4.1 Final publishable summary report

Executive Summary

All living organisms contain proteins and rely on them to sustain life. Proteins have many specialized purposes including structural, metabolic or enzymatic, and signalling amongst others. Some proteins are highly conserved and related from an evolutionary point of view, yet every organism also have a unique protein signature. This is forcefully exploited by our immune system to encounter and respond to foreign organisms. Hence, proteins are subject to intense research and investigation by academia and industry. Proteins consists of 20 different amino acids (building blocks) that can combine in numbers up to several thousand for large proteins. Proteins can be divided into shorter fragments called peptides, typically consisting of 15-20 amino acids. Peptides have emerged as indispensable tools for discovery of protein function and interaction. Chemical synthesis of peptides is a standard procedure, however, the costs of peptide libraries required for e.g. a screening program, is often prohibitive. Thus, pharmaceutical and biotechnology industries, as well as science, face significant economic and logistic barriers when attempting to apply large scaled peptide based screening programs in the exploitation of omics information. In the HiPAD consortium, partners have developed state-of-the-art technology within the field of label free detection of molecules and integrated it with High Density peptide microarrays, also developed by a HiPAD partner. Three label free platforms, based on three different detection principles, but all based on gold surfaces, have been developed in parallel. MALDI-imaging (iMS) is an existing technology, but here it has been adapted to peptide chips by developing a new smart-beam laser capable of acquiring 5 spot/sec and software to process and analyze very large amount of data. Although the iMS were adapted to HD peptide chips, the technology is compatible with other arraying technology (e.g. micro spotting). Surface Plasmon Resonance imaging is a known technology, where molecules in close proximity of the gold surface can be detected due to an interaction with the gold surface. In HiPAD, a new nanoplasmonic principle have been exploited to develop a new instrument (iNPx) capable of simultaneously measuring 100,000 spots (i.e. individual ligands) in real-time. The acquired data can be used to determine molecular kinetics for the molecules of interest for each of the 100,000 spots in the software developed in parallel. Also the iNPx instrument was adopted to the HD pepide microarrays, but it is also compatible with other micro arraying techniques. The third label free detection principle developed in the consortium is electrochemical detection of molecular interactions on gold pads, ultimately built into CMOS-based MOSFETs. In the project, we have demonstrated that MOSFETs can be used to detect the specific interaction between peptides and antibodies; it is possible to obtain quantitative data as well as kinetics if the MOSFTEs are associated with a microfluidics system. The HD peptide microarrays have been developed and optimized during the project; routinely HD peptide microarrays contain 220,000 individual peptides but good results have been obtained with 500,000 peptides. With the latest improvements in surface chemistry, 2,000,000 individual, and addressable, peptides can be synthesized on a single chip. This makes it possible to represent the entire human proteome on a single chip measuring 11x22 mm – the size of a thumb nail. Bioinformatics tools have been developed to allow analysis and interpretation of the very large amounts of data that can be generated. Thus the HD peptide chips have been exploited to address B cell (antibody) responses in human donors against a common virus (cytomegalo virus, CMV). Similarly, human MHC class II specificities have been addressed in details. The amount and quality of data obtained from MHC class II binding studies to HD peptide microarrays have enabled us, not only to address the specificity, but also to develop algorithms capable of predicting the binding of unknown (or not tested) peptides with great precision. These tools should be extremely valuable in the study of T cell responses, and a prerequisite for the development T cell based immunotherapies. Altogether, has the HiPAD project contributed with new state-of-the-art technology and analysis tools, which can transform the use of – omics information.

Summary description of project context and objectives

The central idea behind the HiPAD project was that peptides can be exploited in proteomics and immunomics. To study molecular interactions, it was crucial to have "eyes to see with", traditionally this is achieved through direct or indirect labelling of the molecules of interest. In HiPAD, we combine new state-of the-art high-density peptide microarrays with state-of the-art label-free detection techniques to enable yet unprecedented use of peptides in proteomics and immunomics.

The main scientific and technological objectives were to:

1) Develop high density peptide microarrays, which should be compatible with label-free detection techniques
    a. MALDI Imaging (iMS)
    b. Surface Plasmon Resonance imaging (SPRi)
    c. MOSFETs
2) Implement the developed technology and address immune specificities
3) Develop Bioinformatics tools to interpret and predict specificities in immunology and proteomics
4) Use bioinformatics to deduce e.g. immune responses to signature responses which can be represented on biomicroarrays (MOSFETs)

The project was organised in seven scientific/technological work packages aimed at developing and maturing the peptide microarrays, the label-free platforms and use the techniques to address primarily immune specificities (figure 1). A separate work package was dedicated to project management.

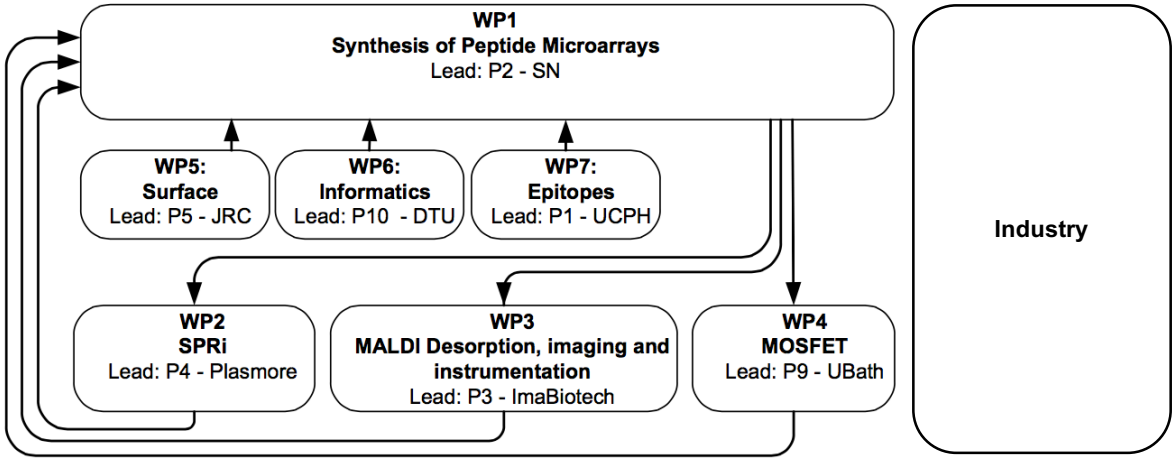


Figure 1. HiPAD work packages

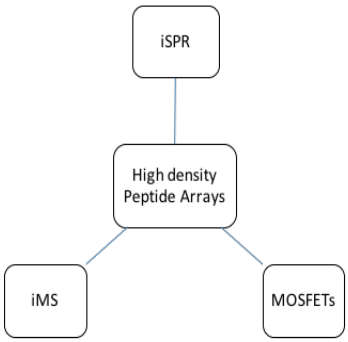


Figure 3. Three label-free platforms are developed to read HD peptide microarrays

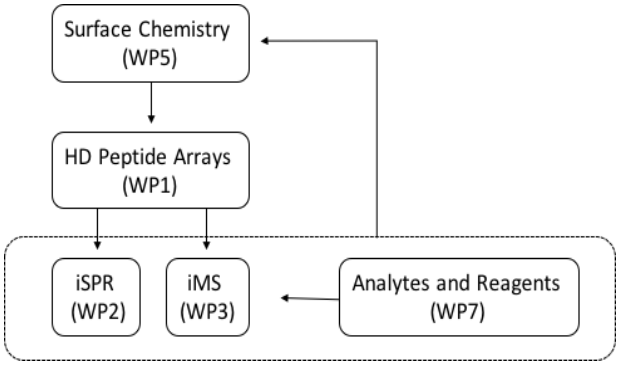


Figure 2. Iterative work flow to integrate HD peptide microarrays to SPRi and iMS

The high-density (HD) peptide microarray technology is the central platform, which should supply the peptide microarays in formats that are compatible with the label-free detection platforms (figure 2). This integration of the HD peptide microarrays with label-free detection platforms was an iterative work process where the microarray surface chemistry is being optimised (WP5) to allow solid-state peptide synthesis (WP1) subsequently being read and analysed by SPRi (WP2) and iMS (WP3) (figure 3).

Hardware development and optimisation of surfaces and chemistries lies within WP1-5 with these main objectives:

WP1:

- To implement high-density peptide microarrays to be compatible with SPRi, iMS and MOSFETs
- To develop software for the integration of design and analysis with bioinformatics data formats

WP2:

- To develop an instrument using localized Surface Plasmon Resonance Imaging (SPRi) on nanostructured surfaces for the monitoring of biomolecule interaction kinetics on peptide microarrays.

- To develop software capable of reading the HD peptide arrays and calculating the kinetics constants for each spatially pre-addressed peptide (i.e. position). The software should preferably be compatible with the IMS system.

