Review of immunoinformatic approaches to *in-silico* B-cell epitope prediction

Rick Reitmaier (reitmr@gmail.com)

ABSTRACT

In this paper, the current state of *in*-silico, B-cell epitope prediction is discussed. Recommendations for improving some of the approaches encountered are outlined, along with the presentation of an entirely novel technique, which uses molecular mechanics for epitope classification, evaluation and prediction.

INTRODUCTION

The relatively new field of immunomics, which focuses on the interactions between a pathogen and a host, erects a bridge across the endeavors of informatics, genomics, proteomics, immunology and clinical medicine.

Genomics and proteomics have both arisen to become independent arenas for research and development whist immunomics is lagging in this effort (*De Groot 2006*).

Perhaps part of the issue is that many immunologists prefer traditional approaches to solving biological problems, believing that 'hands-on' results are somehow more significant than *in-silico* techniques.[9]

Thus one finds that the computational tools available for the genome researcher have reached a level of maturity and recognition that an immunologist can only admire.

The average immunologist is overwhelmed with a broad array of tools that are often highly specific in use, are not well understood nor characterized, have undergone fairly limited testing, and are often not publicly accessible.[9]

Despite the relative immaturity of the current suite of tools, advances in the area of vaccine design have shown some early success (*Korber et. al 2006*).

As immunoinformatics tools begin to mature and see wider adoption, there are numerous opportunities for applying these techniques; problems in protein therapeutics, autoimmunity, transplantation and allergy to name a few.

IMMUNE RESPONSE

In order to discuss the computational basis for many of the immunoinformatics tools, an appreciation of the biology involved in the immune response must first be elaborated.

The underlying goal of any immune response is to remove disease causing matter (or pathogens) from the host. The pathogens take the form of viruses, bacteria, fungi and parasites. The immune system mounts an attack on these microorganisms whenever they trigger the appropriate mechanisms.

Any entity possessing the ability to engage an immune response can be referred to as an antigen.

Many types of cells are involved in the immune system but two are of particular interest to many immunologists as they play a pivotal role in much of the immune response.

The T-cell (being derived in the thymus) acts to directly kill cells infected with viruses or activates other cells of the immune system.

The B-cell (being derived in the bone marrow) secretes antibody which is directly and indirectly used to destroy pathogens.

B-cells, Immunoglobulin and Antibodies

B-cells are able to directly bind with antigen through their B-cell receptors or are able to secrete antibodies possessing this ability.

The antigen-binding portion of the antibody is equivalent to the B-cell receptor and thus antibodies can be thought of as the secreted form of a B-cell receptor.

The immune system operates by generating a large variety of antibodies, also known as immunoglobulins, each able to recognize a different antigen.

Each B-cell generates only one form of immunoglobulin, the variety required by the immune system being created by the vast numbers of B-cells that are generated.

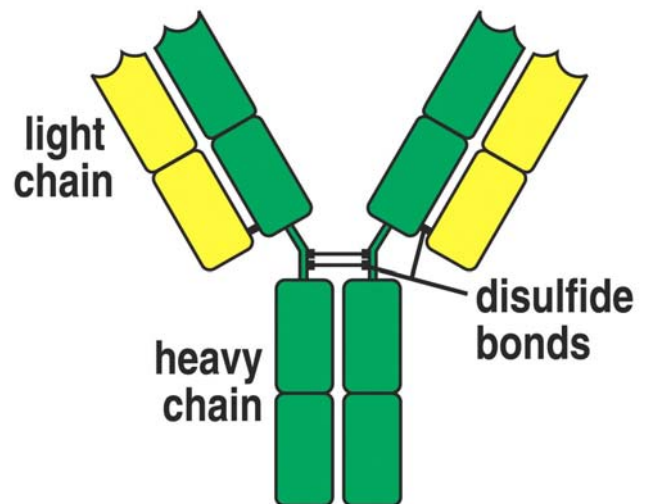Immunoglobulin molecules are roughly Y-shaped and consist of four polypeptide chains.

