

Method for the characterization of peptide:MHC binding polypeptides

The present invention relates to a method for the characterization of peptide:MHC binding polypeptides, e.g. by mass spectrometry and an analysis of the recognized peptide space, i.e. in order to identify peptides that can be bound in the context of their presentation by MHC, and those who can not be bound.

Background of the Invention

Immunotherapy has obtained a prominent role in the field of oncology and has proven valuable in the treatment of different types of tumors. A range of immunotherapies varies from chimeric antigen receptors (CARs), expanding tumor-infiltrating lymphocytes (TILs) and T-cell receptor (TCR)-transduced effector cells. Various studies successfully make use of TCR-engineered T-cells to enhance patients' adaptive immune responses against malignancies, demonstrating potent anti-tumor reactivity. The efficacy of genetically modified T-cells against cancer is greatly improved at the expense of enhanced toxicities, however.

Off-target toxicity occurs when a transduced T-cell population unexpectedly attacks an antigen other than the intended one or activates themselves independently from their specificity.

US 2018/0125889 indicates that a gamma delta T-cell may be engineered to express an inhibitory CAR, which minimizes activation in off-target cells, e.g. non tumor cells, in which the cell surface target is a tumor-associated, but not tumor-specific antigen. WO 2018/053374 describes a T-cell epitope screening method to predict or study the toxicity and/or off-target effects of T-cells, TCRs, or TCR-like molecules, including contacting an engineered target cell, or a population of engineered target cells, with a T-cell, a TCR, or a TCR-like molecule, and performing an assay to determine whether the T-cell, TCR, or a TCR-like molecule binds to the engineered target cell, or population of engineered target cells, and/or to measure the strength of any such binding.

Bijen *et al.* (Bijen *et al.*, 2018) found off-target recognition by the 7B5 T-cell clone of human fibroblasts and keratinocytes that are not expressing the HA-2 antigen. Bijen *et*

a/. discloses a combinatorial peptide library scanning approach to identify an off-target peptide, i.e., CDH13-derived peptide, recognized by the 7B5 T-cell clone.

There remains a need to reduce the off-target toxicity in immunotherapy.

There are two classes of MHC molecules. MHC I or II that can be found on most nucleated cells. MHC molecules are composed of an alpha heavy chain and beta-2-microglobulin (MHC class I receptors) or an alpha and a beta chain (MHC class II receptors), respectively. Their three-dimensional conformation results in a binding groove, which allows for a non-covalent interaction with specific peptides.

MHC class I receptors present peptides that result from proteolytic cleavage of predominantly endogenous proteins, defective ribosomal particles (DRIPs) and larger peptides. MHC class II receptors can be found predominantly on professional antigen presenting cells (APCs), and primarily present peptides of exogenous or transmembrane proteins that are taken up by APCs during the course of endocytosis and are subsequently processed.

Complexes of peptide and MHC class I molecules are recognized by CD8-positive cytotoxic T-lymphocytes bearing the appropriate T-cell receptor (TCR), whereas complexes of peptide and MHC class II molecules are recognized by CD4-positive helper-T cells bearing the appropriate TCR. During this recognition process, it is well known that the TCR, the peptide and the MHC are present and form a complex in a stoichiometric amount of 1:1:1.

For a peptide to trigger (elicit) a cellular immune response, it must bind to an MHC-molecule. This process is dependent on the allele of the MHC-molecule and specific polymorphisms of the amino acid sequence of the peptide. MHC-class-I-binding peptides are usually 8-12 amino acids in length and usually contain at least two conserved residues ("anchoring residues", AR) in their sequence that interact with the corresponding binding groove of the MHC-molecule. In this way, each MHC allele has a binding motif that controls the peptides' ability to specifically bind to the binding groove. Nevertheless, as mentioned above, in the MHC class I dependent immune reaction, peptides not only have to be able to bind to certain MHC class I molecules being expressed by tumor cells, they also have to be recognized by T-cells bearing

specific T-cell receptors (TCRs). The antigens that are recognized by the tumor specific cytotoxic T-lymphocytes, that is, their epitopes, can be molecules (peptides) derived from all protein classes, such as enzymes, receptors, transcription factors, etc. which are expressed and, as compared to unaltered cells of the same origin, up-regulated in cells of the respective tumor.