WP3:

- To develop MALDI-iMS methods for the analysis of peptides and interacting proteins on HD peptide arrays

- To develop software platforms for the analysis of HD peptide microarrays
- To develop a smart-beam technology to increase analysis speed and integration of SPRi directed screening of HD peptide arrays

WP4:

- To develop CMOS based MOSFET arrays for the electrochemical detection of peptide-protein interactions

WP5:

- To develop advanced surfaces optimized for solid-state peptide synthesis and detection of protein-peptide interactions using the developed label-free techniques (SPRi and iMS) – this entails:
    - development large-area nanostructured gold surfaces with a uniform geometry supporting SPRi and MALDI-iMS
    - development of protocols for functionalisation of gold surfaces compatible with peptide synthesis
    - development of 3D-polymers for MALDI-iMS analysis
    - optimization of nanostructured surfaced for matrix free iMS

The HD peptide microarrays and instruments developed in the HiPAD consortium would be used by work packages 6 and 7 to address immune specificities (B cell responses, MHC specificities and subsequently T cell specificities):

WP6:

- To develop bioinformatics methods to interpret data read from HD peptide microarrays and apply these in a biologically relevant context

WP7:

- To provide analytes suitable to develop and optimise the HD peptide microarrays and instruments within the consortium.
- To use the developed technology to address immune specificities, in particular:
    - B cell responses to virus
    - MHC and T cell specificities

Prototypes of the HD peptide microarrays were developed in a FP7 supported project (acronym: PepChipOmics), however, the study of interacting proteins of interests entirely relied on direct labelling of the proteins, or on indirect labelling with specific antibodies. The development and integration of label-free detection technologies would be an enabling step towards uncovering the true potential of the peptide microarray technology in –omics projects and it might even allow the study or identification of unknown ligand interactions. The label-free detection techniques we sought to integrate with the HD peptide microarrays technology in the HiPAD project all relied on properties offered by gold surfaces (or in the case of MALDI other conductive materials). Nobel metals, Gold included, are challenging due to their very inert chemical nature. Hence, the initial work was focused on functionalisation of gold surfaces with chemistries that both preserved the physiochemical properties, upon which the label-free detection of molecular interactions depends, and allowed solid-state peptide synthesis.
The HD peptide microarray prototypes were synthesized on amino-functionalised microscope glass slides; hence the initial work was focused on developing protocols for applying nanostructured coatings to microscope glass slides (large area). The anticipated integration of SPRi and MALD-iMS techniques entailed developments of the SPRi prototypes, which preferably should allow reading full size microscope glass slides and 3D-solid phase structures or polymers suitable for both SPRi and MALD. For each platform, intrinsic requirements of the surfaces and chemistries should be honoured

in order to detect molecular interactions, or bound peptide ligands; hence the integration of HD peptide microarrays with these label-free platforms was a iterative work process involving several partners as indicated by figure 3.

The other main objectives of the HiPAD project were to use the developed platforms and instrumentation to address immune specificities. This work also involves several partners and is illustrated in figure 4.

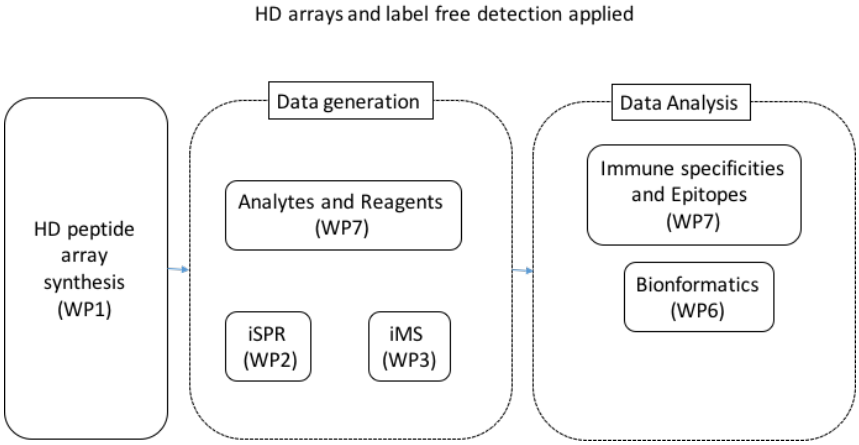


Figure 4. Work flow for generation of binding data

Protein and peptides are frequent immune targets and the immune system itself is a particularly rich source of peptide-dependent analytes. which could be used to evaluate and demonstrate the utility of a the HiPAD project. The results could readily be used to develop useful applications (bioprobes, immuno-therapy targets, vaccine candidates etc.). In the recent years it has become increasingly clear that the exploitation of our immune system has a tremendous potential to treat a number of severe diseases including cancer, virus infections and perhaps autoimmune diseases. To be able to exploit our immune system, it is of considerable importance to understand how it recognises and discriminates between self and non-self-targets (i.e. understand the specificity of the immune system).

A canonical B- or T cell target is typically a peptide consisting of 9-15 amino acids. They consist of the 20 standard amino acids. The length and sequence of the resulting polypeptides are encoded by our DNA This sequence space is enormous. To encompass all possible targets any given T or B cell can encounter, we should consider 9^{20}-15^{20} = $1x10^{19}$ – $3x10^{23}$ different peptide targets. For T cells, where the peptides are presented in association with MHC molecules, the target structures are further diversified by the extreme polymorphism of the MHC. It is clear that high capacity and high-throughput methods will be required to handle the complexity of the immune system.

WP6 developed bioinformatics tools capable of analysing the large amounts of data generated by HD peptide microarray experiments. Some of the tools are based on Artificial Neural Networks (ANNs) which is capable of explaining not only the observed molecular interactions, but also, based on copious experimental observations, to predict the outcome of similar other molecular interaction with a precision that is statistical satisfactory. Hence, bioinformatics analyses are ideally suited to extract the patterns of interactions of peptide receptors of the immune system such as antibodies and MHC.

Main Results

The majority of the results obtained within the HiPAD consortium are related to current and future exploitation by the participating SME's. Significant results have been reported during the project, but details are subject to confidentiality.

The photolithographic driven solid-state peptide synthesis exploited by Schafer-N allows synthesis of peptides on microscope glass slides theoretically containing up to 2 mill. individual peptides, but routinely used to generate 220,000 peptides in nine replicates. The photolithographic synthesis is directed by a digital mirror device (DMD) projected onto a microscope glass slide, hence the resulting peptide microarray is contained in an area of approx. 11x22 mm (figure 5).

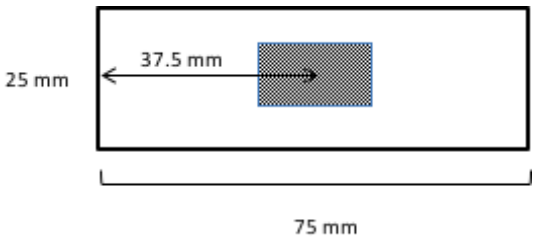


Figure 5. Relative size and location of the peptide microarray (grey hatch area) synthesized on microscope glass slides.

The spatial resolution of the MALDI-imaging (iMS) instruments theoretically allows detection of 10,000 individual peptide fields (100 x 100 µm) within the area of the peptide synthesis. The integration of HD peptide microarrays with iMS requires a process to guide the desorption laser to the area of interest and to ensure that the obtained mass spectrometry data emanates from the correct position of the microarray and thus, by inference, from the intended peptide sequence. To this end, a teaching process was developed, which included visible and specific mass markers (teach marks) in and around the peptide microarray (figure 6) to allow correct positioning of the desorption laser (this new teaching process is patent pending).

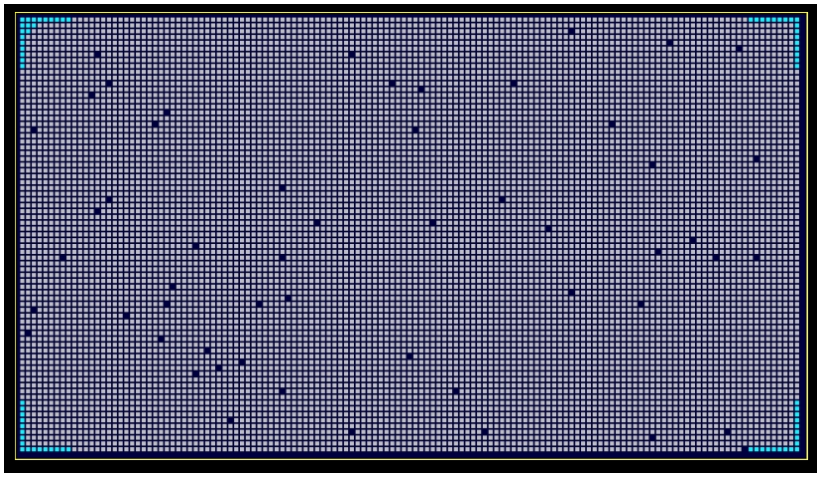


Figure 6. Layout of peptide microarray for MALDI imaging. Teach marks incorporated inside and outside the microarray for correct positioning of the desorption laser beam.

