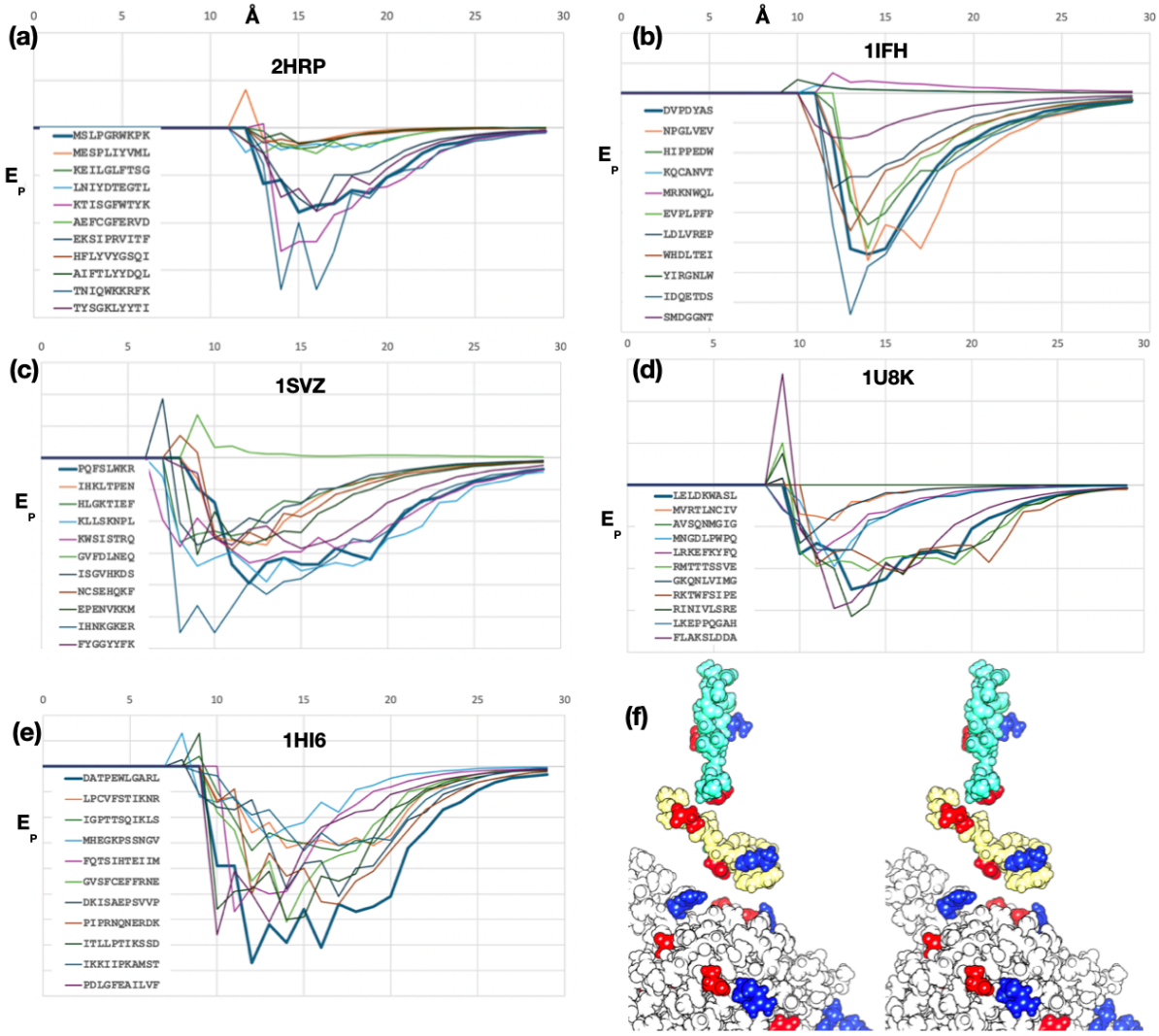

Figure 3. Simulations of peptide docking to known immunoglobulin structures using electrostatic potential, excluding distances and orientations where steric interactions are present. (a) 2HRP versus cognate peptide sequence MSLPGRWKPK (thick line) and ten random sequences. (b) 1IFH versus cognate peptide sequence DVPDYAS (thick line) and ten random sequences. (c) 1SVZ versus cognate peptide sequence PQFSLWKR (thick line) and ten random sequences. (d) 1U8K versus cognate peptide sequence LELDKWASL (thick line) and ten random sequences. (e) 1HI6 versus cognate peptide sequence DATPEWLGARL (thick line) and ten random sequences. (f) Cognate peptide docking for 1HI6, showing the lowest energy orientations at 30Å (cyan) and 15Å (yellow) from the bound position. The peptide reorients to the immunoglobulin surface at 15Å in this simulation.