Many cancer immunotherapies that are currently in development rely on the administration of peptide:MHC binding polypeptides to a subject in the form of a soluble polypeptide molecule or a transfer of cells, preferentially T cells, expressing such polypeptides as membrane bound molecules.

While the actual target peptide sequence for such peptide:MHC binding polypeptides is usually established/defined, there may be an unknown number of additional peptides that these molecules can bind to. These so called "off-target peptides" constitute a significant safety risk for the application *in vivo*, due to potentially serious side-effects. The reason for these side-effects is usually the presentation of such off-target peptides on healthy tissues other than the cancer tissue, and respective fatal outcomes have been reported previously (see, for example (Linette *et al.*, 2013)).

There are furthermore case reports of problems and potentially fatal serious adverse events upon administration of T cells transduced with certain non-optimised T-cell receptors (van den Berg *et al.*, 2015), and with receptors targeting Mage-A4 and Mage-A10 (Adaptimmune products ADP-A2M4 and ADP-A2M10).

Consequently, the precise knowledge about the identity of these off-target peptides as well as their safety relevance is of highest relevance for the proper development of cancer immunotherapies involving peptide:MHC binding polypeptides.

Current strategies to identify such peptides comprise searching of protein sequence databases for peptides with similarity to the target peptide. These approaches usually result in a large number of potentially ten-thousands of peptides, depending on the applied search parameters, all of which need to be tested in downstream assays for their potential to be bound by respective peptide:MHC binding polypeptides. A modification of these searches by integration of additional characteristics from the peptide:MHC binding polypeptide, such as mutational scanning data, can potentially

decrease the number of peptides that need to be tested in downstream assays, but still do not provide any information whether these peptides are also presented by MHC molecules in a physiological setting, and therefore still pose a considerable safety risk to the application *in vivo*.

In addition, all these approaches require certain assumptions on how similarity to the target sequence is evaluated or which amino acids are tolerated in certain positions of the peptide sequence based on mutational scanning data. Peptides that do not fulfill these assumptions, for example because their mode of binding to the MHC molecule or the peptide:MHC binding polypeptide varies from that of the target peptide sequence, cannot be identified by these approaches. This is particularly relevant for peptides with a different length than the target peptide (e.g. nonameric off-target peptide from a decamer target peptide) which may show a completely different amino acid sequence not related to the target sequence (Ekeruche-Makinde *et al.*, 2013).

In view of the above, there is a need in the art for efficient methods of identifying the target epitopes of TCRs with the aim to come as close to the situation *in vivo*. This furthermore requires the identification of "off-target" epitopes that are cross-reactive with the TCRs (or TCR-like molecules) under analysis, so that therapeutics can be developed that are not only highly specific but that also do not target normal healthy tissue.

It is therefore highly desirable to provide an alternative and more direct method for the identification of relevant MHC binding peptides that are bound or recognized by a peptide:MHC binding polypeptide by avoiding the use of inaccurate prediction algorithms and cumbersome testing of potentially hundreds to thousands of potential off-target peptides. It is therefore an object of the present invention to provide such a method for the characterization of peptide:MHC binding polypeptides in order to identify the off-target peptides (i.e. MHC presented ligands) that can be bound by these molecules in a comprehensive and direct manner. Other objects and advantages of the present invention will become readily apparent for the person of skill when studying the following description as provided.

In a first aspect of the present invention, the object of the invention is solved by a method for characterizing the binding of a polypeptide molecule comprising at least

one defined peptide binding domain ("pMHC binding polypeptide") to a peptide ("target peptide") of a peptide:MHC complex, comprising

- a) providing a sample comprising at least one peptide:MHC complex to be analyzed,
- b) contacting said sample with said polypeptide molecule ("pMHC binding polypeptide") and allowing said at least one peptide binding domain of said polypeptide molecule to bind to said at least one peptide:MHC complex,
- c) isolating said at least one peptide:MHC complex bound to said at least one peptide binding domain, and
- d) identifying said peptide of said at least one peptide:MHC complex as isolated in step c), and thereby identifying the binding of said polypeptide molecule to said peptide of said at least one peptide:MHC complex.

In one embodiment, the amino acid sequence of said peptide binding domain is, or is derived from, a T-cell receptor (TCR), a T-cell receptor-like polypeptide, and/or an antibody or a mere binding domain thereof.

As used herein, the term "polypeptide molecule comprising at least one defined peptide binding domain" is used interchangeably with the term "pMHC binding polypeptide".

