Review: Sagacious epitope selection for vaccines, and both antibody-based therapeutics and diagnostics: Tips from virology and oncology.

Running Title: Sagacity in Epitope Selection

Authors: Samuel Ken-En Gan¹,²,*, Ser-Xian Phua¹, Joshua Yi Yeo¹

Affiliation:
¹ Antibody & Product Development Lab, EDDC-BII, Agency for Science, Technology and Research (A*STAR), Singapore 138672, Singapore; phuasx@gmail.com (S.-X.P.); joshua_yeo@eddc.a-star.edu.sg (J.Y.Y.)
² APD SKEG Pte Ltd, Singapore 439444, Singapore

*Corresponding author: samgan@apdskeg.com; samuel_gan@eddc.a-star.edu.sg;

Keywords: Antigen selection, Epitope selection, Antibody targeting, Epitope Accessibility

Abbreviations: Ig, VH, CH, VL, CL, FWR, CDR, Fc, Fab, V-region, ADCC, AR.
Statement of Significance: The culmination and advances in protein engineering and antibody development has allowed zooming in on the target antigen to epitope selection. Relevant to vaccines, antibody-based therapeutics and diagnostics, the target epitope plays a major role in the success of the interventions. This review discusses these epitope factors, and how with sagacious epitope selection, better antibody-based therapeutics and diagnostics, and vaccines can be potentially developed with higher success.
The target of an antibody plays a significant role in the success of antibody-based therapeutics and diagnostics, and to an extent, that of vaccine development. This importance is focused on the target binding site – epitope, that sagacious epitope selection in a form of design thinking beyond traditional antigen selection using whole cell or whole protein immunisation can positively improve success. Intrinsic factors that can affect the functioning of resulting antibodies can be more easily selected for with purified recombinant protein production and peptide synthesis to display limited/selected epitopes. Many of these factors stem from the location of the epitope that can affect the accessibility of the antibody to the epitope at a cellular or molecular level, direct inhibition of target antigen activity, conservation of function despite escape mutations, and even non-competitive inhibition sites. Through the incorporation of novel computational methods for predicting antigen changes to model-informed drug design development, superior vaccines and antibody-based therapeutics or diagnostics can now be more easily designed, mitigating failures. With detailed examples, this review highlights the new opportunities, factors and methods of predicting antigenic changes for consideration in sagacious epitope selection.
Introduction

Antibodies and their fragments are increasingly important in diagnostics and therapeutics development as evidenced in the ongoing COVID-19 pandemic. An already expensive process, diagnostics can fail due to escape mutations on the epitope that compromise primer-based kits or diagnostic antibody binding even with sagacious rational antibody design and engineering. On therapeutics, antigenic epitope changes leading to escape mutations contribute to the expensive drug failures. To improve success, Model Informed Drug Discovery and Development (MID3) was in pilot by the U.S. Food and Drug Administration since 2018 to support drug development.

The ability to selecting the right single antigen for diagnostics, therapeutics, and to an extent that for vaccines targeting (e.g. choosing only the Spike over a whole virus), was augmented through recombinant technology, where purified target antigens could be produced and either injected into animals or used with in vitro antibody display methods e.g., phage display for antibody selection. The same technology also supported the targeting of specific epitopes on the antigen, where having antibodies specific to an epitope in a diagnostic kit can improve selectivity and specificity. This specificity is useful when differentiating between similar antigens e.g., between viruses such HIV and Hepatitis B (HBV) even when targeting Reverse Transcriptase (RT) or distinguishing SARS-CoV-2 from MERS, and SARS when targeting the RNA-dependent RNA polymerase (RdRp). Yet, being too specific in diagnostics, as opposed for vaccinations, can also result in false negatives when the target epitope on the antigen mutates beyond antibody recognition.

With advances in peptide technology, short/stapled peptides can also be used without going to the DNA level for recombinant expression and over the commonly used whole cells or antigens. Recent developments in cyclic peptides and peptide vaccines allow for immunization against specific conformational structures as epitopes instead of whole antigens.
through mimotopes \(^\text{14}\) as B cell peptides \(^\text{15}\). Such more selective methods can support the development of therapeutics to reduce side effects due to off-target effects, although some level of lower specificity could be of value for vaccines and to an extent, diagnostics to target variants.

The selection of epitopes is thus a paradigm shift from chance-dependent antibody development to a more rational purposeful approach, and it is dependent on the intended application in guiding towards high specificity or to have inclusion of minor changes in which this selection can occur as the two categories \(^\text{16}\) of 1) linear/continuous, defined as a stretch of amino acids sequences; 2) conformational/discontinuous, defined as sequence distal residues in close proximity through protein folding, with the latter conformational type more prevalent as B-cell epitopes.