References

1. Malonis, R. J., Lai, J. R., & Vergnolle, O. (2019). Peptide-based vaccines: current progress and future challenges. Chemical reviews, 120(6), 3210-3229. Chicago
2. Lee, C. H., Salio, M., Napolitani, G., Ogg, G., Simmons, A., & Koohy, H. (2020). Predicting cross-reactivity and antigen specificity of T cell receptors. Frontiers in Immunology, 11, 565096.
3. Hopp, T. P., & Woods, K. R. (1981). Prediction of protein antigenic determinants from amino acid sequences. Proceedings of the National Academy of Sciences, 78(6), 3824-3828.
4. Berzofsky, J. A. (1985). Intrinsic and extrinsic factors in protein antigenic structure. Science, 229(4717), 932-940. Chicago
5. Lennert, K., & Stein, H. (1982). The germinal center: Morphology, histochemistry, and immunohistology. In Lymphoproliferative diseases of the skin (pp. 3-15). Berlin, Heidelberg: Springer Berlin Heidelberg.
6. Casola, S., Otipoby, K. L., Alimzhanov, M., Humme, S., Uyttersprot, N., Kutok, J. L., ... & Rajewsky, K. (2004). B cell receptor signal strength determines B cell fate. Nature immunology, 5(3), 317-327.
7. Dill, K. A. (1990). Dominant forces in protein folding. Biochemistry, 29(31), 7133-7155.
8. Van Oss, C. J. (1995). Hydrophobic, hydrophilic and other interactions in epitope-paratope binding. Molecular immunology, 32(3), 199-211.
9. Rockberg, J., & Uhlén, M. (2009). Prediction of antibody response using recombinant human protein fragments as antigen. Protein Science, 18(11), 2346-2355.
10. Zeiler, M., Straube, W. L., Lundberg, E., Uhlen, M., & Mann, M. (2012). A Protein Epitope Signature Tag (PrEST) library allows SILAC-based absolute quantification and multiplexed determination of protein copy numbers in cell lines. Molecular & Cellular Proteomics, 11(3).
11. Voloshin, V. P., & Medvedev, N. N. (2021). Orientation of water molecules near a globular protein. Journal of Structural Chemistry, 62(5), 692-703.
12. Lehnert, E., Mao, J., Mehdipour, A. R., Hummer, G., Abele, R., Glaubitz, C., & Tampé, R. (2016). Antigenic peptide recognition on the human ABC transporter TAP resolved by DNP-enhanced solid-state NMR spectroscopy. Journal of the American Chemical Society, 138(42), 13967-13974.
13. Mantel, I., Sadiq, B. A., & Blander, J. M. (2022). Spotlight on TAP and its vital role in antigen presentation and cross-presentation. Molecular immunology, 142, 105-119.
14. Winter, M. B., La Greca, F., Arastu-Kapur, S., Caiazza, F., Cimermancic, P., Buchholz, T. J., ... & Craik, C. S. (2017). Immunoproteasome functions explained by divergence in cleavage specificity and regulation. Elife, 6, e27364.
15. Dönnes, P., & Elofsson, A. (2002). Prediction of MHC class I binding peptides, using SVMHC. BMC bioinformatics, 3, 1-8.
16. McShan, A. C., Devlin, C. A., Morozov, G. I., Overall, S. A., Moschidi, D., Akella, N., ... & Sgourakis, N. G. (2021). TAPBPR promotes antigen loading on MHC-I molecules using a peptide trap. Nature communications, 12(1), 3174.
17. Radwan, J., Babik, W., Kaufman, J., Lenz, T. L., & Winternitz, J. (2020). Advances in the evolutionary understanding of MHC polymorphism. Trends in Genetics, 36(4), 298-311.
18. Altmann, D. M., & Trowsdale, J. (1989). Major histocompatibility complex structure and function. Current Opinion in Immunology, 2(1), 93-98.
19. Mariuzza, R. A., Phillips, S. E. V., & Poljak, R. J. (1987). The structural basis of antigen-antibody recognition. Annual review of biophysics and biophysical chemistry, 16(1), 139-159.
20. Asuero, A. G., Sayago, A., & González, A. G. (2006). The correlation coefficient: An overview. Critical reviews in analytical chemistry, 36(1), 41-59.

21. Chipman, D. M. (2004). Solution of the linearized Poisson–Boltzmann equation. The Journal of chemical physics, 120(12), 5566-5575.
22. Hasted, J. B., Ritson, D. M., & Collie, C. H. (1948). Dielectric properties of aqueous ionic solutions. Parts I and II. The journal of chemical physics, 16(1), 1-21.
23. Anisov, S. (2001). Towards lower bounds for complexity of 3-manifolds: a program. arXiv preprint math/0103169.
24. Gardiner, J. D., Behnsen, J., & Brassey, C. A. (2018). Alpha shapes: determining 3D shape complexity across morphologically diverse structures. BMC evolutionary biology, 18, 1-16.
25. Tan, T., Frenkel, D., Gupta, V., & Deem, M. W. (2005). Length, protein–protein interactions, and complexity. Physica A: Statistical Mechanics and its Applications, 350(1), 52-62.
26. Bystroff, C., & Baker, D. (1998). Prediction of local structure in proteins using a library of sequence-structure motifs. Journal of molecular biology, 281(3), 565-577.
27. Bystroff, C., & Garde, S. (2003). Helix propensities of short peptides: molecular dynamics versus bioinformatics. Proteins: Structure, Function, and Bioinformatics, 50(4), 552-562.
28. Bystroff, C., Thorsson, V., & Baker, D. (2000). HMMSTR: a hidden Markov model for local sequence-structure correlations in proteins. Journal of molecular biology, 301(1), 173-190.