Such pMHC binding polypeptide is for example, an antibody, or a functional fragment or derivative thereof, as disclosed elsewhere herein, or a T-cell receptor (also called "TCR" herein), preferably a soluble T-cell receptor, or a functional fragment or derivative thereof, as disclosed elsewhere herein.

The term "defined peptide binding domain" within the binding polypeptide refers, for example, to the variable domains of the heavy and light chain of an antibody or the variable domains of the α - and β -subunit of a TCR.

The term "defined peptide binding domain" within the binding polypeptide may also refer to the at least one complementarity determining region (also called "CDR" herein) of at least one variable domain of an antibody or a TCR. As used herein, the term

“peptide” in a peptide:MHC complex is used interchangeably with the term “target peptide”.

The term “peptide:MHC complex” can be abbreviated as “pMHC complex”, or even “pMHC”.

According to one embodiment, in the claimed method, the polypeptide molecule is optionally attached to a matrix material, and/or wherein said polypeptide molecule further comprises at least one attachment site binding to or being attached to a matrix material.

It is important to mention that this embodiment only relates to the method per se, in which the polypeptide molecule is characterized, e.g., for cross reactivity. As discussed elsewhere herein, the polypeptide molecule per se can be soluble.

According to one embodiment, characterizing the binding of a polypeptide molecule to a peptide of a peptide:MHC complex comprises identifying the binding of the polypeptide molecule to said peptide of a peptide:MHC complex.

In another preferred aspect, the disclosure provides for methods for identifying an off-target peptide of an off-target peptide:MHC complex capable of binding to a defined T-cell receptor (TCR) and/or antibody peptide binding domain, including, for example:

- a) providing a sample, e.g. a cellular lysate, comprising an off-target peptide:MHC complex, wherein said sample does not necessarily contain the target peptide:MHC complex defining the peptide binding properties of said peptide binding domain,
- b) affinity purifying said sample, comprising contacting said sample with a polypeptide, optionally coupled or attached to a matrix material, wherein said polypeptide comprises at least one peptide binding domain binding to said target peptide:MHC complex, and wherein said polypeptide is a T-cell receptor (TCR) and/or an antibody that binds to the target peptide of said target peptide:MHC complex,
- c) isolating said off-target peptide:MHC complex bound to said at least one peptide binding domain; and

d) identifying said off-target peptide of said at least one off-target peptide:MHC complex as isolated in step c).

In the context of the present invention, said sample can be selected from any suitable natural or artificial sample comprising at least one peptide:MHC complex to be analyzed, such as a cellular lysate, or a sample comprising purified or enriched peptide:MHC complex(es). The composition/properties of the peptide:MHC complex(es) as well as the concentration/amount of the molecules can be known or unknown. One example of the sample is a library of peptide:MHC complexes, wherein the sequences of the peptides as bound are defined and/or similar in their length and amino acid sequences.

In one embodiment, the polypeptide molecule comprising said at least one peptide binding domain does not comprise a further binding specificity, i.e., it is monospecific for the peptide or peptide:MHC complex.

In different embodiments of the method according to the present invention, the polypeptide molecule comprising said at least one peptide binding domain is selected from bispecific, trispecific, tetraspecific or multispecific molecules.

The term "bispecific" in context of the method of the invention refers to a polypeptide molecule with at least two valences and binding specificities for two different antigens and thus, comprises two antigen binding sites. In accordance with the present invention one of those antigen binding sites is the defined peptide binding domain. The term "valence" refers to the number of binding sites of a polypeptide molecule, e.g. a bivalent polypeptide relates to a polypeptide that has two binding sites. It should be noted, that, the term valence refers to the number of binding sites, wherein those binding sites may bind to the same or different targets, i.e. a bivalent polypeptide molecule may be monospecific, i.e. binding one target, or bispecific, i.e. binding two different targets.

The peptide:MHC binding polypeptide in this case is, or is derived from, a T-cell receptor (TCR), a T-cell receptor-like polypeptide, and/or an antibody, or a fragment of these molecules that is able to bind or mediate binding of said molecule to specific/defined peptide:MHC complexes.