About 96% of monoclonal antibody therapeutic candidates fail to make it to market \(^\text{17}\), costing close to tens of millions of dollars \(^\text{18}\), augmenting the MID3 approach that includes epitope prediction. Epitope prediction, traditionally based on amino acid physicochemical properties such as hydrophobicity, flexibility, solvent accessibility and antigenicity \(^\text{19-22}\) have seen augmentation by machine learning methods to show promise for cancer \(^\text{23}\) and even hybrid experimental-computational approaches \(^\text{24}\) involving deep neural network for MHC binding \(^\text{25}\) and attention-based LSTM networks \(^\text{26}\).

**Epitopes factors**

**Accessible Epitopes (Cellular)**

For epitope selection particularly for therapeutics, accessibility of the epitope by the antibodies is perhaps the first and foremost to be considered. For maximising success, extracellular targets are typically picked for antibody therapeutic candidates with the exception for ‘intrabodies’
where target cells produce the antibodies against intracellular antigens within itself. Apart from intrabodies, whole antibodies against intracellular targets often have limited penetration of membranes (knowledge known to flow cytometry users) which will require intracellular delivery methods that are not always feasible for therapeutic use. Although unknown in prevalence, there are instances where intracellular oncogenic markers can be targeted because of externalization e.g. PRL3. Nonetheless, antibodies do have intrinsic advantages for undruggable targets that are not suitable for small molecules and the selection and effectiveness of targeting intracellular proteins would require assessment on a case-by-case basis.

With antibody intracellular delivery hurdles yet to be overcome, most current therapeutics and diagnostic whole antibody targets are extracellular antigens. With extracellular targets, there are many factors that can underlie the selection of epitopes. In one example, glycosylation has been shown to impede antibody recognition in both the ongoing COVID-19 pandemic and HIV, with clear impact when occurring at epitope sites. Thus, even with good epitope bioinformatics prediction scores or good experimental results from bacterial produced proteins (where there are rare occasional exceptions), care must be taken for the effects of post-translational modification on occluding or interfering with antibody recognition.

Epitope occlusion in the form of cryptic epitope or cryptotopes is reported in viruses such as the Norovirus, Influenza, Ebola and HIV (Figure 1A), where apart from post-translational modifications methods, HIV utilised the hypervariable regions in the Env to occlude gp41 epitopes from immune detection. On self-antigens, such occlusion of inducible cryptic epitopes play a role in reducing autoimmunity, and in some cases, differentiating disease states when exposed during pathogenesis e.g. prion disease. In the unlikely situation where targeting the viral receptor is not easily achieved, blocking the viral target (host cell
receptor) from viral spike binding can be performed, as in the case for SARS-CoV-2 \(^{40,41}\) and poliovirus \(^{42}\), although care needs to be taken not to end up affecting host cell activity by the antibody by over-activation (e.g. mimicking receptor stimulation and inducing of altered signalling triggering uncontrolled microthrombosis, cell lysis and neutrophil activation \(^{43}\)) or by preventing activation (e.g. inhibiting hyaluronan clearance by liver cells \(^{44}\)).

Antigens have the potential to bind antibodies in non-conventional CDR-dependent manner by inducing binding pockets or the formation of stretches/patches on antibodies. This phenomenon was observed for non-conventional immunogenic molecule such as nickel \(^{45}\) binding to Trastuzumab and Pertuzumab that can possibly underlie the disease pathogenesis of nickel type-I allergy. The other example of induced binding is in the molecular dynamics (MD) simulation of Trastuzumab binding to Her2, inducing a cryptotopes that facilitated Pertuzumab binding \(^{46}\). While the synergism between the two clinical therapeutics were shown experimentally to be due to the different epitopes on the same antigen without evidence of an induced Pertuzumab epitope \(^{47}\), the possibility of occluded epitopes ought to be sagaciously considered depending on the desired application and utility.
Figure 1. Accessibility of Epitopes. (A) Targeting cryptotopes. (B) Accessibility of epitope to differing antibody isotypes. The location of the epitope on the target antigen can affects its accessibility by antibodies that are conjugated or multimeric in nature due to steric hindrances. This is evident in internal epitopes that would not be accessible to multimeric IgMs but accessible to monomeric Igs or antibody fragments. (C) Steric hindrances for multi-valent binding by antibodies due to the location of the epitope on the target antigen. Created with BioRender.com.
Accessible Epitopes (Molecular): Lessons from IgM for multi-specific antibodies

Apart from access at the intra/extracellular level and occluded/induced epitopes, obvious steric hindrances at the molecular level to access epitopes can impact the efficacy of the antibody (Figure 1B). While the earlier example \(^{46,47}\) showed synergistic binding of Trastuzumab and Pertuzumab simultaneously to their different epitopes on Her2, steric hindrances resulting from multiple whole or conjugated or multimeric antibodies are known to arise in flow cytometry\(^{48}\).

