

## **1.- PEPTCELL AS A COMPANY - WHO WE ARE**

PepTcell is a young biotechnology company which commenced operations in April 2004 in order to capitalise on several new scientific concepts established by its founder. The aim of the company is, therefore, to develop and commercialise these concepts into novel analytical tools and clinical products in the fields of autoimmune and infectious diseases as well as organ transplantation.

Since starting its operations PepTcell has been engaged in developing and completing the code for its new T cell epitope prediction software, the filing of patents to protect its intellectual property, establishing collaborations with several research groups in order to independently test several of its candidate drug products and successfully raising capital from a range of investors.

## **2.- EPITOPES AND THE IMMUNE RESPONSE - THE PROBLEM**

After more than a century as a purely experimental science, Immunology has reached a turning point in which it must re-address its scientific procedures in order to capitalize on the wealth of data available in this post-genomic age. Current experimental procedures, despite the incredible advances witnessed in the last two decades, are ill placed to cope with the volume of data available and have clearly failed in delivering faster and more efficient methods of identifying candidate agents for vaccine development. The cost of failing to meet this target is evident to all humanity in the current reality of old diseases with increasing dominance (e.g. Malaria, Tuberculosis, HIV-AIDS, etc.), and the future threat of potentially devastating pandemics caused by novel pathogens (e.g. SARS, Influenza H5, etc.).

Adaptive immune responses, and by definition vaccine development, has at its core the notion of epitope identification. Both T and B cell responses are clonal by nature, with each individual clone of lymphocytes responding to one specific epitope. Almost any molecule can be a source of epitopes for the adaptive immune system, however it is those derived from proteins, and in particular those capable of driving T cell responses, that remain the key to the development of most successful vaccine therapies.

## **3.- IDENTIFICATION OF T CELL EPITOPES - WHAT MODELS ARE CURRENTLY AVAILABLE**

Current experimental methods for the identification of T cell epitopes, require a large amount of investment in terms of both labour and time to single out the few reactive epitopes present in amongst the over 200 peptides that an average protein can yield. If one adds to this the fact that current genome analysis tools can identify in excess of 4,000 potential target proteins in a single pathogen, the magnitude of the problem we face becomes painfully evident.

The application of computational tools to the analysis of protein sequences based on existing genetic, biochemical and immunological data, in order to identify reactive T cell epitopes, constitutes the best currently available option to break through the vaccine development bottleneck described above.

Initial attempts to predict T cell epitopes by computational means, which relied on the identification of purely biochemical “hot-spots” (e.g. hydrophilicity) along a protein sequence, have been shown to be wholly unreliable (*Deavin et al, 1996*). Their Achilles’ heel proved to be their failure to take into account the complex processes leading to the induction of an epitope-specific T cell response.

Broadly speaking, the induction of T cell responses is a sequential process that involves the following three steps:

1. Degradation and processing of the protein antigen into peptides of different lengths,
2. Binding of the resulting peptides to a MHC molecule,
3. Recognition by a T cell of the peptide bound to the MHC molecule as foreign (i.e. non-self) and subsequent activation of the T-cell response.

A whole spectrum of predictive models has been developed by academic groups and commercial organisations by concentrating their efforts in only one of these three steps. None of these models, as it will be described below, has been proven effective in rapidly and reliably identifying reactive T cell epitopes.

### **3.1. - Prediction models for peptides created by degradation of antigenic proteins.**

Several non-commercial statistical models have been developed to predict peptides produced by degradation and processing of antigenic proteins. The accuracy of these models is variable (*Saxová et al, 2003*) and hence a fraction of the peptides produced are identified. Moreover, of the peptides identified, there is no indication of what percentage are able to bind a MHC molecule, let alone induce a T cell response. This situation arises from two factors.

1. Antigen degradation involves several enzymes and pathways, some of which are yet to be identified.
2. The process of antigen degradation is by itself probabilistic in nature (i.e. not all copies of a protein will be degraded in the same manner).

In view of this level of processing uncertainty, it is widely acknowledged that any prediction model applied at this level of the process can only, at best, provide some indication of the stability of candidate peptide epitopes identified by other means.