The at least one peptide binding domain can be a molecule or is derived from a molecule selected from an antibody, a simultaneous multiple interaction T-cell engaging (SMITE) bispecific, a bispecific T-cell engager (BiTE), an scFV, a diabody, a dual-affinity re-targeting antibody (DART), a tandem antibody (TandAb), a soluble TCR, an scTCR, a mutated TCR, for example comprising S-bridges, a truncated TCR, and a bispecific T-cell receptor (TCR)-antibody fusion molecule.

In one example, the at least one peptide binding domain can be an antibody or is derived from an antibody.

In the context of the present invention, a “defined” or “defining” a peptide:MHC binding polypeptide shall mean a polypeptide (constituting or comprising a binding domain) that binds to a selected (“targeted”) MHC peptide in the context of MHC. In a preferred embodiment, said binding of the peptide:MHC binding polypeptide to the selected MHC peptide takes place with the highest affinity and/or selectivity, when compared to other (known) MHC peptides.

Examples as used in the context of the present invention are the defined peptide:MHC binding polypeptide of example 1, 3, 4 and 5 showing an affinity to the PRAME-004 peptide SLLQHLIGL (SEQ ID NO: 1), and the defined peptide:MHC binding polypeptide of example 2, which shows an enhanced affinity to the MAGEA4/A8 derived peptide having the sequence KVLEHVVRV (SEQ ID NO. 24), herein also referred to as MAGEA4/8 peptide.

A binding of the defined peptide:MHC binding polypeptide may comprise binding to the targeted peptide in the context of MHC or binding to both the targeted peptide and the MHC polypeptide.

While some methods for screening TCRs and TCR-like molecules against peptide ligands have been described in recent years, to date such methods have met with limited success. For example, Birnbaum *et al.* developed a peptide:MHC (“pMHC”) yeast display library of $\sim 2.1 \times 10^8$ antigen minigenes (Birnbaum *et al.*, 2014). Using Birnbaum's system, cells that bound to soluble TCRs were purified with magnetic beads and then subject to high throughput sequencing. After four rounds of selection, hundreds of peptides were identified that were cross-reactive with five distinct mouse

TCRs. However, the original epitopes, to which the TCR was known to bind, were not detected.

The invention preferably uses peptide:MHC binding polypeptides to isolate or enrich specific peptide:MHC molecules from a mixture of organic molecules, like peptides and proteins (e.g. protein lysates produced from tissues or cell lines, libraries of recombinantly produced peptide:MHC molecules) and, e.g. subsequent mass spectrometry to analyze the isolated peptide:MHC molecules and to identify the sequence of bound peptides.

The present method has a number of advantages compared with methods of the state of the art, which can be designated as "off-target prediction approaches".

For the actual identification of off-target binders, in the prediction approaches a binder motif is used to predict a large list of peptides, followed by laborious *in vitro* testings. In the present invention, a direct identification from samples, like cellular lysates, is possible. This does not require cumbersome testing of predicted peptides in *in vitro* assays. Furthermore, it was found that the method is highly sensitive, so that even weakly cross-recognized peptides can be identified.

In the prediction approaches, the identification of the source of unknown off-target recognitions is not possible. In contrast, with the present invention, this is possible by creating a sample, e.g. lysate, and using said lysate in affinity chromatography experiments with the peptide:MHC binding polypeptide.

In the prediction approaches, the binder motif generation is inferred from positional scanning data, whereas in the present invention it is inferred from identified off-targets, and further considers multiple substitutions at different positions in the peptide(s) amino acid sequence, without prior *in vitro* testing necessary.

Additionally, the presented method is not limited to the analysis of a specific HLA allotype (e.g. HLA-A*02:01) and can be also used to identify off-targets presented by different HLA allotypes. These alloreactivities are difficult to assess with the currently available standard approaches and are not amenable to prediction methods.

Finally, with respect to length variants and modifications of the peptide sequences, in the prediction assay no prediction of length variants or naturally occurring peptide modifications is possible, the present assay allows a direct identification of all length variants and naturally occurring peptide modifications such as, for example, phosphorylations, and glycosylations.

In one embodiment, the present method specifically and reliably identifies peptides bound by peptide:MHC binding polypeptides, comprising an enrichment or isolation of peptide:MHC molecules recognized by said peptide:MHC binding polypeptides from mixtures of peptide:MHC molecules derived from biological samples or biotechnological production, and the identification of the enriched/isolated peptide:MHC molecules, e.g. by mass spectrometry, and a subsequent testing of identified peptides for their potential to be bound by the same peptide:MHC binding polypeptides *in vitro*.