Figure 3-2 Immunobiology, 6/e. (© Garland Science 2005)

Figure 1 - Immunoglobulin molecules are composed of four polypeptide chains; two heavy (green) and two light (yellow) joined by disulfide bonds (*Janeway 2005*)

Due to the relative molecular weight difference between them, the polypeptide chains are named heavy and light and are attached to each other via disulfide bonds.

The tips of the upper two arms of the Y contain the amino acids that anchor the immunoglobulin to the antigen.

In order to recognize the variety of antigens encountered by the host, the tips of the arms are not fixed in composition but instead are highly variable, allowing each immunoglobulin molecule to be specific for a given antigen.

It is generally acknowledged that antigen binding occurs in a variety of ways and thus leads to a differing arrangement of antibody conformations.
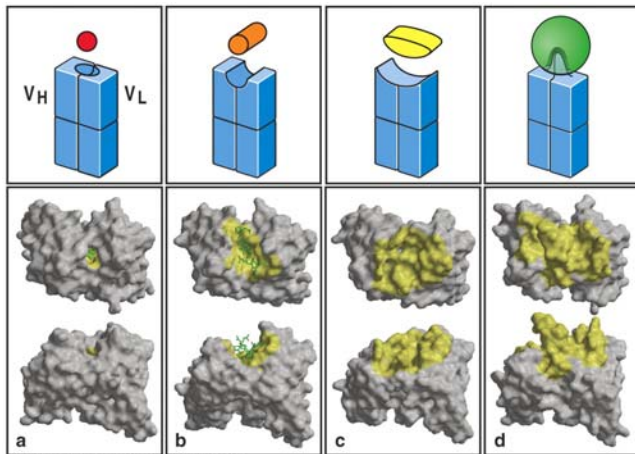

Figure 3-8 Immunobiology, 6/e. (© Garland Science 2005)

Figure 2 - Antigens can bind in pockets, grooves or on extended surfaces with antibodies. (*Janeway 2005*)

Exploring these variable domains further, it can be seen that there are specific positions in the peptide chain in which variability is highest. There are twelve such regions of hypervariability per immunoglobulin molecule.

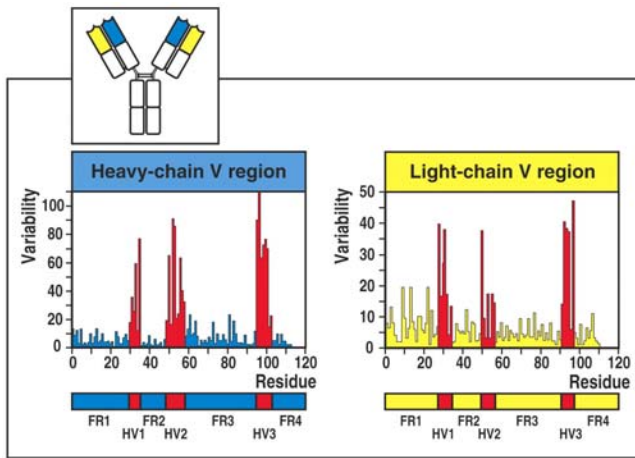As each arm of an immunoglobulin molecule is identical, there are actuality only six unique hypervariable units.

Figure 3-6 Immunobiology, 6/e. (© Garland Science 2005)

Figure 3 - Discrete hypervariable (HV) regions in both light and heavy chains (*Janeway 2005*)

Each of the hypervariable regions is surrounded by less variable stretches, which are structurally β-sheets.

Each of the chains contains three of the six loops that form the binding region. The hypervariable portions of the loops on the light chain are designated L1, L2 and L3 whilst those on the heavy chain are H1, H2, and H3.

These regions are also known as complementary determining regions (CDRs) and this term is useful when discussing the general mechanism of binding to antigen or the means by which these regions are created by the host cells.

The CDRs of antibodies, as noted earlier, are derived from the combination of two chains. This mechanism produces variability in relation to the binding site, seeing as combining different variants of the chains results in different sets of CDRs.

Despite the large number of permutations that are afforded by this technique, the immune system has yet a more robust means of generating dramatically larger numbers of unique CDRs.