### **3.2. - Prediction models for MHC binding of epitope peptides - Competitors**

Peptide binding to the MHC molecule is seen by many as the key step in the induction of an epitope-specific T cell response. As a result, a large number of models, both commercial and non-commercial have been developed. These models use a broad range of analytical methods (e.g. neural networks, hidden markov models, support vector machines, etc) and are based either on available peptide binding or structural data. Despite their popularity and competing accuracy claims, the reality is that the outcome of all these algorithms is characterised by the same fundamental flaw:

- Binding to the MHC is not the only key requisite for T cell reactivity as clearly proven by all the tightly bound epitopes derived from self proteins and the recognised escape variants of many pathogens that never elicit T cell reactivity.

Indeed, our competitors acknowledge in their websites that peptides identified by their algorithms must still be extensively tested experimentally in order to establish their individual T-cell reactivity. As a result, these models, at their best, fail to provide the promised significant labour and time economies compared to traditional purely experimental protocols. At their worst, they fail to identify both the binding and the reactive T cell epitopes in a protein, as shown in one of our competitors peer reviewed publications (*Brander et al, 2001*). In this manuscript, the algorithm used identified 8 binding peptides in Kaposin, a Human Herpesvirus 8 protein. Amongst those, only 3 passed their experimental binding test with only one being later experimentally identified as reactive. However, after further experimental analysis of the identified peptide, Brander and

colleagues state that “unexpectedly, the data revealed that the best recognized epitope is a 9-mer peptide (...) which was not predicted by the algorithm”.

### **3.3. - Prediction models for T cell recognition of, and activation by, epitope peptides - Competitors**

In recognition of the intrinsic flaw of epitope prediction models based uniquely on binding data, non-commercial algorithms have been developed that attempt to predict directly the reactivity of T cell epitopes. These models use analytical methods based on amphipathicity and hydrophobicity data from the protein sequence. However, similarly to the binding only based models they attempted to supersede, they all share a fundamental flaw:

- Direct prediction of reactive T cell epitopes from a protein sequence without considering the binding requirements of different MHC alleles results in the identification of peptides unable to bind to the MHC.

## **4. - WHAT IS PEPTCELL’S MODEL ALL ABOUT?**

The goal of computational T cell epitope prediction has remained up to now the identification of reactive peptides within a protein sequence without the need to resort to cost and labour intensive experimental methods. This aim is somewhat restricted by the probabilistic and unpredictable nature of the initial step in the induction of a specific T cell response (i.e. degradation of the antigenic protein into peptides of different lengths). In an attempt to bridge this problem, several commercial and non-commercial algorithms have been developed that claim to be able to identify a group of peptides from a protein sequence that, if produced by the natural process of antigen degradation in the cell, are reactive T cell epitopes. However, as discussed above, some of these algorithms limit their search to identifying peptides capable of binding to the MHC, under the false assumption that strong binders are also inherently strong reactors. Others aim at identifying reactive peptides under the equally false assumption that strong reactors will also be strong binders.

PepTcell believes in a different approach to the challenge of reactive T cell epitope prediction. The company recognises that the nature of the binding of a peptide to an MHC, and by implication that of a TCR to the MHC plus peptide complex, is identical in nature to those of other molecular interactions such as antibody to antigen, enzyme to substrate or drug to receptor. In all these cases the activity and stability of the complex, together with the cellular or chemical outcome of the interaction depends entirely on the structural organisation of the contributing parties. PepTcell has applied this principle to analyse the event at the core of immune reactivity at the T cell level (i.e. discrimination of self vs foreign via a short peptide sequence). Using cellular, biochemical, peptide binding and crystallography data collected over the last 30 years by research groups worldwide, the company has established the different levels of interaction between the MHC, peptide and TCR that control the discriminatory process. The data associated to each of these levels is analysed through several interconnected statistical models in order to establish:

1. A list of peptides that, if produced by the natural process of antigen degradation in the cell, are capable of binding to specific MHC alleles.
2. Identify amongst the binding peptides those that will be recognised as foreign by a T cell and hence elicit the activation of a T cell response.

Sample outputs obtained from the algorithm for both these steps and using a fictitious protein and HLA allele are shown in Figures 1 and 2.