Preferred is a method according to the present invention, wherein said polypeptide molecule comprising said at least one peptide binding domain binding to a peptide:MHC complex is selected from bispecific, trispecific, tetraspecific or multispecific molecules.

Further preferred is a method according to the present invention, wherein said polypeptide molecule comprising said at least one peptide binding domain is a bispecific molecule comprising a peptide binding domain that is derived from a T-cell receptor (TCR).

In another aspect, the polypeptide containing said at least one peptide binding domain binding to the target peptide:MHC complex may be bispecific molecules comprising a binding domain that is derived from a T-cell receptor (TCR).

In other aspects, methods as described herein may further include selecting a set of polypeptides that differ in their binding properties.

Advantageously, the presented method as described in the independent claim does not necessarily require determination of binding motifs by substitution of single amino acids at individual positions of the target peptide, subsequent testing the recognition of

these peptide variants in functional assays, then predicting potential off-targets based on said binding motifs employing human proteome databases and subsequently testing these (potentially) large numbers of peptides in *in vitro* assays, as explained above.

In another aspect, methods as described herein preferably reduce the number of identified off-target peptides by at least 2-fold, preferably at least 5-fold, and most preferred at least 10-fold, as compared to prediction-based approaches by focusing on the relevant peptides which are presented on a biological sample by the appropriate HLA allotype as explained above and in the comparative examples herein.

In another aspect, the at least one binding domain may contain a detectable marker or a label.

According to one embodiment of the invention, the cross-reactivity for said peptide:MHC binding domain is further explored by at least one cytotoxicity experiment. This can be done by e.g. T2 cells loaded with the respective peptides, which are co-incubated with human CD8+ T cells (50,000 cells/well) in the presence of indicated concentrations of the PRAME-004 specific peptide:MHC binding polypeptide (see Figure 4 and respective description elsewhere herein).

The present invention may also relate to a kit containing materials for performing the method including a) a material, such as a matrix material described herein, and b) a polypeptide containing at least one binding domain binding to a target peptide:MHC complex.

The present invention further relates to a polypeptide molecule comprising at least one defined peptide binding domain ("pMHC binding polypeptide"), wherein the binding of which to a peptide ("target peptide") of a peptide:MHC complex was characterized with a method according to any one of claims 1 – 29.

Such polypeptide molecule can be an antibody or a T cell receptor, or a modified format, as discussed in the context of the method elsewhere herein.

Such polypeptide molecule can be mono-, bi- or multispecific, as discussed in the context of the method elsewhere herein.

Such polypeptide molecule can be soluble, as discussed in the context of the method elsewhere herein.

Such polypeptide molecule can have an affinity to the peptide ("target peptide") of the peptide:MHC complex, or to the peptide:MHC complex of $KD < 500$ nM.

All preferred embodiments of the polypeptide molecule discussed in the context of the method apply here as well and will not be repeated for the sake of economy.

The present invention further relates to a method of manufacturing a cell population expressing a target specific antigen recognizing construct, including providing a host cell, providing the target specific antigen recognizing construct comprising a coding sequence encoding the second polypeptide, introducing into said host cell said target specific antigen recognizing construct, and expressing said target specific antigen recognizing construct by the host cell. In another aspect, the expressing may include presenting the antigen recognizing construct on the cell surface.

In another aspect, the target specific antigen recognizing construct may be an expression construct containing a promoter sequence operably linked to said coding sequence. In yet another aspect, the target specific antigen recognizing construct may be of mammalian origin, optionally of human origin. The target specific antigen recognizing construct may further be a modified TCR, wherein said modification comprises addition of a functional domain comprising a label, or an alternative domain comprising a membrane anchor domain.

In another aspect, the target specific antigen recognizing construct may be an alpha/beta TCR, gamma/delta TCR, or a single chain TCR (scTCR). In another aspect, the target specific antigen recognizing construct may be introduced into said suitable host cell by retroviral transduction. In yet another aspect, the methods described herein may further include isolating and purifying the target specific antigen recognizing construct from the host cell and, optionally, reconstitution of the target specific antigen recognizing construct in a T-cell.

In an aspect, the present invention relates to a cell, in particular T-cell, population produced by methods of the present invention.