The CDRs are generated not by a single or multiple genes, but are instead built from a repertoire of gene segments which are rearranged in a random way thus generating great diversity.

This process alone accounts for roughly 10^6 different specificities (*Janeway 2005*), but in addition to this combinatorial diversity we need to overlay something called junctional diversity in which the recombination of the gene segments creates unique nucleotides along the splicing joints.

In fact, H3 and L3 are significantly affected by this process since both of these regions are coded by a gene segment that falls directly on one of the junctions. This property is manifest in figure 3, where greater variability is visible within these units.

When a B-cell responds to an antigen, a process called somatic hypermutation commences in which point mutations are introduced to CDRs in an attempt to generate a higher affinity binding to the antigen.

This process is subject to positive and negative selection pressures, and ultimately immunoglobulins that have a higher binding affinity to the antigen will dominate. Possibly resulting in molecules that have even higher affinities then that associated with the original immunoglobulin that was involved in activating the B-cell.

T-Cells and MHC

T-cells are also involved in the recognition of antigen but do not directly attach to antigens.

Instead T-cells recognize peptide fragments that have been internalized by other host cells via endocytosis, chopped into pieces, and subsequently exported to the cell surface.

Besides obtaining peptide fragments from extra-cellular space, host cells are also able to present intercellular matter, such as viruses and bacteria, to the T-cells.

The peptide pieces that are presented (also known as epitopes) are on the order of 10 amino acids and consist of a contiguous fragment of the antigen protein.

The complex that is responsible for the presentation on the cell surface is called major histocompatibility matrix (MHC), the

genes for which are continuously active in almost every cell of the human body.

Just as in the case of immunoglobulins, MHC molecules require a highly variable binding arrangement in order to accommodate the great variety of peptide fragments that may be encountered.

While not as uniquely prolific as immunoglobulin generation, the MHC genes allow for a great variety of MHC molecules to be constructed.

As noted earlier MHC molecules are brought to the cell surface where T-cells interact with the peptide fragment contained within the complex.

This configuration of MHC with T-cell allows for much less structural diversity than is witnessed in B-cell antigen binding.

Additionally, the requirement that peptide fragments be held and somewhat firmly positioned within the MHC molecule imparts a stronger set of structural requirements on the overall binding complex.

It is for these reasons and others that will be presented shortly, that immunoinformatics techniques applied to T-cells and MHC proteins have advanced to a greater degree than those dealing with B-cells.

EPITOPE PREDICTION

One of the principal goals of informatics research in immunology is the development of algorithms to assist in the creation of new vaccines. Reliable epitope identification via computational means would lessen the burden required for laboratory analysis of viral, bacterial and parasitic gene products.

An informatics driven approach would allow the immunologist to greatly reduce the experimental work, providing a valuable starting point for the exploration of potential binding sites.

In examining the nature of T-cell binding, it can be seen that the amino acids presented to T-cells consist of short linear peptide segments that have been cleaved from antigenic proteins. Because of the constraint that the epitopes must be linear along the polypeptide chain, many sequence based algorithms are highly applicable to T-cell investigation.

The case is markedly different for B-cells, wherein the amino acids involved in the antibody-antigen binding may be very far away from each other in the polypeptide chain, but due to the folding of the protein they may be physically very close. This critical dependence on the three-dimensional structure of the antigen makes informatics approaches to B-cell immunology problems much more difficult than those dealing with T-cells.

In fact, advances surrounding T-cell epitope identification have already been integrated into the field of drug discovery.[9] Unfortunately, the same cannot be said for informatics techniques as applied to B-cell epitope identification.

The remainder of this paper presents the computational methods currently employed in B-cell epitope prediction. From this point forward we are dealing strictly with B-cell epitope prediction methods, so all uses of the term epitope refer only to B-cell epitope, not T-cell. Likewise, for any other ambiguity encountered.

But prior to diving into the informatics side of the field, a stronger appreciation of the biological basis surrounding the antigen-antibody binding complex will prove fruitful during the subsequent analysis of the algorithms.