PepTcell's algorithm has been extensively tested and cross-validated against available data and is equally applicable to MHC Class I and class II alleles. In all tests carried out it has proven superior to all its commercial and non-commercial competitors. For example, when applied to Human Herpesvirus 8 Kaposin protein PepTcell's algorithm successfully identified the prime reactive peptide which was not identified by the algorithm used by Brander and colleagues (i.e. peptide LL9, *Brander et al, 2001*). Equally relevant, PepTcell's algorithm, when tested against a list of 360 peptides known to bind to HLA-A\*0201, was able of correctly identifying all 106 non-reactive and 254 reactive epitopes without the need for any experimental procedure. That corresponds to an accuracy level of over 99.98% for the overall population.

## **5.- WHAT DOES PEPTCELL'S ALGORITHM OFFER?**

PepTcell's algorithm offers, for the first time in the field of immunology, the possibility of rapidly, and accurately identifying reactive and non-reactive T cell epitopes within a protein, eliminating the need to resort to labour and cost intensive experimental methods. The resulting savings in R&D time, and hence costs, are significant. Discrimination between reactive and non-reactive peptides known to bind an MHC allele can be carried out in hours not weeks or months, whilst identification of reactive T cell epitopes from one or several proteins can be carried out in days or weeks, not months or years.

The applications of the PepTcell's algorithm are not, however, limited to the field of vaccine design. Due to its accuracy in discriminating between reactive and non-reactive peptides, the algorithm can equally successfully be applied to the development of immunotherapies for the treatment of chronic autoimmune diseases (e.g. identification of modified epitopes lacking autoimmune reactivity). Moreover, following FDA's recognition that animal models are ill suited to predict immune responses to therapeutical proteins in humans, the algorithm can also be used to identify reactive sequences within your product and hence support your IND filings, strengthen the value of your intellectual property and greatly reduce the time scales and cost of placing a new product in the market.

As a result, PepTcell's algorithm and its commitment to service quality and scientific accuracy constitute a unique opportunity for companies to improve the success rate of their drug development projects.

Should you wish to corroborate the accuracy and applications of PepTcell's algorithm we would suggest that the three most appropriate options would be for your company to either:

1. Provide us with list of peptides of which you have experimentally ascertained the reactivity without the results having been made public, or
2. Provide us with a protein from which you have experimentally identified the reactive, and perhaps non-reactive, T cell epitopes and for which these results are not publicly available, or
3. Enter a Confidentiality Agreement with PepTcell in order to carry out appropriate due diligence on our model, systems and operations.

### ***References:***

- Brander, C. et al (2001): Definition of an optimal cytotoxic T lymphocyte Epitope in the latently expressed Kaposi's Sarcoma-Associated Herpesvirus Kaposin Protein; The Journal of Infectious Diseases, 184:119-26*
- Deavin, A.J. et al. (1996): Statistical comparison of established T-cell epitope predictors against a large database of human and murine antigens; Molecular Immunology, 33:145-155*
- Saxová, P. et al (2003): Predicting proteasomal cleavage sites: a comparison of available methods; International Immunology, 15:781-787*

**Figure 1:** List of potential nonamers derived from protein X (fictitious) capable of binding to HLA allele X\*XXXX. The complete final output for a protein would include all the peptides of different lengths capable of binding to the HLA alleles predominant in the human population under consideration.

Protein Details		Downloaded Protein : MTEKGLFLPWHRLFLKMHPRITFLDHLKRVMTNPPPIVFNKAHFSGCMAMNESTYRYYTKWETFL																						
No. acids	66	9	N-mer (2-20)	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	
1	M																							
2	T																							
3	E																							
4	K																							
5	G																							
6	L																							
7	F																							
8	L																							
9	P																							
10	W																							
11	H																							
12	R																							
13	L																							
14	R																							
15	F																							
16	L																							
17	L																							
18	K																							
19	M																							
20	H																							
21	P																							
22	R																							
23	T																							
24	F																							
25	L																							
26	D																							
27	H																							
28	L																							
29	K																							
30	R																							
31	V																							
32	M																							
33	T																							
34	N																							
35	N																							
36	N																							
37	P																							
38	I																							
39	P																							
40	V																							
41	P																							
42	N																							

  

Binding Tests		Results																				
Y/N	Y/N	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Test
1	Y																					1
2	Y																					1
3	Y																					1
4	Y																					2
5	Y																					2
6	Y																					8
7	Y																					8
8	Y																					9
9	Y																					9
10	Y																					
11	Y																					
12	Y																					
13	Y																					
14	Y																					
15	Y																					