The target specific antigen recognizing constructs, and cells, in particular T-cells, produced by methods of the present invention are expected to show improved properties in a treatment against diseases, wherein said treatment comprises immunotherapy. Because of the fact that the present method resembles a situation closer/more similar to the one *in vivo*, less off-target effects and side effects will be encountered. This has benefits for medical and clinical uses; in the clinical trial testing a modified anti-MAGE-A3 TCR, derived from immunization of HLA-A*02:01 transgenic mice, two out of nine cancer patients developed fatal on-target neurological toxicity, due to recognition of a peptide derived from the same gene family that is expressed in the brain (Morgan *et al.*, 2013).

In another trial, where an affinity-enhanced anti-MAGE-A3 TCR was tested in myeloma and melanoma patients, two patients died of off-target toxicity caused by recognition of a completely different peptide, resulting in severe myocardial damage (Linette *et al.*, 2013; Raman *et al.*, 2016). These clinical cases show how difficult it is to predict the exact specificity and the resulting effects of TCRs that did not undergo optimal thymic selection. It is crucial to develop strategies to extensively validate the exact specificity of TCRs, particularly because TCR-engineered T-cells are highly sensitive (Jahn *et al.*, 2016; Stone and Kranz, 2013).

In another aspect, the present invention relates to a method of treating a patient who has cancer, including administering to the patient a composition comprising the cell population as above, in which the cancer is selected from non-small cell lung cancer (NSCLC), small cell lung cancer (SCLC), renal cell cancer, brain cancer, gastric cancer, colorectal cancer, hepatocellular cancer (HCC), pancreatic cancer, prostate cancer, leukemia, breast cancer, Merkel cell carcinoma, melanoma, ovarian cancer (OC), urinary bladder cancer, uterine cancer, gallbladder and bile duct cancer, esophageal cancer (OSCAR) acute myeloid leukemia, cholangiocellular carcinoma, chronic lymphocytic leukemia, glioblastoma, head and neck squamous cell carcinoma, non-hodgkin lymphoma, and endometrial cancer.

In another aspect, the host cell may be obtained from the patient. In another aspect, the host cell may be obtained from a healthy donor. In another aspect, the host cell may be a CD8+ T cell.

In another aspect, the MHC molecule may be an MHC class I molecule. In another aspect, the MHC molecule may be an HLA-A*02 molecule.

Preferred is a method according to the present invention, wherein said polypeptide molecule comprises at least one second binding domain that is selected from a domain binding to a cell surface molecule known to induce the activation of immune cells, or is selected from the group consisting of immune response-related molecules, preferably a second binding domain binding to CD3 or one of its chains, like the CD3 γ , CD3 δ , and CD3 ϵ chains, CD4, CD7, CD8, CD10, CD11b, CD11c, CD14, CD16, CD18, CD22, CD25, CD28, CD32a, CD32b, CD33, CD41, CD41b, CD42a, CD42b, CD44, CD45RA, CD49, CD55, CD56, CD61, CD64, CD68, CD94, CD90, CD117, CD123, CD125, CD134, CD137, CD152, CD163, CD193, CD203c, CD235a, CD278, CD279, CD287, Nkp46, NKG2D, GITR, Fc ϵ RI, TCR α/β , TCR γ/δ , and HLA-DR.

According to one embodiment, the polypeptide molecule comprising said at least one peptide binding domain is a soluble molecule. Such soluble polypeptide molecules, e.g., antibodies, T cell receptors or derivatives therefrom as discussed elsewhere herein, are a preferred therapeutic modality.

According to one embodiment, the polypeptide molecule has an affinity to the peptide ("target peptide") of the peptide:MHC complex, or to the peptide:MHC complex of $K_D < 100$ nM.

K_D is the dissociation constant, i.e., the equilibrium constant that measures the propensity of a target-binder complex to separate (dissociate) reversibly. One way to determine K_D is with surface plasmon resonance (SPR).

Preferably, the polypeptide molecule has an affinity to the peptide ("target peptide") of the peptide:MHC complex, or to the peptide:MHC complex of $K_D < 100$ nM; < 50 nM; < 20 nM; < 10 nM; < 5 nM; < 2 nM; < 1 nM; < 500 pM; < 400 pM; < 300 pM; < 200 pM; < 100 pM; < 50 pM; < 20 pM; < 10 pM; or < 1 pM. Most preferred in the range of 10 nM to 500 pM.