A discussion of the CDR regions and associated antigenic sites as it relates to B-cell epitope prediction is first presented below.

Antigen-Antibody Binding Complex

An extremely detailed analysis of the 26 antigen-antibody complexes available at that time was performed by *Macallum et. al 1996*. Their results highlighted the fact that the binding of antigen and antibody was much more complex than previously thought.

The authors defined distinct classes of antibody-antigen for different types of molecules. Haptens formed one group (small); peptides, carbohydrates and nucleic acids another (medium), and finally proteins in the last (large).

Large antigens were found to contact residues at the extremities of the combining site along with the apical loop residues.

Additionally and most importantly they proposed a "contact definition" for the interaction and in doing so redefined which amino acids make up the CDR positions, generally choosing to increase the length of the CDRs based on their analysis of the antibody-antigen interactions.

More recently, it has been demonstrated (*Collis et. al. 2003)* that the mixture of lengths of the CDRs is the determining factor for the nature of the combination site. Specifically, L1, L3, H2 and H3 play a pivotal role is this mechanism. For example long L1, L2, H and H2 with a short L3 and H3 would result in a groove. A long L1, H2 with a medium to long H3 and a short L3 would result in a pocket. Likewise a very long H3 could be the result of a protuberance from the surface.
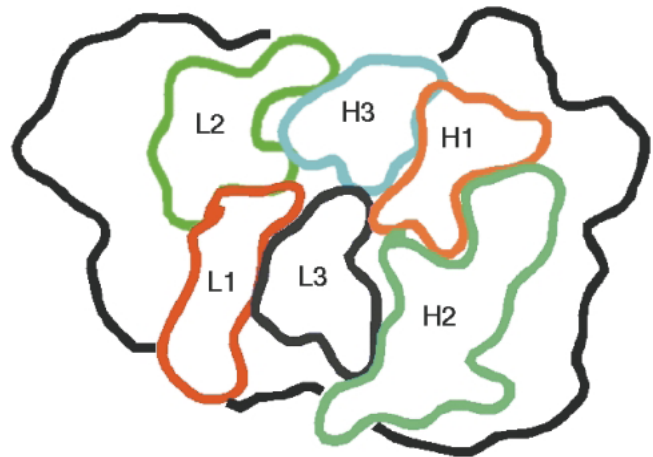

Figure 4 - A representation of the antigen combining site and the associated CDR positions (*Collis et. el. 2003*).

It was also found that some of CDRs were more likely to vary than others in length. The authors found that for the 1788 antibodies studied, the lengths of L1, L3, H2 and H3 appeared most affected by the topography of the combining site.

An analysis of sequence composition revealed that "in all cases the distributions of amino acid frequencies in antibody combining sites were totally different from that of normal loops."

The above result has significant impact on the ability to predict epitopes, as one cannot make hypotheses regarding the combining site based on the distributions found in unbound antigen-antibody complexes. That is, the proteins should only be examined when they are bound together.

Additionally, as is pointed out by v*an Regenmortel (2006)*, some of antibody-antigen interactions that are being used for study may in fact be artifacts of the process in which they were obtained.

The assay in which the antigen was prepared might contain denatured molecules and in doing so create conditions for a binding site that would not be available *in-vivo*. This has been shown in the unique forms of binding that occur when an antibody recognizes a peptide in free solution as opposed to conjugation with a carrier or absorbed as a solid phase.

Further exploration of 59 antigen-antibody interactions was performed by *Almagro 2004* with the intention of identifying and characterizing the CDR residues that directly interact with the antigen; the so called specificity determining residues (SDRs).

The observations were mostly in agreement with that of *Macallum et. al., 1996*, with only minor differences in the locations of the CDR regions.

Now that the biological basis surrounding B-cell antigen-antibody biding has been elucidated, we can turn our attention to the *in-silico* prediction techniques which have accompanied these discoveries. The techniques are presented in chronological order so that an appreciation of the advancement in the field can be obtained.