The light directed peptide synthesis is based on Fmoc-chemistry, hence the surface subjected to solid-state peptide synthesis requires free amino groups (-NH2) to initiate synthesis. This posed a challenge to the label free detection techniques, which all depends on a (conductive) metal surface; in particular, the Surface Plasmon Resonance imaging (SPRi) principle rely on gold substrates. Since Gold fulfil all the individual requirements of all the three label free platforms developed in the consortium, an entire work package (WP5) was dedicated to develop advanced surface chemistries on gold substrates. To fully exploit the full area of HD peptide microarrays in the SPRi instruments developed in WP2 it was required to expand the area of Gold nanostructures (nAu) to cover an entire microscope slide, thereby ensuring that both HD peptide microarray synthesis and subsequent analysis was conducted on glass slides measuring 75x25 mm (figure 7).

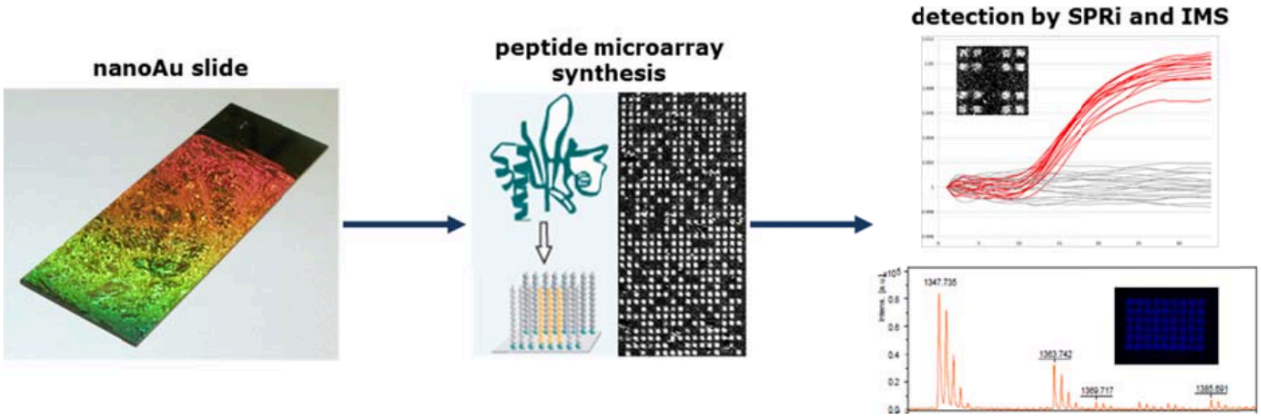


Figure 7. Gold-coated nanostructured slide devoted to array peptide synthesis and detections of interactions by SPR imaging and iMS.

Large area nAu structures were achieved through a nanofabrication process based yielding an area corresponding to entire microscope slides and functionalised with nAu structures (figure 7, right). The optical performance of the nanostructured surface showed a reflectance spectrum with a specific minimum at well-defined wave lengths compatible with nano-SPRi.

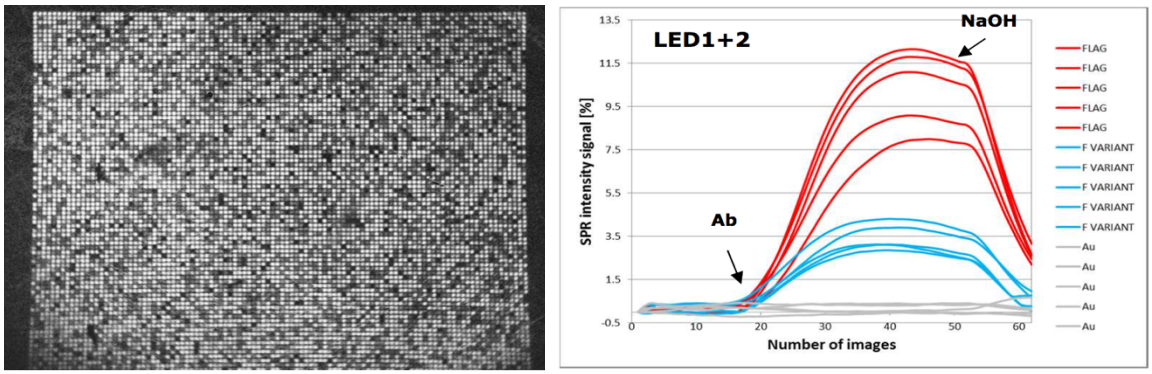


Figure 8. Gold-coated nanostructured slide devoted to array peptide synthesis and detections of interactions by SPR imaging and iMS.

Functionalization of the Gold surfaces were required to enable solid-state-peptide synthesis. Organo-functionalization was achieved by deposition of inorganic compounds, which were further derivatised with functional amino groups. Only one organo-functionlization showed sufficient

chemical resistance to the peptide synthesis chemistries to enable synthesis of full scale HD peptide microarrays. Solid-state peptide synthesis of FLAG peptides and FLAG peptide variants on nAu coated and organofunctionalised microscope slides was analysed with the SPRi prototype instrument developed in WP2. The resulting reflectance image (snapshot) of a monoclonal anti-FLAG antibody reacting with the synthesized FLAG peptides is seen in figure 8, left, and the resulting sensograms of several individual peptide fields is seen in figure 8, right.

Label free detection: SPRi

Two SPRi instruments were developed in WP2. The first developed and assembled prototype was capable of reading 10,000 fields simultaneously. The instrument was integrated with microfluidics and temperature control. The first prototype was not designed to integrate with full HD peptide microarrays synthesized on nanoplasmonic chips. Molecular Interaction Tracking (MIT) software were developed to support the user interaction with the instrument. The MIT software allows control of instrument parameters and data acquisition. The second generation and final high-end instrument was developed on the principles of the first prototype, but with improved hardware specifications. Instrument optics and microfluidics were designed to read full HD peptide microarrays. The instrument highlights are summed in table 1.

Table 1. Comparison of the first SPRi prototype and the final instrument.

Module	First Prototype	Final Instrument
Optics	Resolution of 10^4spots. Limited aberrations	Resolution of 10^5spots. Limited aberrations. Better signal to noise ratio. Better performance of the camera.
Fluidics	Basic liquid cell design. Minimized dead volumes for rapid injection of samples. Continuous flow with automatic injection valve.	Cell "ad hoc" for the microscope glass. Minimized dead volumes for rapid injection of samples. Continuous flow with automatic injection valve, but the flow rate was extended from few nl/min to few ml/min with a new syringe pumping system
Software	Integrated all the electronic control module Software for real-time image analysis and plotting of sensorgrams for up to 10^4ROIs	Integrated all the electronic control module Software for real-time image analysis and plotting of sensorgrams for 10^5ROIs. Possibility to modify LED current

While the SPRi instruments are able to measure the kinetics of molecular interactions, it is not able to identify any modifications of the peptide ligands present on the HD microarrays, nor is it able to identify the composition of the interacting molecules if they are unknown.

Label free detection: iMS

With mass spectrometry it should be possible to detect modifications of the peptides present on the peptide microarrays, e.g. if they are subjected to enzymatic modification. It should also be possible to identify any molecules bound to the peptide microarray, e.g. by peptide mass fingerprinting after tryptic digest.

In situations where peptides (on the peptide microarrays) are subjected to modifications after enzymatic activity, it is necessary to break the peptide linkage to the surface after exposure to the enzyme of interest, but before the MALDI analysis. To solve this, a significant amount of effort was dedicated to the design of a linker chemistry, resistant to the peptide synthesis chemistry and de-protection (removal of side-chain protecting groups), yet readily cleavable after e.g. enzymatic treatment. Several chemical designs were designed and tested in the process resulting in the design of

a chemical linkers with the desired properties. One linker is readily incorporated between the chip surface and the peptides of interest, it is resistant to the UV-light during peptide synthesis, but easily converts to an UV labile linker upon a simple chemical treatment. Another linker is sensitive to gas-phase cleavage. An illustration of a model peptide (substance P) synthesized on a peptide chip, and subsequently cleaved off the peptide chip surface in gas-phase before being subjected to iMS analysis (figure 9). The mass corresponding to the mass of Substance P is green revealing the geometry of the peptide array.