In another aspect of the method according to the present invention, said attachment site binding to or being attached to said matrix material is positioned in the at least one

binding domain, the at least one second domain or is a separate attachment group, and does not essentially interfere with the binding of said molecule.

Preferred is a method according to the present invention, wherein the isolation of peptide:MHC molecules is performed by affinity chromatography or immunoprecipitation after coupling of said peptide:MHC binding polypeptide to a solid matrix.

Preferred is the method according to the present invention, wherein said matrix material is selected from Sepharose® or agarose. Nevertheless, the binding can also be performed in solution without a solid matrix (as a batch), and the complexes can be suitably isolated, e.g. using antibodies, precipitation, filtration or the like.

In another aspect, said method may further comprise contacting said sample in step b) and/or c) with one or more other binding domain molecules, such as, for example, broad specific TCRs and/or antibodies.

Such broad specific binding domain molecule is preferably a binding domain molecule that has binding specificity against one or more MHC molecules alone, irrespective of the peptide bound thereto. Such a binding molecule is called MHC pan-specific herein.

In one embodiment of this method,

- (i) a first fraction of the sample is contacted with the pMHC binding polypeptide, and
- (ii) a second fraction of the sample is contacted with another pMHC binding polypeptide or the MHC pan-specific binding molecule

and wherein further, after isolating the peptide:MHC complexes bound to the pMHC binding polypeptide and the other pMHC binding polypeptide or the MHC pan-specific binding molecule, the isolation efficiencies achieved in the two fractions are compared.

This can for example be accomplished by combining mass spectrometry data sets of the respective isolations performed in parallel with the different binding molecules. The quantitative data for each identified peptide allow for a comparison of the isolation

efficiency of peptides between different pMHC binding polypeptides or pMHC binding polypeptides and MHC pan-specific binding molecules, as illustrated in example 2.

This enables a correction for the natural abundance and any unspecific binding of each peptide in the analyzed biological sample by calculation of the fold change/enrichment factor of the peptide:MHC binding polypeptides and the broad specific TCRs or antibodies.

In one embodiment of this method, in step b) or c), a large excess of pMHC binding polypeptides and optionally MHC pan-specific binding molecules is applied. In such way, the recovery of each peptide not limited by the number of available binding sites.

In one further embodiment of this method, the binding affinities of the binders (binding polypeptides or binding molecules) used in the first and second fraction to the different peptides in the sample are determined based on the comparison of the isolation efficiencies, as e.g. illustrated in example 3.

The inventors have surprisingly shown that the ratio of recoveries (=contacting and isolating) of individual peptides between the MHC pan-specific binding molecule (e.g., antibody BB7.2 and the pMHC binding polypeptide) correlates with the binding affinity of the pMHC binding polypeptide to these peptides (targets and off-targets), as seen by the comparative experiments performed via bio-layer interferometry (in this case Octet measurements; see Fig. 12, the latter expressed as greyscale coloring and Fig. 13).

Preferably, the isolation thus may include contacting the peptide:MHC complexes with a pan-specific antibody against the MHC molecule. In another aspect, the antibody may be at least one selected from W6/32, B1.23.2, BB7.2, GAP-A3, Spv-L3, Tü39, L243, or IVD-12. Further preferred is a method comprising an isolation of said peptide:MHC molecule in parallel to a broad-specific (or multi-specific or unspecific) peptide:MHC binding polypeptide, which is directed against the MHC component of the peptide:MHC molecule, and therefore allows an isolation of all peptides bound to a certain MHC allotype, irrespective of the nature of the peptide sequence of the bound peptide (Figure 1). As mentioned, such a broad-specific peptide:MHC binding polypeptide can be, but is not limited to, an antibody, such as the HLA-A*02 pan-

specific antibody BB7.2 or the pan-HLA-A,B,C specific antibody W6/32. Additional mouse hybridoma derived antibodies and associated pan-specificities for human MHC molecules are listed in Table 1.