B-CELL EPITOPE PREDICTION

Immunologists, inspired by the success of informatics as it was applied to the human genome, turned to similar algorithms in order to solve the epitope prediction problem.

One of the early and most influential papers on epitope prediction (*Hopp and Woods 1981*) devised a method based on the observation that charged hydrophilic amino acids are common features of antigen determinants. They used a combination of *Chou and Fastman's 1974* technique of averaging values across a polypeptide chain with the solvent parameters, obtained from *Levitt 1976*.

Curiously, they adjusted the values obtained from Levitt in order to build a model that had a higher success rate. But of course in doing so, they then lacked a control set of data over which to verify the validity of the model.

The conclusion of the authors was a 100% success rate out of the 12 proteins that formed the data set. It would have been most interesting to build the model on six of the proteins and then attempt to predict the other six, but unfortunately it would take another ten years for this to happen.

A full review of epitope prediction as was current for the time was presented by *Pellequer et al. 1991*. Many of the propensity scales in use at the time were compared, Hopp included, against 9 extensively studied epitopes from 11 proteins. The results demonstrated a very low success rate using any of the techniques described in the paper. It was also pointed out that many of the techniques were widely popular at the time and that numerous authors reported high success rates. It is most likely that these high reported success rates are possibly attributed to small

datasets and a prevalence of skewed testing methodologies, as will be described later.

Interesting, though, the most striking comment from the authors apparently would not be fully appreciated for yet another 10 years.

> "These globally mediocre results reflect the fact that no single scale contains enough information to allow a transformation from primary structure data to a tertiary structure entity (the epitope). Clearly, available prediction methods based on unidirectional analysis do not cope satisfactorily with the three-dimensional reality of antigenic sites."

Despite the authors compelling arguments which included data from other papers showing a 52-61% success rate and noting that that at the time only 26 complexes were available for study, the so called one-dimensional techniques would continue to be explored for many years.

The debate of epitope prediction using one dimensional sequence analysis (as termed by *van Regenmortel 1996*) continued with even more fervor as *van Regenmortel* implied that even a three dimensional model would be insufficient for correct prediction as the antigen-antibody interaction is a dynamic event with time, the fouth-dimension, entering the equation.

A terse two page article appearing in Protein Science (*Blythe & Flower 2005*) provided extremely compelling evidence on the futility of using amino acid scales to derive epitope predictions. The authors examined 50 proteins using 484 unique amino acid scales with 17,100 combinations of algorithm parameters. After choosing the 150 most likely candidates a further 228,900 parameter combinations were performed. They concluded that there was no significant correlation between the sequence profiles and the location of known epitopes. Clearly, the technique of epitope prediction must include more than just statistical data about polypeptide strands.

It may be co-incidence, but it appears that after the publication of this paper there has been an expansion in the number of novel techniques being applied to the problem of epitope prediction.

The CEP prediction server (*Kulkarni-Kale et. al. 2005*) introduced the concept of "accessibility of residues" based on a metric computed using the three dimension nature of the antigen. This approach was reported to achieve a 75% correct prediction rate, based on 21 complexes. The authors acknowledge in their paper that even approximate prediction is useful in helping to design experiments and that their technique locates probable antibody-antigen binding sites.

More recently, the CEP server (accessed March 2007; web site last updated August 2006), reports 52 complexes showing 39 correctly predicted. Wherein a correct prediction is yielded if >60% of the residues identified match the known binding sites.

The epitope mapping tool (EMT) described by *Batori et. al. 2006*, was constructed around a database of antibody motifs. These 1502 5-10 segment long amino acid sequences were classified into general motifs by sequence alignment. The authors do not specify what algorithm was used to do the alignment nor the criteria of what constitutes a motif. A single antigen was presented against the model and the authors reported that there is favorable evidence of the existence of epitope motifs.

It would be interesting to see how the hypothesis fairs under the stress of more than a single antigen.

Machine-learning techniques are also just now being applied to this problem, apparently in hopes that one of them will provide positive results.