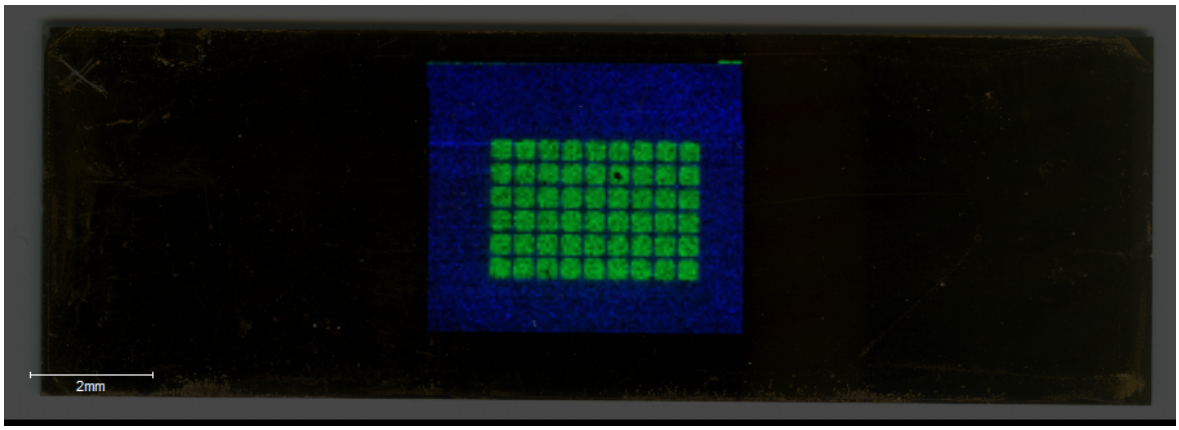


Figure 9. Substance P visualization (green) on a peptide array (6x9 fields) by MALDI imaging gas-phase cleavage of the linker.

As proof of principle figure 10 shows the synthesized peptide (substance P), which was allowed to oxidize before cleavage and iMS analysis; the array geometry is visible (top left) and the mass spec shows the original synthesized peptide as well as oxidized forms.

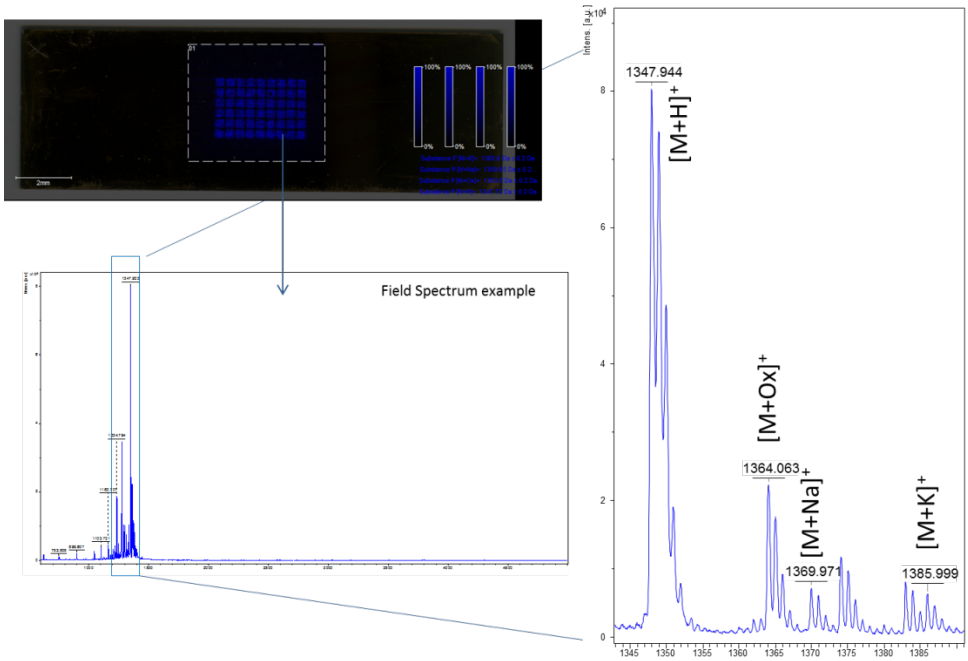


Figure 10. Modifications of of Substance P detected by iMS.

Label free detection: MOSFETs

The third label free detection platform developed by the consortium was metal oxide semiconductor field effect transistors (MOSFETs). The development and assessment of gold electrodes arranged in CMOS-based MOSFET arrays was conducted in WP4. Electrochemical Impedance Spectroscopy (EIS) and Potentiometry (measuring the open circuit potential OCP) were employed to assess the binding of antibodies to peptide epitopes, identified in WP7 by using HD peptide microarrays: four peptides showing high, medium and low reactivity with polyclonal rabbit anti-human serum albumin was selected from a HD peptide array representing all possible 15-mer peptides derived from human serum albumin. The peptide reactivity of the four peptides were measured in a classic ELISA assay confirming one peptide AU367 as a strong binder, Au533 and Au161 as intermediate binders and AU5 as a weak binder of polyclonal rabbit anti-sera (figure 11).

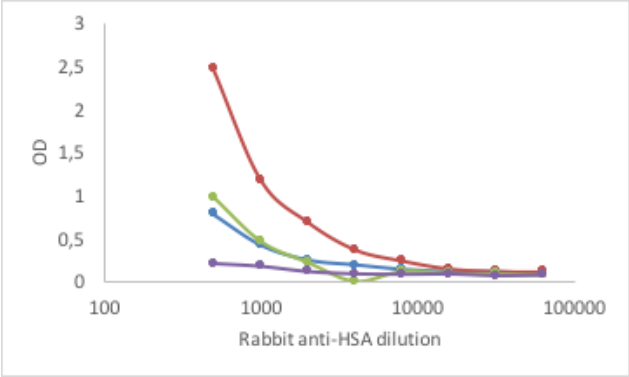


Figure 11. Titration of polyclonal rabbit anti-human HSA against four peptides AU367 (red), AU533 (blue), AU161 (green) and AU5 (purple).

	Interaction of Anit-bodies with different peptides					
	Electrode	**Experiment Stage**	**R_{ct}**	**α**	**% Δ Rct**	**% Δ C**
AU 367 High Binder Peptide	**1**	Ethanolamine	12620	0.999		
		Antibody	16500	0.935	30.74	-6.20
	2	Ethanolamine	9200	0.980		
		Antibody	12300	0.960	33.70	-13.48
	3	Ethanolamine	10100	0.970		
		Antibody	13000	0.960	28.71	-5.41
	4	Ethanolamine	8600	0.986		
		Antibody	11200	0.923	30.23	-6.79
	5	Ethanolamine	12633	0.980		
		Antibody	16200	0.920	28.24	-4.52
AU 5 Low Binder Peptide	**1**	Ethanolamine	11740	0.99		
		Antibody	12450	0.93	6.05	-1.84
	2	Ethanolamine	12690	0.93		
		Antibody	13360	0.93	5.28	-0.83
	3	Ethanolamine	13960	0.98		
		Antibody	14900	0.97	6.73	-3.00
	4	Ethanolamine	13069	0.93		
		Antibody	14090	0.88	7.81	-1.60
	5	Ethanolamine	12633	0.91		
		Antibody	13650	0.86	8.05	-0.64

Table 2. Titration of polyclonal rabbit anti-human HAS against four peptides AU367 (red), AU533 (blue), AU161 (green) and AU5 (purple).

To test whether the differences in binding strength could be measured by gold electrodes the strong and weak binders were immobilised on gold electrodes and changes in charge transfer and double layer capacitance were measured on five individual electrodes for each peptide (table 2).
It was observed that upon binding of the antibody to the peptide (high binder) the charge transfer resistance, changed between 28-33% and correspondingly showing a decrease of 6-13% in the double layer capacitance while for the low binder the change in the charge transfer resistance was calculated to be approx. 5-8% with a decrease in capacitance by 0.6-3% as shown in Table 2**.** The decrease in capacitance can generally be explained as a result of antibody molecules displacing high dielectric constant solvent from the surface. In agreement, the more or stronger antibody bound to the immobilized AU367 peptide is causing a larger change in charge transfer resistance and capacitance than the weaker or lower amounts of antibody bound to the AU5.

