The Challenge of Antigen Sequence Diversity: Solutions with ULTRA-Peptide Libraries

U. Reimer

JPT Peptide Technologies, Volmerstrasse 5 12489 Berlin, Germany

Sequence diversity is a hallmark of many pathogenic viruses including HIV and Influenza, rendering the control of such infections difficult. However, recent DNA sequencing efforts have revealed similar sequence diversity among eukaryotes, including humans. Additionally, certain diseases, in particular cancer, are linked to increased numbers of somatic mutations. More complexity is added by post-translational modifications which further alter proteins and give rise to potential neo-epitopes.

Sequence diversity plays an important role in immune recognition and has to be taken into account for development of immunomonitoring or -therapy approaches. Peptide libraries are ideally suited for accommodating such diversity. Here we describe approaches to representing this diversity with a minimum number of peptides.

Introduction

Sequence diversity on the level of the genome and proteome is known for many pathogens such as HIV or Hepatitis B Virus (HBV) and poses a tremendous challenge for immunological research.

Chemically synthesized peptides are broadly used for the monitoring of T and B-cell immune responses, for the stimulation of effector cells and as vaccines. Ideally, such peptides reflect the exact sequence of the targeted antigen. Although, recent NGS advances allow identification of such sequences for development of state of the art personalized therapeutics and/or diagnostics labour, cost and time remain prohibitive for broad application. Therefore, comprehensive, chemically produced peptide libraries targeting the sequence space of all clades or serotypes of a given pathogenic antigen are a powerful approach to address the complexity of sequence diversity. Here we present a concept for the design of generic, all-inclusive peptide libraries with an optimized coverage of all sequence variations in a given genetically heterogeneous population.

The resulting sequence design approach is applicable for the generation of peptide libraries and pools to stimulate antigen specific T-cell responses as well as for peptide microarrays and ELISA to profile broadly humoral immune responses at the epitope level.

Materials and Methods

PepMix™ peptide pools are frequently used for T-cell stimulation in T-cell assays such as ELISPOT, ICS or Flow Cytometry. PepStar™ peptide microarrays are efficient tools for the characterization of the antibody repertoire of serum samples with a consumption of 1µl serum per experiment. Our standard products make use of a common library format in which a single antigen is represented by 15 amino acid peptides overlapping by 11 amino acids. Wheras PepMix™ overlapping peptide pools typically represent a single antigen, the **PepStar™ peptide microarrays** may represent multiple antigens encompassing up to a maximum of ~7000 peptides.

Library Design: Our aim was to generate libraries with optimal sequence coverage for a given set of sequences. The developed algorithm creates all possible peptides and scores these according to their frequency of occurrence across all sequences. It so finds the optimal overlap required to provide homogeneous overall coverage. However, in some cases the set of input sequences is too large or too diverse (e.g. 6225 sequences for HIV-1 gag) to be covered in a library for PepMix™ peptide pools or PepStar™ peptide microarrays. In such a case, the library size may be limited by selecting peptides according to their contribution to overall coverage.

The resulting peptide collections are termed ULTRA peptide libraries. Several such ULTRA peptide libraries have already been designed and manufactured (Table 1). The main target organisms to which the concept has been applied are viral pathogens. For HIV numerous sequences of many different clades are known. For HIV-nef a peptide library generated from the model sequence HXB2 (49 peptides) achieves a coverage of < 10 % for 3903 complete nef sequences (Fig. 1). The coverage for the native clade B of HXB2 the coverage is with 13.5 % little better. An Ultra library of 150 (PepMix) or 667 peptides (Array) achieves a total coverage of 51.8 or 56.3 %, respectively. The coverage for the B-clade is 57.3 and 58.5 % for these libraries.

Table 1: Overview of available ULTRA products.

Product	Source Organism	Antigen	No. of peptides
PepMix™	HIV	env	150
PepMix™	HIV	gag	150
PepMix™	HIV	pol	150
PepMix™	HIV	nef	150
PepStar™	HIV	Proteome*	6564
PepMix™	HBV	C	154
PepStar™	HBV	Proteome	4782

* immunogenic fraction

Results

The PepMix™ ULTRA peptide pools for HIV gag, pol and nef were recently used in a proof of concept study for the expansion of broadly specific cytotoxic T-cells from HIV positive donors (3). Patients were not pre-selected in terms of HLA-type or disease prognosis. Functional, broadly-specific cytotoxic T-cells were successfully expanded from all 7 investigated patients. This is a first step towards the development of generic, all-inclusive cellular therapies for HIV.

For B-cell immune profiling a comprehensive PepStar™ ULTRA peptide Microarray providing almost complete coverage of the immunogenic fraction of the HIV proteome was also recently used (4). Such microarrays provide a method for measuring the magnitude, depth and breadth of IgG binding to linear HIV epitopes.