An artificial neural network (*Saha & Raghava, 2006*) appears to have produced slightly better results than the 'classical' propensity scales approach. An accuracy rate of 66% was achieved by the authors which is somewhat marginally better than prior prediction attempts. Besides being the first application of a neural net to this problem domain, what is notable about this study is the testing methodology. A dataset consisting of an equal number of known epitopes and non-epitopes were chosen. This appears to be the first instance with B-cell epitope predictors in which know negatives have been inserted into the dataset. Additionally, the authors randomly chose three-fifths of the set for training; leaving the remaining two-fifths for validation and testing. This process was then repeated five times and the final results were obtained from an averaged of the testing sets. This exemplary use of blind and randomized testing will hopefully become the norm for the validation of epitope predictors.

The use of hidden Markov models (HMMs) was also recently explored by *Larsen et. al. 2006*. A combination of propensity scales and HMM trained on epitopes, reported to result in slightly better performance than previous methods.

Undeterred by *Blythe and Flower's* compelling paper, a subsequent effort by *Anderson et. al. 2006,* chose to combine propensity scale metrics with the effects of spatial proximity and surface exposure. Choosing to incorporate some amount of three dimensional criteria in their scale was wise, but apparently insufficient to produce any significant advantage when compared against earlier techniques.

Epitope Interaction Databases

One of the issues with epitope prediction is that there is no consensus on what interactions are antigenic. Thus comparing techniques is difficult, even those that work against the same protein (*Schlessinger et. al. 2006*).

The construction of an antigen-antibody database is one means to alleviate this issue, as there would be a consistent definition of what constitutes a binding complex for a given antigen. Epitome, Bcipep and AntiJen are examples of such efforts.

The goal of the Epitome (*Schlessinger et. al. 2006*) database is to categorize all known antigen-antibody complexes in the PDB. A semi-automatic process produces a detailed description of the CDRs and the antigenic residues. One interesting note about this approach is that the authors choose to use a structural alignment tool (SKA) to perform the initial step of locating the CDRs. Curiously, the authors noted that usage of sequence similarity is problematic when applied to CDR detection, yet they choose to use a structure alignment tool as part of the first phase of antibody-antigen analysis.

CDRs are highly variable by definition and structural alignment programs work by locating regions of similarity, one wonders about the efficacy of this approach. It is quite possible that using this approach all interaction sites were located with the currently know complexes, but as the authors are intent on automating the entire process, it is quite conceivable that a new form of antigen

variant will be encountered. This approach would then possibly fail to locate this new pattern.

An alternative technique would be to utilize the fact that immunoglobulin is involved in all these interactions in only a limited region of the entire protein. One could use this biological fact to narrow the search down to only the loops of the antibody. The PDB structures could then be directly interrogated to identify which amino acids are in reasonable proximity to the antigen, thus determining which peptides are likely involved in the complex.

The benefit of this approach is that the full resolution of the three dimensional relationship between antigen and antibody is utilized to determine the combining site. Additionally, as new and possibly unique antigen-antibody complexes are encountered, this method would remain viable.

In any event, it is quite disappointing that the implementation abandons the information rich, three-dimensional data contained within the PDB so early during the processing.

The Antijen database (*Blythe et. al. 2002*) is the most extensive epitope database with over 24,000 entries. Although, it contains peptide binding data for a number of interactions including B-cell epitopes, the main focus of the resource is linear epitopes.

Bcipep (*Saha et. al. 2005*) contains B-cell epitope data derived from 1216 antigens described in PDB. This database also exclusively focuses on linear epitopes.

The Immune Epitope Database (IEDB) contains a highly annotated set of B-cell epitopes that have been curated from PDB antigen-antibody complexes. Currently an effort is underway to provide a non-redundant set of data that is tailored for use with epitope prediction tools (*Sette, A. unpublished work*). These datasets should be available by mid-2007.

All of these efforts will hopefully lay the groundwork for a unified, consistent and controlled series of datasets that can be utilized for building, testing and comparing predictors.

NEW DIRECTIONS

The influence of antibody structure on function cannot be understated in the context of antigen-antibody interaction.