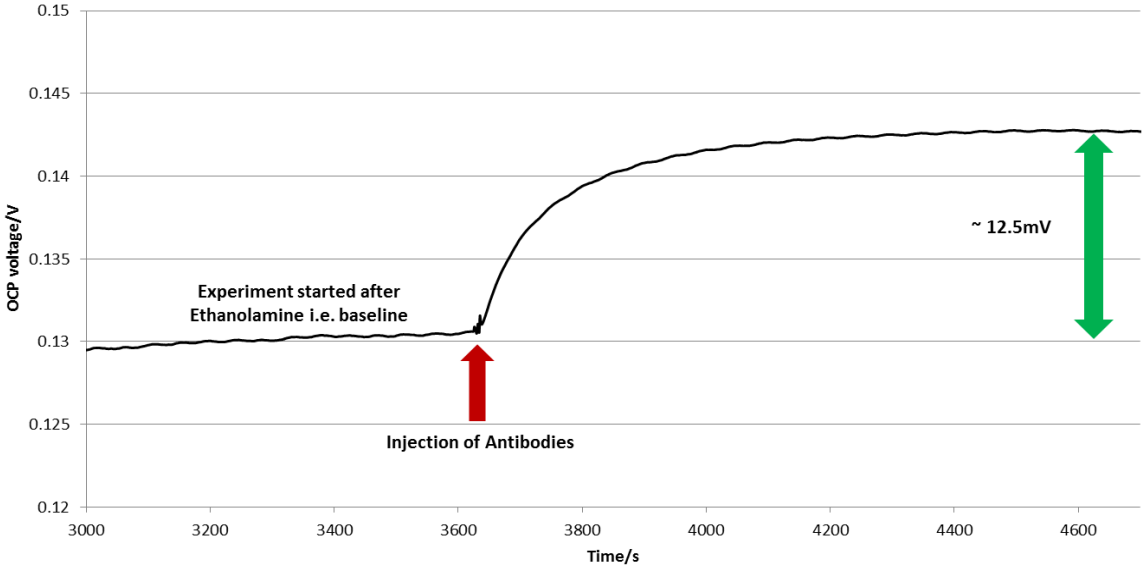


Figure 12. Interaction of polyclonal rabbit anti-HSA sera with peptide AU367 as measured by OCP.

When a bio-layer is immobilized on an electrode bathed in an electrolyte, the change in the charge density and/or distribution at the electrode-electrolyte interface brought about by a bio-molecular interaction induces a change in the open circuit potential (OCP) of the system. For the present setup, reliable measurement of OCP was achieved by employing an ultra-low input bias current instrumentation amplifier charge distribution at the gold electrode. Figure 12 shows open circuit potential measurement of polyclonal rabbit anti-HSA serum binding to the peptide AU367 reaching a maximum of 12.5 mV measured in real-time, using an ultra-low input bias current instrumentation amplifier INA116 (Texas Instruments, USA). The concentration of the antibody employed was 0.1 nM. A measurable signal was also obtained even for the low binder (AU5) peptide of the same concentration, but was consistently measured in the area 0.1 ± 0.1 mV, two orders of magnitude lover than AU367.

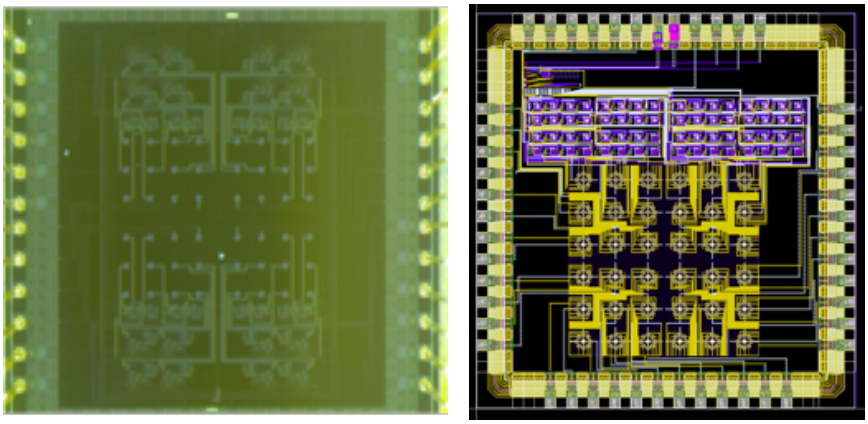
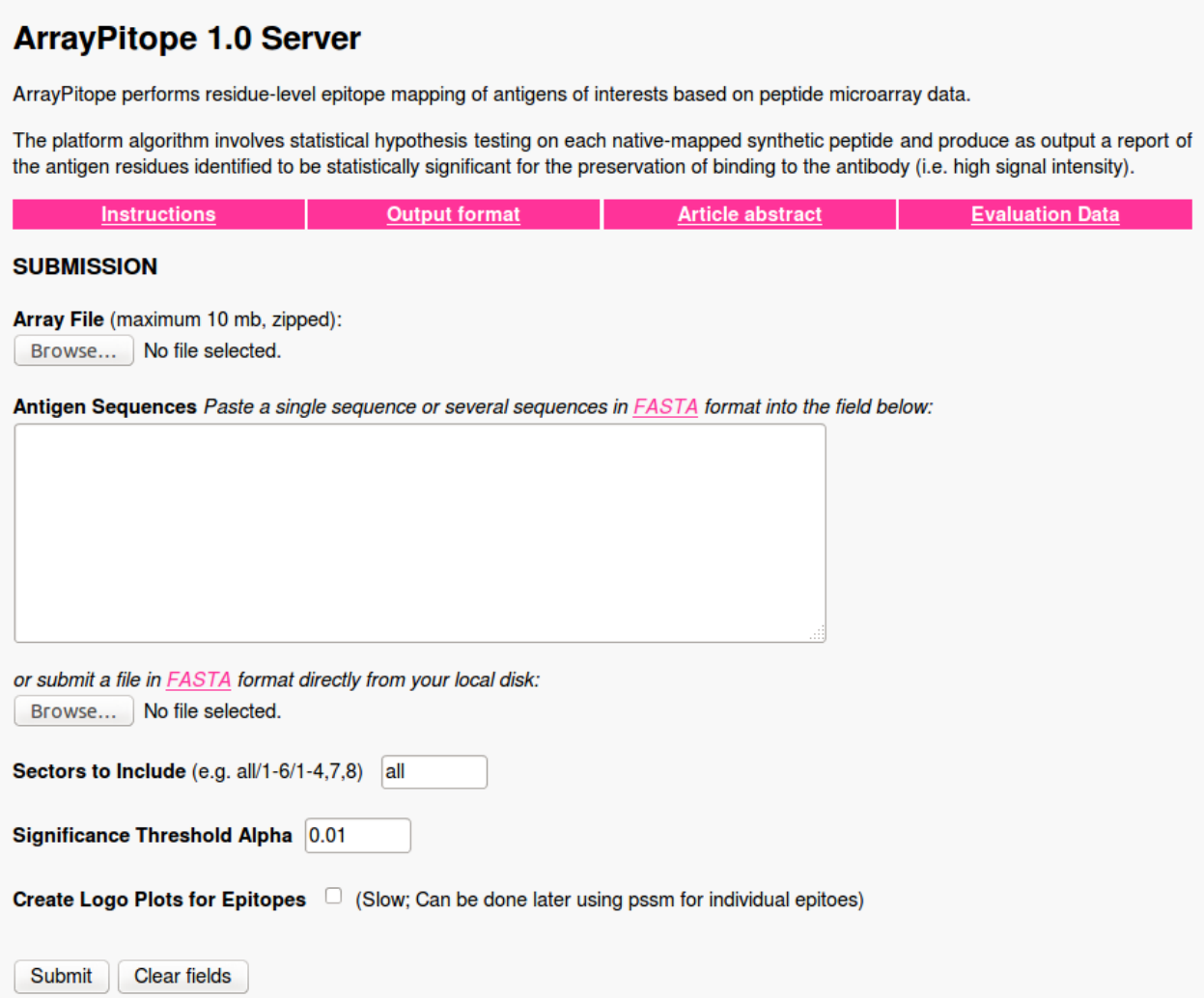


Figure 13. Left 1st generation of CMOS-based MOSFET. Right, 2nd MOSFET generation.

These experiments demonstrate a clear correlation between the results first obtained by HD peptide microarrays, conventional ELISA and subsequently by electrochemical measurements. Hence, both EIS and OCP was considered very useful methods to study molecular interactions between peptides and proteins (exemplified by antibodies). Two generations of CMOS-based MOSFETs were designed and evaluated to assess the use of integrated circuits as label free detection of molecular interactions (figure 13). The first generation of MOSFETs was designed in a 0.35 µm architecture containing 48 electrodes (PADs): 4 for Potentiometric Detection and 2 References (for Differential Measurement), 4 for Electrochemical Impedance Spectroscopy and 2 Test Structures. Subsequently, a second generation MOSFET generation was designed and manufactured in 0.35µm architecture containing 288 electrodes and 36 references with the possibility of on-chip multiplexing. By way of example, we showed that it is possible to use the full scale HD peptide microarrays to identify peptide epitopes, transfer them to a MOSFET setup and measure the same binding. The overall conclusion was that MOSFETs can be used to study molecular interactions between peptides and proteins and the technology may offer advantages over conventional detection approaches in terms of cost, portability, ease of use and possibility to do real time measurements (which enable studying the kinetics of the peptide interactions). Arrays of 256 electrodes can be easily fabricated and expanded to slightly larger sizes, using standard immobilisation procedures (e.g. micro-spotters or microfluidics).