When made into multimeric IgMs for multiple antigen binding, Trastuzumab IgM could not have full occupancy of its Fab regions due to steric hindrances binding to multiple Her2. This was however, not in the case for Pertuzumab IgM \(^{49}\), which could then show higher avidity effects \(^{50}\). The sheer size of multimeric antibodies, while advantageous in agglutination for immune clearance and result in a better therapeutic or diagnostic than the traditional IgG, as evidenced in a nasal delivery for SARS-CoV-2 \(^{51}\), need to be watched for possible steric hindrance (Figure 1C). While IgM is used in hemagglutination assays \(^{52-55}\), care needs to consider that the accuracy of such IgM based assays may be affected.

For this reason, the checklist to selecting any epitope for diagnostic or therapeutic application needs to ensure molecular accessibility, especially when intending towards large antibodies (whether conjugated or multimeric or whole). It should be noted that the textbook primary antibody response is typically IgM and steric hindrances may underlie (in part at least) why IgMs tend to have lower affinities for the antigen.

This concern can apply to the development of bi-specific antibodies requiring additional optimization \(^{56}\). Potentially limiting the effectiveness of multi-specific antibodies, especially those are intended to engage whole cells \(^{57}\) despite the promises of doing so in bispecific T-cell engagers (BiTEs) for oncology \(^{58}\), care must be taken of accessibility to maximize success. It is in this area that perhaps a more flexible antibody hinge at the CH1 of antibodies may alleviate some structural constraints as evidenced in IgA \(^{59}\) and other isotypes \(^{50}\).
Conservation of Epitopes: Lessons from Viruses and Omalizumab-IgE

Since escape mutations in the epitope results in antibody recognition failure, one key main criterion of epitope selection is its conservation and this occurs in at least two levels of conservation within the antigen even in the presence of mutations, or conservation within the family to allow broad-spectrum targeting (Figure 2). The earlier level of within the species or viral type is exemplified in the recent COVID-19 pandemic where SARS-CoV2 spike mutations led to decreased effectiveness of the early vaccines to novel variants. On the level for broad-spectrum protection, other viral vaccines such that against the human papillomavirus (HPV) are found to induce cross-neutralizing antibodies.

Likely due to faster escape mutations in microbial pathogens, there are more monoclonal antibody therapeutics against cancer than infectious diseases, where those targeting the latter tend to be polyclonal to cover more epitopes to mitigate risk of escape.

Nonetheless, in the search for conserved regions in pathogens, large sequences databases can provide some insights (such as HBVdb, Los Alamos HCV, HFV, HIV, Stanford HIV Drug Resistance, GISAID, and Nextstrain) but bearing in mind the constraints of extracellular targets for both neutralization of viruses for therapeutics and detection in diagnostics that do not require pre-processing to release internal contents. Though it should be noted that broad-spectrum targets of such nature are expectedly limited due to viral tropism.
Figure 2. Conservation of epitopes across highly mutable target antigens (top) would reduce the chance of escape mutations even as the pathogen accumulates mutations in the target antigen to prolong the effectiveness of the antibody in either detecting or neutralizing the target. Similarly, targeting conserved epitopes that are conserved in similar proteins (e.g., Reverse Transcriptase) across viruses or pathogens can allow broad-spectrum detection or therapeutic intervention (bottom), improving its market use while also useful against emerging viruses utilizing enzymes of the same protein family. Created with BioRender.com.
Foresightedly, the overcoming of cellular level accessibility issues e.g. intrabodies or novel delivery methods to target intracellular targets, conserved viral epitopes can expand to viral enzymes \(^1\) for better search of potential broad-spectrum antivirals. This can be performed through a search of substrate analogues sites that indicate the presence of a functional domain \(^2\) that would be more conserved due to the preservation of enzyme functions.

In antibodies, the broad conserved target of Omalizumab (Xolair®) to IgE Fc allows it to be effective against various IgE type I allergies \(^4\). Since the C-region of antibodies is generally more conserved, Omalizumab was sagaciously raised against the C-region, overlapping with the FcɛRIα binding site. While the existence of allotypes in the Cε-regions \(^5\) can be an issue, another possible effects recently reported that can affect binding is the allosteric communication from the V-region that could influence FcɛRIαa interaction \(^6\). Despite overlapping with that of FcɛRI, the V-region distal allosteric effects had negligible effects on Omalizumab binding. This make the Omalizumab binding side a sagacious epitope that displayed the following benefits: 1) at a conserved region across a variety of pathogenic factors; 2) at the allosterically insulated regions that is robust to changes at hypervariable regions; and 3) relatively immutable.