Epitopes are highly context dependent and cannot exist without a corresponding antibody. At the same time, the presence of somatic hypermutation results in variations at the binding site. It is therefore more apropos to regard the antigenic surface as a landscape of epitopic regions without well defined borders (*Greenbaum et. al. 2006*).

Given the right set of circumstances any region on the surface can act as a binding site and therefore an enormous number of epitopes exist for a given antigen. The difficulty with this definition is that tools that choose to divide regions into epitopes and non-epitopes may therefore not reflect the biological reality of the situation. And in doing so, are then prone to produce inaccurate and incomplete epitope predictions.

For these reasons an epitope prediction algorithm must be based on the structure of the antigen first and foremost. The particular amino acids at a given position on the polypeptide chain of an antigen are somewhat immaterial when first considering a binding complex. Instead they should be viewed as of a consequence of the three dimensional activity of a docking event. Characterized in these terms, it is clear that the structure of the antigen is

paramount and should be considered the foundation of any prediction tool.

So far, B-cell epitope related studies of antigen-antibody binding sites have focused on amino acid sequence, gross patterns of antigen surface topology or some combination of the two. All attempts, thus far, to use the information derived from these studies has resulted in very poor epitope predictors.

It appears that in attempts to reduce the complexity of the antigen-antibody interaction, vital information is lost. Thus, it may be worthwhile to tackle the problem in a completely different manner.

One possibility is to explore the antigen-antibody interaction as a fully three dimensional physical entity similar in fashion to how an antibody ultimately binds with an antigen. Calculating the atomic forces that come into play within the antigen-antibody binding site and then performing a statistical analysis across a number of known antigen-antibody complexes, might reveal some hitherto unknown aspect of the interaction.

Using this approach one could imagine a catalog (similar to motifs) of molecular force patterns that signify a strong correlation with a potential epitope; the catalog, being built from known antigen-antibody complexes. Upon presentation of a new antigen the force fields would be computed for the entire protein and then the catalog searched for a similar field configuration.

The next step of presenting an antibody complex could also be computed. As the antigen site is located, a number of potential CDR configurations and amino acid sequences could be built up using charge information from the potential binding locations. Each of these antibodies could then be placed in a molecular dynamics simulation with only the relevant portions of the molecule involved being part of the computation. The binding state (or lack thereof) could then be modeled to high degree of accuracy. This approach would also allow for the proteins to be placed in solute so that the interactions of the surrounding medium could also be taken into account.

Although solving a related but somewhat different problem, *Cachau et. al. 2003*, utilized molecular modeling techniques to map peptide sequences to three dimensional structures. So, the use of physical models has some grounding and is not completely foreign for this endeavor.

The use of molecular modeling techniques as applied to this problem domain may seem like a radical approach and quite possibly there is little expertise in the immunoinformatics arena with this technology, but in order to make any advancement in epitope prediction some fundamentally new idea is required. The techniques so far employed have achieved little advancement in this area over the last 15 years, so it is clear that the assumptions upon which these models are built are flawed and something entirely new needs to be tried.

CONCLUSION

I have presented the current state of immunomics techniques applied in the B-cell epitope discovery problem and have provided the underlying biological framework in which to evaluate the effectiveness of each of these approaches.

While the field of T-cell epitope prediction has made some significant advances, unfortunately despite much effort and activity the same cannot be said of B-cell prediction.[9]

This is most likely due to the fact that many of the ubiquitously available genomics tools / algorithms are applicable for T-cell – MHC binding, but that interactions associated with the possibly seemly related antigen-antibody binding problem are significantly more complex.

Ample evidence was provided that these two types of binding are indeed distinct and that similar algorithmic approaches cannot be utilized.

A critique of the current techniques being applied to the problem of B-cell epitope prediction was also presented. Notes for potential improvement opportunities were set forth, as the discussion warranted.

Finally, a radically new approach of B-cell epitope characterization and prediction, based upon molecular mechanics, was elucidated. This technique fully takes account the three-dimensional molecular nature of an antigen-antibody interaction.