In the HiPAD consortium we have used the HD microarrays to address immune specificities. The amounts of data generated by experiments based on HD peptide microarrays is unprecedented; with each experiment routinely generating > 200,000 data points, it classifies as big data. To analyse and interpret such data, bioinformatics tools are a requirement. To ensure un-biased designs and that right information can be extracted from the experiments, designing experiments (i.e. the layout and contents of HD peptide microarrays), should also be given bioinformatics considerations. Finally, despite the high capacity of the HD peptide microarrays, it is important not to waste capacity by just including e.g. every overlapping peptide, but to ensure that redundancy is filtered out etc. The bioinformatics was dedicated to work packages 6.

Peptides are ideal to study linear B cell epitopes. By way of nature, antibodies originating from immune responses raised against protein antigens during a natural immune response will be polyclonal and recognize both discontinuous and continuous epitopes. For a number of reasons, it is desirable to know the specificity of antibodies, i.e. targets as well as off-targets; in particular to identify antibody epitopes, preferably at high resolution, suggesting which specific residues are directly (e.g. contact residues) or indirectly (e.g. framework determinants) involved in the antigen-antibody interaction. Such information could be vital to the identification of antibody targets as well as off-targets. HD peptide microarrays are ideally suited to search for linear B cell epitopes; not only is it possible to identify single peptide as targets, but through systematic substitution of every single position in a peptide, it is possible to identify important specificity residues. In WP6, we have built and released a public accessible web-server which accepts raw data obtained from HD peptide arrays, performs statistical hypothesis testing on every native-mapped peptide and produces an output of the antigen residues identified to be statistical significant to preserve antibody recognition. Screen caps of the server http://www.cbs.dtu.dk/services/ArrayPitope-1.0/ and output can be seen in figure 14.



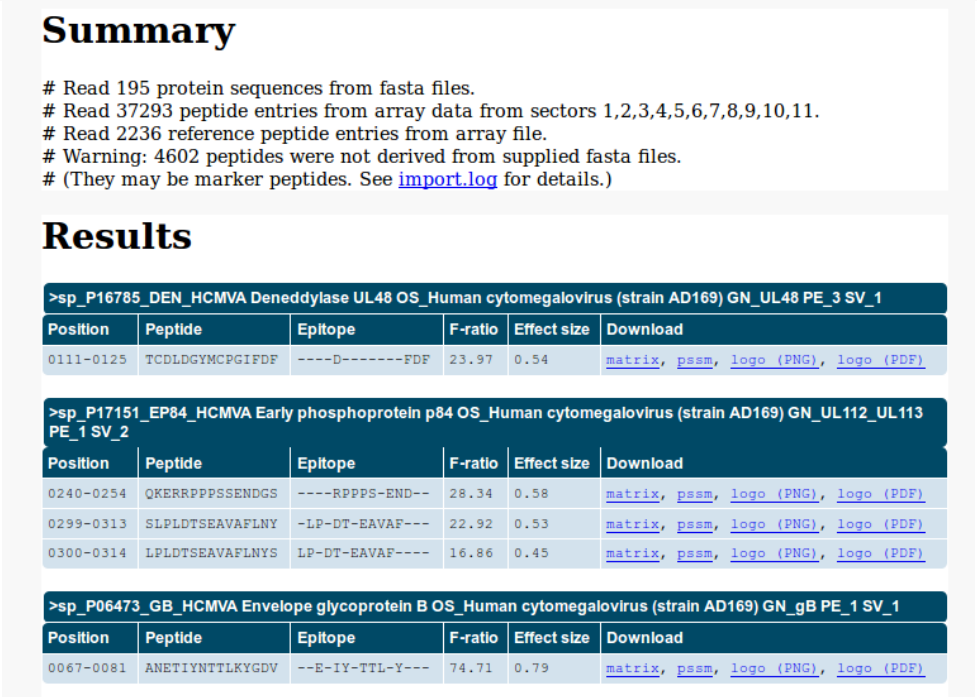


Figure 14. Scree caps of the ArrayPitope server (top) and an output example (bottom).

The capacity of HD peptide microarrays allows entire proteomes to be represented on a single microarray. By way of example, we examined the entire cytomegalovirus (CMV, HHV-5) proteome against a pool of serums from CMV-infected donors. Each protein is represented by overlapping peptides as illustrated in figure 15. Following incubation with serum of interest, the signals from each peptide are recorded and target peptides are identified. This does, however not provide information about the actual epitope or residues upon which the specificity of the antibody binding depends on.

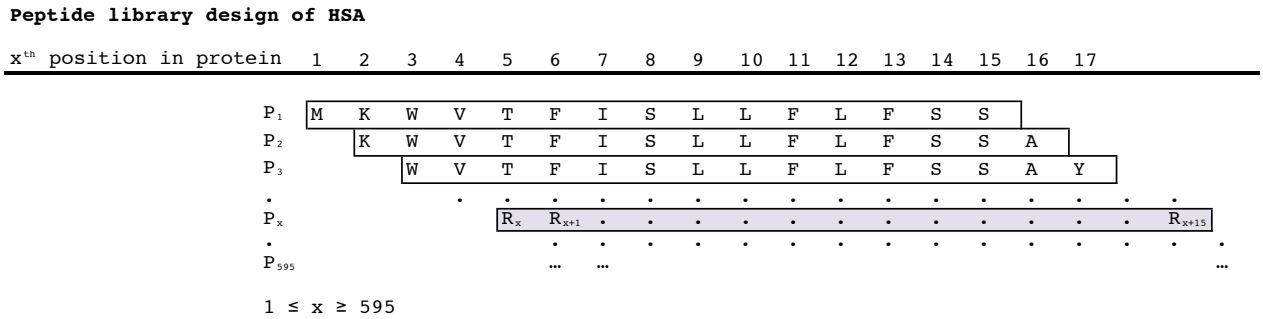


Figure 15. Design principle of overlapping peptides, here covering the entire Human Serum Albumin

To identify these specificity residues, which functionally approximates "contact residues", we have employed a strategy entailing a full substitution of each residue, i.e. each position (residue) in the peptide of interest is substituted by the other 19 amino-acids, one by one. Hence, each 15-mer peptide of interest gives rise to 19*15 = 285 single substitution variants, which are examined for antibody binding. By submitting the raw data obtained from a HD peptide microarray containing signal from native and single amino-acid substituted peptides to the ArrayPitope server, it returns a high-resolution mapping of each peptide of interest.

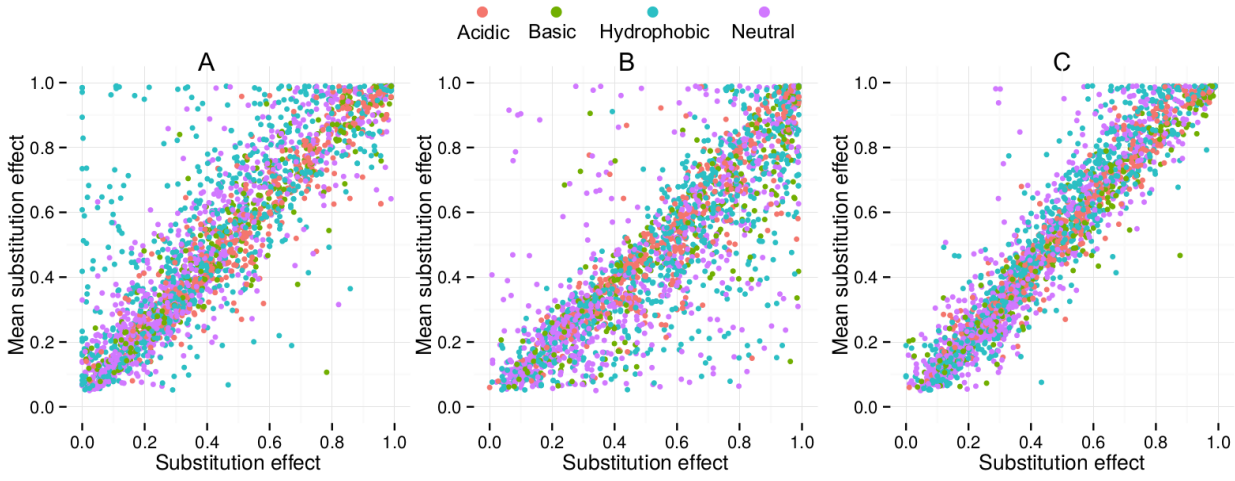


Figure 16. Correlation of the substitution effect (i.e. the relative loss in signal upon substitution) of individual replacing amino acids (x-axis) with the mean substitution effect of all 20 natural amino acids (y-axis) on epitope-residues of cytomegalovirus proteins. Substitutions to alanine (A), D-alanine (B) and in situ mixture of all 20 natural amino acids (C). The original amino acid type is displayed in color according to its chemical properties as follows: Acidic [DE], basic [HKR], hydrophobic [ACFILMPVW] and neutral [GNQSTY]

The central idea behind single amino-acid substitution is, that if a key contact residue is substituted with another amino-acid, the interaction with the specific antibody is weakened or lost, reflected by a ditto change in signal strength. Since each putative peptide epitope under investigation requires additionally 285 peptide variants (preferentially in triplicates), the number of peptides needed quickly grow and exhausts even a HD peptide microarray. To increase the capacity, other substitution schemes have been evaluated, including a basic Alanine scan, an Alanine scan with the stereo-isomer D-Alanine and multiplexing all 20 amino acids in the position subject to substitution. A comparison of the different substitution schemes against the mean substitution effect for all 20 amino-acids is seen in figure 16.