**Predicting New Mutated Epitopes: Lessons from Viruses**

The need for conserved epitopes in the face of mutations is amplified and shown to be crucial by the challenge to existing COVID-19 vaccines by new variants \(^7\). In the lack of conserved epitopes, the next best is to pre-emptively predict possible escape mutations in the epitopes and design interventions against them. This is not only easily performed through recombinant or peptide methods discussed above, but even easier to implement as vaccines with the rise of mRNA vaccines \(^8\).

Prediction attempts are seen in the polio virus RdRp in vitro platform leading to the development of its vaccine \(^9\) with similar attempts for HFMD causative Enterovirus (EV) A71
These methods of predicting new mutations exhibit great promise despite being *in vitro* and can benefit from *in silico* augmentation involving network analysis, robust statistical model building, time series, stacking models and random signature analysis. Nonetheless, it should be noted that EVs and lentiviruses like HIV, mutate through recombination whereas Influenza use the Orthomyxovirus family method of assortment, requiring further research to get to a unified computation method capable of dealing with RNA viruses of different mutation methodologies.
Figure 3. Schematic of mutation prediction of RNA viruses that can be generated in vitro (adapted from\textsuperscript{87}) to determine innate viral polymerase mutation rates, hotspots and biases for the incorporation into computational methods for better prediction. Such a platform would not only provide insights to the viral polymerases for augmentation towards lethal mutagenesis, but also improve preparedness to identify conserved regions and identify potential mutations of concern that can lead to escape to guide drug and diagnostic kit development. Created with BioRender.com.
Experimentally, there is already notable progress in predicting mutations through the understanding of molecular biology and the innate biases of pathogen enzymes involved in replication (Figure 3). While such predictions are often deemed stochastic and unpredictable, recent evidence has shown certain propensities in the genetic code that can confer a certain degree of predictability. It is certain that with more knowledge about the target biology, the less stochastic and more predictable mutations in the given target can be. This is demonstrated recently where even without immune or drug selection pressures, HIV reverse transcriptase displayed clear mutational biases to induce certain types of mutations in specific locations in HIV genes within a mimic of a single replication cycle. Given that a high percentage of the generated mutations were found in patients, the rise of drug-resistant mutations in the HIV genes were not entirely stochastic nor required intense immune or drug selection pressures to emerge. As the ‘Godzilla’ of fast mutating viruses, studies showed that the cross resistance for protease inhibitors and reverse transcriptase inhibitors could be linked by network analysis. Such analysis showed how mutations can confer not just resistance to one inhibitor, but to others in cross-resistance. Thus, in the vein of MID3, structural modelling studies can study the possible effects of likely emerging mutations with respect to drug resistance for this is not an inexhaustive range of possibilities given that the mutations would be constrained by the function of the target. And it is also in this area that therapeutics that augment lethal mutagenesis for error catastrophe could be an alternative strategy for undruggable targets.
Enabling better direct inhibition: Lessons from Her2 and IgM

While alternative strategies exist as in non-competitive non-nucleoside RT inhibitors or NNRTIs, the related above-mentioned lethal-mutagenesis-based methods, direct inhibition of targets is the most common approach of therapeutics. While it is expected that epitopes at activation sites are key to this, consideration of the antibody can augment the success of this where larger molecules may inhibit binding better.

In the above mentioned example of steric hindrances of Trastuzumab and Pertuzumab IgM to Her2, Pertuzumab IgM inhibited cell proliferation better than Pertuzumab IgG1. This was interesting because Pertuzumab IgM being larger/multimeric may have prevented Her2 homodimerization/activation, necessary for its oncogenic effects better than its IgG1 counterpart based on the stoichiometry of binding sites. This is especially given that the experiments were performed in vitro without effector immune cell effects and that unlike the Trastuzumab binding site, the Pertuzumab binding site directly inhibits homodimerization.