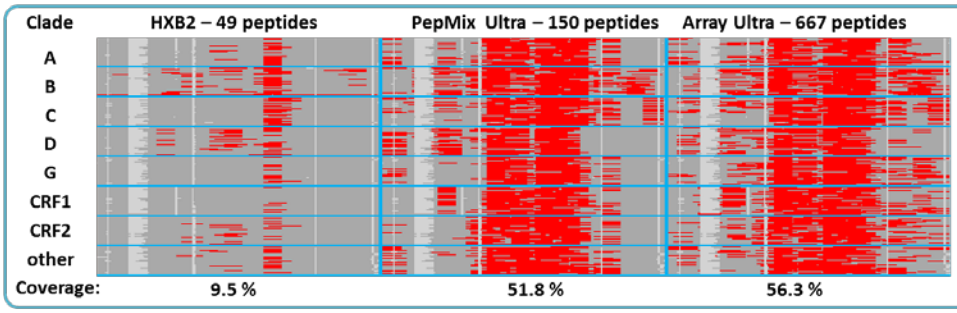

Fig. 1: Sequence coverage for HIV-nef libraries. The alignment of HIV nef sequences is presented on the x-axis. On the y-axis 30 exemplary sequences for each clade are shown. The coverage is calculated for the complete set of 3903 sequences from the 2014 alignment from the Los Alamos National Laboratory. Sequence parts which are covered by the respective peptide library are shown in red. Non-covered regions are shown in dark grey. Light grey regions indicate gaps

Conclusions & Outlook

ULTRA peptide libraries are ideal tools to include sequence diversity into immunological research, both on the T-cell and the B-cell side.

This concept is not only limited to pathogens but also applicable to human diseases connected to high genetic variability such as cancer. Recent advancements in DNA-sequencing technology have facilitated the sequencing of thousands of complete genomes. The 1 000 Genomes and other projects have provided information on the genetic variability between individuals within and across different ethnic groups. On average 10 000 to 12 000 sites with coding mutations have been identified per individual (1). Given about 20k genes per individual this translates into a sequence alteration for one in two genes. These germline mutations are observed uniformly in each cell of an organism. Furthermore, somatic mutations cause sequence heterogeneity within individuals which are of special importance for pathological mechanisms in cancer. Exemplarily, for one of the most important tumor driver genes, TP53, 767 germline and 29 881 somatic mutations have been identified (2).

For the coverage of mutations leading to amino acid variations, an alternative library format can be utilized. The variation is positioned in the center of a 17 amino acid peptide. This design ensures the presence of each possible 9 amino acid peptide that contains the mutated residue. Additionally, such libraries can also be used to cover posttranslational modifications (PTM) of proteins (5).

The presentation of sequence diversity in immunological applications is important to address whole populations which differ by germline mutations and enable the detection of immune responses towards neo-epitopes created by somatic mutations or post-translational modifications.

References

1. The 1000 Genomes Project Consortium (2015) A global reference for human genetic variation. *Nature* **526**, 68-74.
2. Petitjean, A. *et al.* (2007) Impact of mutant p53 functional properties on TP53 mutation patterns and tumor phenotype: lessons from recent developments in the IARC TP53 database. *Hum. Mut.* **28**, 622-9.
3. Lam, S. *et al.* (2015) Broadly-specific Cytotoxic T Cells Targeting Multiple HIV Antigens Are Expanded From HIV+ Patients: Implications for Immunotherapy. *Mol. Ther.* **23**, 387-395.
4. Stephenson, K.E. *et al.*, (2015) Quantification of the epitope diversity of HIV-1-specific binding antibodies by peptide microarrays for global HIV-1 vaccine development. J. Immunol. Methods **416**, 105-123.
5. Masch, A. *et al.*, (2015) Comprehensive Characterization of Antibodies Directed towards Epigenetic Histone Modifications *JPT Application Note at www.jpt.com*

The Author

Dr. Ulf Reimer
reimer@jpt.com
Head of Research & Development
JPT Peptide Technologies
Berlin, Germany

Dr. Ulf Reimer is Head of Research and Development at JPT. He studied Biochemistry at Martin-Luther University Halle/ Wittenberg and completed his PhD Thesis in the "Enzymology of Protein Folding" research group at the Max-Planck-Society at Halle/Saale. From 1998 to 2000 he was a postdoctoral researcher in the same group and in the group of Prof. Griesinger at the University of Frankfurt/Main. Then he joined Jerini AG in Berlin, where he started as a scientist and later became Director of Computational Chemistry at Jerini AG, in which he was jointly responsible for drug discovery and development.

The Company

JPT Peptide Technologies is a DIN ISO 9001:2015 certified and GCLP compliant integrated provider of innovative peptide solutions for: cellular and humoral immune monitoring, seromarker discovery & validation, vaccine target discovery, peptide lead identification & optimization, targeted proteomics, and enzyme profiling.

Contact us for further information!
email: peptide@jpt.com
phone: +49 30 6392 7878

Please visit us online www.jpt.com
Further information: ULTRA Peptide Libraries