A benchmarking plot of the substitution comparison (figure 17) revealed that a multiplexing strategy was the better choice as it had the highest correlation to the exhaustive substitution scheme of using all possible substitutions individually. Although the multiplexing scheme increased the capacity of high-resolution epitope discovery, it comes with cost; e.g. it is not possible to determine heteroclitic responses or to determine which substitutions are neutral, or most effective.

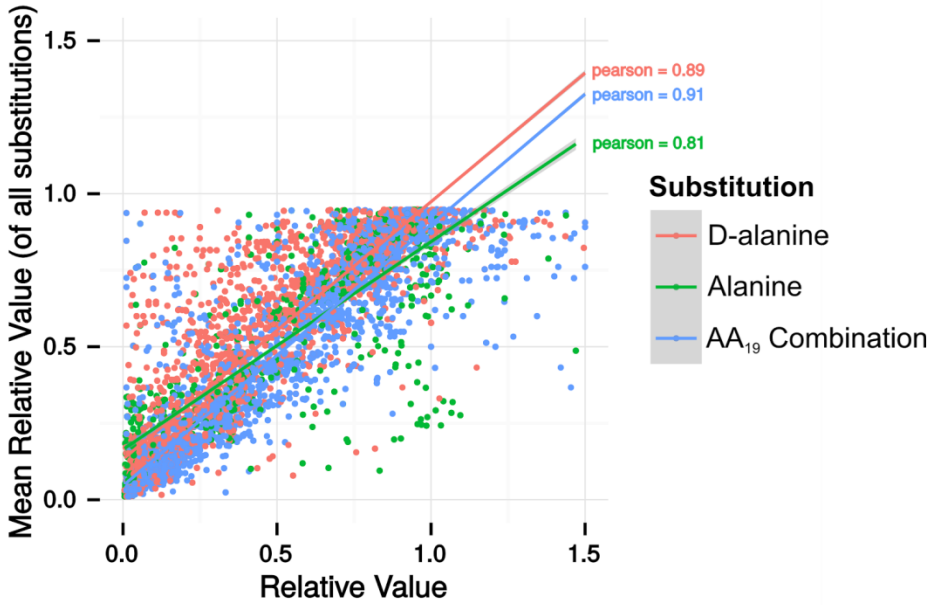


Figure 17. The substitutional effect of three replacement schemes are shown along the x-axis as their binding intensity relative to the median of copies of the original peptide of which an amino acid was replaced, against the mean substitutional effect of individual peptides deriving by all 20 amino acids (y-axis), individually calculated the same way. Pearson correlation coefficients are shown in text for the three schemes.

In WP7, the aim was to use the developed technologies to address immune specificities, in particular B cell responses and MHC specificities. Recombinant MHC class II molecules was expressed, purified and refolded, in the end with at a high quality comparable to the quality previously obtained for MHC class I molecules. We have systematically tested DRB1*03:01 and DRB1*01:01 on several chips and designs. We have assessed the assay (MHC II binding to peptide chips) reproducibility and established synthesis, assay and data interpretation for peptide chips containing non-standard amino-acids (evaluated with DRB1*01:01, DRB1*03:01 and DRB1*04:01). Seamless incorporation of non-standard amino acids in the peptide synthesis chemistry allows us to array peptides which mimic post translationally modified (PTM) proteins. In an immunological perspective it is of significant importance to be able to address PTM proteins; for instance, cancer cells are known to express special phosphorylation patterns (newer drugs are targeting phosphorylation signatures). In the

autoimmune disease, rheumatoid arthritis, auto-reactive T cells are found to react to citrullinated peptides (peptidylarginine deiminases (PADs) enzymatically converts Arginine to Citrulline).

By including non-standard amino acids on high-density peptide arrays, we are able to address the binding of modified peptides to MHC class II, a prerequisite for T cell recognition.

Assay reproducibility and MHC II specificity
A chip design containing 72,000 random 13-mer peptides in triplicates were used to synthesize chips on two different dates to evaluate the assay variance. The raw signals from each chip were transformed (z-scaled) and the signals from each peptide were plotted in a scatter plot (figure 18). For both molecules tested (DRB1*01:01 and DRB1*03:01) we found a Pearson correlation of approx. 0.9 which demonstrates a very good reproducibility. The two different MHC II molecules have very different specificity, this is demonstrated by the Pearson correlation of 0.27 (figure 19). Hence, the binding of MHCII to peptide chips is molecule specific and reproducible.

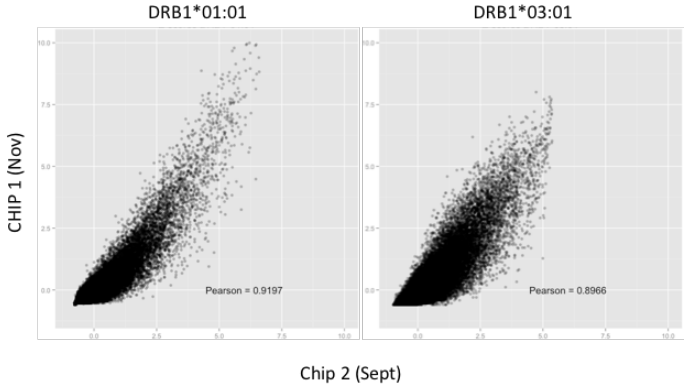


Figure 18. Correlation of z-scaled raw signals from to different chips synthesized in September 2015 and November 2015.

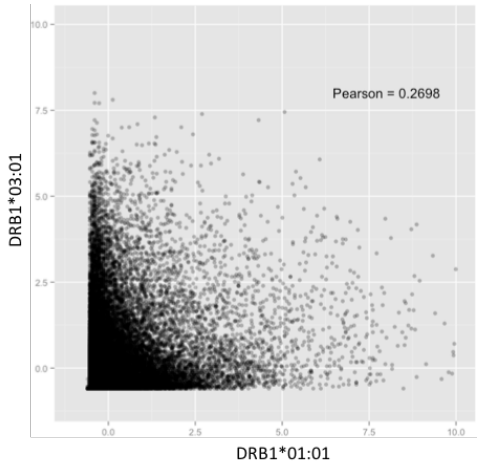


Figure 19. Z-scores from each peptide obtained after binding of DRB1*01:01 and DRB1*03:01 to peptide chips containing 72k random peptides in triplicates.

Binding analysis of peptides with non-standard amino acids
A chip design containing approx. 72,000 random peptides with a natural distribution of amino acids. 10% of Arginine were randomly selected to be substituted with Citrulline which was included as the

21st amino acid during peptide synthesis. The chip was 1) subject to incubation with HLA-DRB1*04:01 and HLA-DRB1*03:01 2) data acquired on a laser scanner and used to 3) train neural networks (ANNs) for each molecule. The resulting ANNs show a very high correlation (Pearson) between predicted and observed signals (binding value) (figure 20)

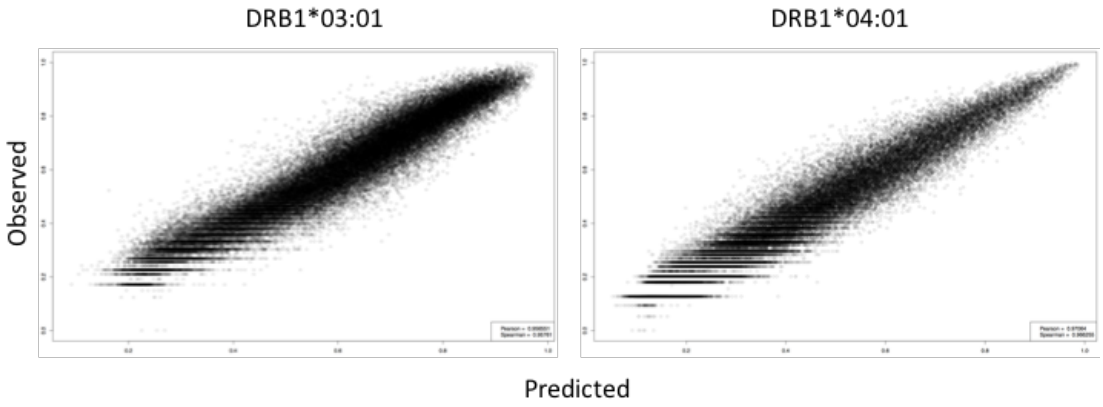


Figure 20. Scatter plot of Predicted versus Observed signal (MHC II binding). Pearson correlations of >0.95 for both molecules suggest the resulting ANNs have a very good "understanding" of the individual MHC II specificities.