Thus, the selection of epitopes directly at the dimerization sites or near them while allowing direct therapeutic inhibition as opposed to simply binding the antigen for just immune tagging for effector cells, can be coupled by the use of multimeric or conjugated antibodies for better direct effects (Figure 4A). And this is relevant not only in oncology where over-expressed biomarkers lead to proliferation, but also in infectious diseases where most therapeutics aim to neutralize virus attachment and entry and now more recently, narrowing to possible superantigenic areas in the spike. With further understanding of viral spike proteins, epitopes that facilitate direct inhibition of potentially deadly viral pathological effects ought to be considered. With recent evidence showing the SARS-CoV2 spike to also possess superantigenic characteristics that can cause hyperinflammation leading to multisystem inflammatory syndrome in children (MIS-C) and cytokine storms in adult COVID-19, such
regions can be blocked through epitope selection of the region or mitigating the clinical morbidity. Such an approach may provide the needed paradigm shift in treating viral infections that have superantigenic root pathogenesis. Although antibodies may already bind to such superantigenic motifs sites, further in-depth consideration of such recognition sites will facilitate epitope selection for direct viral pathogenicity blocking in addition to that of attachment and entry.
Figure 4. Choosing the epitopes for their potential functionality to (A) directly inhibit dimerization in the case of Pertuzumab IgM inhibiting Her2 dimerization better than its IgG counterpart due to its larger size. In such cases, an antibody fragment such as a Fab alone may not necessarily provide sufficient steric hindrance to block dimerization. In the event where direct blocking may not be suitable, the targeting of (B) allosteric epitopes to influence active or functional sites can also be explored to increase target areas in the same target antigen.
Evident from oncology and HIV virology requiring multi-pronged interventions, where in the former combinatorial synergistic use of Trastuzumab and Pertuzumab showed promise[^47] and that combinatorial antiretroviral therapy (ART) remains the dominant intervention for HIV[^104] in the latter, there are additional factors in epitope selection.

To add to additional targetable sites even within the same antigen and in the absence of ideal epitopes for direct intervention, allosteric epitopes can be both an additional and alternative solution. Evident in the use of NNRTI and recent evidence of non-enzymatic subunit being potentially druggable in HIV[^105], distal epitopes can be used to influence the activate of the enzyme/target’s active sites by non-competitive antibodies (**Figure 4B**). While allosteric sites may not necessary elicit obvious shape changes[^106] thus often being modulators rather than activators/inhibitors (see examples: IgE[^76], FcεRI[^107], IgA[^59], microbial targets[^108]), dampening of oncology or virology effects may sometimes provide the necessary gain needed for the immune system. Furthermore, targeting allosteric sites in microbial and viral proteins could drive function-crippling mutations through drug resistance mechanisms at the drug binding site. Evidence of allosteric sites are present in the HIV Gag protein where non-cleavage site mutations contributed to drug resistance associated protein fitness compensation[^109,110]. These sites could be epitopes for intrabodies that in this example, either prevent drug-resistance or reduce protein fitness Such epitopes can be leveraged upon to dissociate already bound complexes or simply to increase dissociation for a modulation of the effects (e.g. in the case of IgE-FcERI[^107]).

With additional areas to target apart from active/functional sites, it may be easier to find common allosteric sites to allow development of broad-spectrum targeting antibodies, and such widening of the search field may overcome constraints present in extracellular localization. While the allosteric analysis is currently most effectively performed computationally, it is beneficial to consider the entire protein structure/model as holistically as possible to study the
distal effects in a counterintuitive manner when raising antibodies based on the reductionist approach in experiments of using purified recombinant proteins, regions or peptides. While many new databases now include annotated posttranslational modification such as PTM-SD\textsuperscript{111,112}, the majority of crystal/NMR structures in structural databases are without post-translational modifications which can have mismatches with \textit{in vitro} expressed cell lines with varying glycosylation patterns\textsuperscript{113}.

Applications, Solutions, and Conclusion - Holistic Epitope selection

Sagacity in target and epitope can maximize the success for vaccine, therapeutics, and diagnostic development. Apart from deep understanding of the target biology, there are numerous factors from accessibility at the cellular and molecular level, the inducibility, occlusion, conservation at family and species levels, influence by distal allosteric sites, that can now be considered not only individually but together. Coupling such efforts with sagacious antibody development\textsuperscript{6} that can involve swapping of elements\textsuperscript{114}, the very expensive developmental processes can be made more cost and time efficient.

Competing Financial Interests: The authors declare no competing financial interests.

Acknowledgements: The understanding that contributed to the insights were funded by Joint Council Office, Agency for Science, Technology, and Research, Singapore under Grant number JCO1334i00050. The staff are funded by the National Research Foundation (NRF) Singapore grant to Experimental Drug Development Centre (EDDC). The authors would like to thank J. Derrick for his insights and discussion. Figures were created with BioRender.com.
References


52 Duffy, K. T., Wharton, P. J., Johnson, J. D., New, L. & Holliman, R. E. Assessment of immunoglobulin-M immunosorbent agglutination assay (ISAGA) for detecting