REFERENCES
[1] Almagro, J. C. (2004) "*Identification of differences in the specificity-determining residues of antibodies that recognize antigens of different size*" J. Mol. Rec. 17, 132-143

[2] Anderson, P. H., Nielsen, M. & Lund, O. (2006) "*Prediction of residues in discontinuous B-cell epitopes using protein 3D structure*" Protein Science 15:2558-2567

[3] Batori, V. B., Friis, E. P., Nielsen, H. & Roggen, E. L. (2006) "*An in silico method using an epitope motif database for predicating the location of antigenic determinants on proteins in a structured context*" J. Mol. Rec. 19, 21-29

[4] Blythe, M. J. & Flower, D. R. (2005) "*Benchmarking B cell epitope prediction*" Protein Science 14, 246-248

[5] Blythe M. J., Doytchinova I. A., Flower D. R. (2002) "*JenPep: a database of quantitative functional peptide data for immunology."* Bioinformatics 2002 18 434-439

[6] Cachau, R.E., Gonzalez-Sapienza, G., Burt S. K. & Ventura, O. N. (2003) "*A new addition to the structural bioinformatics toolbox*" Cell Mol. Biol. 49(6), 973-983

[7] Chou, P. Y. & Fasman, G. D. (1974) Biochemistry 13, 222-224.

[8] Collis, A. V. J., Brouwer, A. P. & Martin, A. C. R., (2003) "*Analysis of the Antigen Combining Site*" J. Mol. Biol. 325, 337-354.

[9] De Groot, A. S. (2006) "*Immunomics: discovering new targets for vaccines and therapeutics*" DDT 11, 5/6 203-209

[10] Greenbaum, J. A. et. al. (2007) "*Towards a consensus on datasets and evaluation metrics for developing B-cell epitope prediction tools*" J. Mol. Rec. (in press)

[11] Hopp, T. P. & Woods K. R. (1981) "*Prediction of protein antigenic determinants from amino acid sequences*" PNAS 78.6, 3824-3828.

[12] Janeway, C. A., et al (2005). *Immunobiology*, 6th ed., Garland Science.

[13] Korber, B., LaBute, M. & Yusim, K. (2006) "*Immunoinformatics Comes of Age*" PLOS Comp. Bio. 2:6 484-492

[14] Kulkarni-Kale, U. Bhosle, S. & Kolaskar, A. S. "*CEP: a conformational epitope prediction server*" Nucl. Acid Res., 33, W168-W171

[15] Larson, J. E. P., Lund, O. & Nielsen, M. (2006) "Improved method for predicting linear B-cell epitopes" Imm. Res. 2:2

[16] Levitt, M. (1976) J. Mol Biol 104, 59-107.

[17] Macallum R. M., Martin A. C. R., & Thornton J. M. (1996) "*Antibody-antigen Interactions: Contact Analysis and Binding Site Topography*" J. Mol Biol. 262, 732-745.

[18] Pellequer, J. L., Westhof E. & Van Regenmortel M. H. V. (1991) "*Predicting Location of Continuous Epitopes in Proteins from Their Primary Structures*" M. in Enzym. 203.8, 176-201.

[19] Saha, S., Bhasin, M. & Raghava, G. (2005) "*Bcipep: A database of B-cell epitopes*" BMC Genomics 6:79

[20] Saha, S. & Raghava, G. P. S. (2006) "*Prediction of Continuous B-Cell Epitopes in an Antigen Using Recurrent Neural Network*" Proteins 65,40-48

[21] Schlessinger, A., Ofran, Y., Yachdav, G. & Rost, B. (2006) "*Epitome: database of structure-inferred antigenic epitopes*" Nucl. Acid. Res. 34, D777-D780

[22] Van Regenmortel, M. H. V. (1996) "*Mapping Epitope Structure And Activity*" M. in Enzym. 9, 465-472.

[23] Van Regenmortel M. H. V. (2006) "*Immunoinformatics may lead to a reappraisal of the nature of the B cell epitopes and of the feasibility of synthetic peptide vaccines*" J Mol. Rec. 19, 183-187.