To illustrate the power of the combined use of high-density peptide chips and bioinformatics driven ANNs, we predicted (pure in silico experiment) the binding of overlapping peptides from a known Rheumatoid arthritis antigen, Vimentine, and the corresponding Citrulline modified peptides. Figure 21 shows that several peptides which have undergone post translational modification (Arg > Cit) are very strong DRB1*04:01 binders compared to the native peptides. Persons carrying the DRB1*04:01 allele have an increased susceptibility to develop Rheumatiod arthritis and a few DRB1*04:01 restricted T cell epitopes have been identified in RA patients. In figure 21, these T cell epitopes (blue) have been overlaid the predicted binding scores. Firstly, it is clear that a Arg > Cit modification is potentially antigenic since a stronger DRB1*04:01 binding is predicted for several peptides; secondly, it is also clear that the known RA associated T cell epitopes coincides with the predicted potential immunogenic regions.

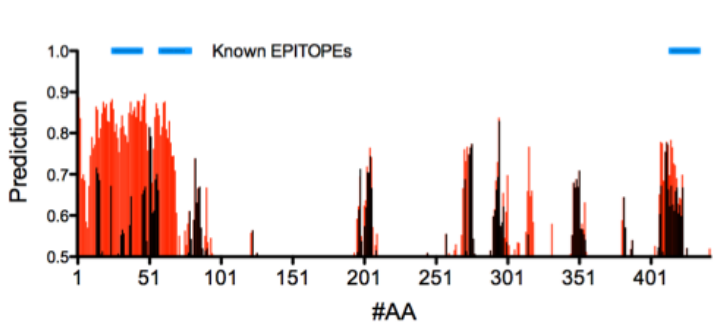


Figure 21. Binding scores of overlapping peptides derived from Vimentin. Native peptides represented by black bars and citrullinated peptides in red. Blue bars indicate know DRB1*04:01 restricted T cell epitopes found in RA patients.

Conclusion and outlook

The HiPAD project was a complex multidisciplinary project with several high-risk high-gain aims. Not all goals were achieved completely, however the overall output from the project has contributed with new instruments setting new standards within their respective fields, and the cornerstone of the project, the HD peptide microarrays, has proven to be a tool for the future within proteomics. Several results and achievements are already being commercialised whereas other achievements need additional developments. Gold surfaces have proved to be particular difficult to functionalize to a level where the organofunctionalization can resist the harsh chemical conditions during peptide synthesis. It is possible - but a further optimisation of surface chemistries and polymers to ensure consistently durable surface coating lies ahead.

Potential impact, main dissemination activities and exploitation

HiPAD was a complex and multidisciplinary project aimed at extending the limits of proteomics (the discipline of studying proteins and their function), achieved through development and improvement of new label free detection platforms integrated with High Density peptide microarrays.

Proteins are essential to all life and all living organisms contain and depend on expression of a large palette of proteins with specialized functions and purposes. Proteins are indispensable to us in many ways, from an essential component in food, targets used in treatment of diseases and infections to enzymes used in households and industry, to mention a few.

The advancements made within the HiPAD consortium has pushed the limits within the field of studying protein interactions with peptides, including tools to determine patterns recognised by proteins (in the project demonstrated through use of antibodies and MHC molecules). In this perspective, any industry which exploit proteins can benefit directly from the HiPAD project.

New instrument hardware has been developed during the project. Plasmore (a SME partner) have developed a new Surface Plasmon Resonanse imaging device based on the iNPx technology developed. The instrument pushes the number of probes which can be monitores simultaneously in real time with several order of magnitudes. The final instrument can report kinectic data from up to 100,000 ligands (small molecules, peptides etc) on a single chip. The instrument is compatible with, but not limited to HD peptide microarrays since the iNPx chips developed are compatible with e.g. micro spotting.

Industrial partner Bruker, has developed a new smart-beam laser, with a capacity to acquire 5 spots/sec it sets a new standard MALDI imaging. SME partner ImaBiotech, has developed MALDI imaging software based on big data architecture in order to manage large datasets; this software is already being used by ImaBiotech in the FP7-HEALTH CARTARDIS (Identification and validation of novel pharmaceutical drug targets for cardiovascular disease, started in October 2013).

SME partner Schafer-N has optimized the production of HD peptide microarrays with additional instrumentation and improved chemistries resulting in a decreased turn-around time for HD peptide microarray synthesis and increased the commercial activity related to HE peptide microarrays.

Altogether are the participating enterprises in a position where new products have been brought closer to market or already on the market. Further commercial exploitation and expansion of the technology and or services offered to industries founded on proteomics is foreseen by the partners. Five patent applications were filed by the participating partners which are expected to be commercially exploited by the partners.

The CMOS-based MOSFET technology was used to explore whether peptide ligands extracted from e.g. scouting or discovery experiments could be transferred to and observed for molecular interactions of interest. Indeed, we found the MOSFETs to be able to detect the interaction between peptides and antibodies with a remarkable sensitivity. We found that MOSFETs can deliver both quantitative data as well as kinetics when integrated into microfluidics systems. Presently, the discovery or scouting potential for MOSFET technology is not relevant. However, the low production costs, on-board multiplexing and capacity suggest that the technology has potential within point-of-care and as "electronic nose"; the applications, of course, depends on the availability of relevant peptide ligands (e.g. extracted from HD peptide microarrays). Further development of MOSFET driven instrumentation aimed at measurement of biological or biochemical activity is expected to depend on further funding.

By participating in the HiPAD consortium, the academic partners have benefitted from access to emerging technologies. Within the field of immune specificity, we have established the used of HD peptide microarrays as an extremely valuable tool to address specificities of antibodies and MHC molecules, which, by large, is centred around small structures, in particular peptides. New bioinformatics tools have enabled us to fully use the large amounts of data generated and develop algorithms capable of predicting immune specificity and subsequently closing in on immunogenicity. This is highly relevant to the industry dedicated to vaccine development and businesses focusing on new immune (cell) therapies.

Based on the acquired know-how, and the development of MHC class II recombinant molecules, a new activity (www.ImmunAware.com) is being pursued by the University of Copenhagen. Initially, the activity will operate as a commissioned enterprise but anticipated to operate as an independent enterprise within 1-2 years offering recombinant MHC class I and II technology and immunogenicity assessment services

Another Spin-out company Bio-Shape Ltd. www.bio-shape.com was co-founded in November 2015 by Dr Hannah Roberts (UNIMAN), Prof. Sabine Flitsch (UNIMAN), Prof. Perdita Barran (UNIMAN) and Prof. Claire Eyers (UNIMAN/UoL). Bio-Shape Ltd. is a contract structural characterisation and method/instrument development facility for biomolecules operating from The BioHub in Cheshire (UK). Bio-Shape Ltd. provides state-of-the-art native mass spectrometry services utilising ion mobility mass spectrometry (IM-MS). Using this technique, the molecular non-covalent interactions are retained during the ionisation process and subsequent mass analysis. Detailed post-translational modification (PTM) analyses can be performed at the intact protein level or following the enzymatic release of glycans. A more in-depth glycan analysis service is also available for quantitative glycan monosaccharide composition analysis and structural characterisation of glycans including separation of isobaric species.

The HiPAD consortium has disseminated its activities in several ways; in particular, we have reached out to an international scientific audience. Non-confidential activities and results have been published in peer-reviewed scientific journals or is being prepared for publication. General techniques have also been publiched in book chapters or as invited articles. Representatives from the consortium have appeared on several international conferences with oral and/or poster presentations. Lectures where given at PhD courses in Scientific Project Management highlighting HiPAD as a project example of multidisciplinary scientific projects with both academic and business partners.

Dissemination of HiPAD activities outside the scientific have been sparse, but UNIMAN were successful in their bid to host "The Complex Life of Sugars" at BBSRC The Great British Biosciences Festival, 14-16th November 2014, London. The exhibition was visited by over 6,500

visitors including people from the following categories: students, teachers, public, scientists, media, potential donors/key decision makers and celebrities. One of the main exhibition activities was focused on cell surface sugars and visitors were encouraged to build a cell surface sugar and explore its interaction with cell invaders both on a cell surface and also a gold glycan array. This activity was designed to highlight and directly promote the hiPAD work (WP3) at UNIMAN and proved very popular. For more contents please visit our blog http://sugar-complexity.tumblr.com/ and twitter feed @sugarcomplexity

In addition to the scientific dissemination already published, we expect to publish several new potential high impact scientific reports in peer-reviewed journals. A list of publications and other activities can be found at http://hipad.ku.dk.