Table 1: Overview of commonly applied mouse hybridoma derived antibodies specific for the indicated HLA allotype

Clone	Specificity	Reference
W6/32	HLA-A,B,C	(Barnstable <i>et al.</i> , 1978)
B1.23.2	HLA-B,C	(Rebai <i>et al.</i> , 1983)
BB7.2	HLA-A*02	(Parham and Brodsky, 1981)
GAP-A3	HLA-A*03	(Berger <i>et al.</i> , 1982)
Spv-L3	HLA-DQ	(Spits <i>et al.</i> , 1983)
Tü39	HLA-DR, DP, DQ	(Maeda and Hirata, 1984)
L243	HLA-DR	(Lampson and Levy, 1980)
IVD-12	HLA-DQ	(Kolstad <i>et al.</i> , 1987)

Combining mass spectrometry data sets of respective isolations performed in parallel yields quantitative data for each identified peptide allowing for a comparison of the isolation efficiency of peptides between different peptide:MHC binding polypeptides or peptide:MHC binding polypeptides and broad specific TCRs or antibodies as illustrated in example 2. This enables a correction for the natural abundance and any unspecific binding of each peptide in the analyzed biological sample by calculation of the fold change/enrichment factor of the peptide:MHC binding polypeptides and the broad specific TCRs or antibodies.

Further preferred is a method comprising the isolation of peptides in such a way that it contains the additional step for the isolation and depletion of non-specifically bound peptides from the mixture of peptide:MHC molecules which bind for example to surfaces of the material utilized in the isolation procedure. Such a step may be constituted by another affinity-chromatography column which contains the solid matrix (e.g. Sepharose®, agarose) but without the peptide:MHC binding polypeptide and onto

which the mixture of peptide:MHC molecules are applied before the isolation step with the respective peptide:MHC binding polypeptide (see Figure 1). Subsequent extraction (elution) of these non-specifically bound peptides and identification by mass spectrometry allows for their identification and enables the exclusion of these peptides from further analysis.

Preferred is a method according to the present invention, wherein said identifying in step d) comprises a method selected from mass spectrometry and peptide sequencing.

According to one embodiment of the method according to the present invention, characterizing the binding of a polypeptide molecule to a peptide of a peptide:MHC complex comprises identifying a superordinate binding motif for said peptide:MHC binding domain.

Details on how this is accomplished are provided in Figure 7 and respective description, and in example 1.

According to one embodiment of the method according to the present invention, characterizing the binding of a polypeptide molecule to a peptide of a peptide:MHC complex comprises

- a) Identifying a position-based and/or superordinate binding motif of the polypeptide molecule for said peptide:MHC binding domain, and/or
- b) Identifying off-target peptides of the polypeptide molecule for peptide:MHC binding domains using the peptides identified by the presented application.

Details on how this is accomplished are provided in Figure 7 and respective description, and in example 1.

According to one embodiment of the method according to the present invention, characterizing the binding of a polypeptide molecule to a peptide of a peptide:MHC complex comprises exploring and/or identifying the cross-reactivity of said peptide for other peptide:MHC binding domains.

As used herein, the term “superordinate binding motif” is used to describe which amino acids are tolerated in which position of the peptide sequence while still maintaining binding of the peptide:MHC binding polypeptide

As used herein, the term “position-based binding motif” is used to describe which positions in the amino acid sequence of the bound peptide are of relevance for the interaction with the peptide:MHC binding polypeptide

As used herein, the term “off-target peptide sequences” relates to peptide sequences which are capable to bind to a given peptide:MHC binding polypeptide although their sequences deviate from the sequence of the peptide that the peptide:MHC binding polypeptide was originally designed or selected for. Hence, there is a risk that, if such off-target peptide sequences are e.g. displayed on healthy, non-cancerous tissue, that the peptide:MHC binding polypeptide will exhibit cytotoxic activity against healthy tissue (“off-tumor/off-target” toxicity).

Details on how this is accomplished are provided in Figure 7 and Figures 12 – 15, and respective description, and in the presented examples. Further preferred is a method according to the present invention, wherein the identified peptide sequences are used to infer information on the binding characteristics of the peptide:MHC binding polypeptide. Such information can for example be used to generate binding motifs of the peptide:MHC binding polypeptide. These binding motifs are commonly used to describe which positions in the amino acid sequence of the bound peptide are of relevance for the interaction with the peptide:MHC binding polypeptide (“position-based binding motif”) and moreover which amino acids are tolerated in which position of the peptide sequence while still maintaining binding of the peptide:MHC binding polypeptide (“superordinate binding motif”). Analysis of the amino acid sequences of the identified peptides and the presence in selected positions of the peptide sequence facilitates the generation of these binding motifs. The latter can be further used to perform predictions of safety relevant off-targets for example from protein sequence databases using the information contained in the binding motif as search criteria.

Preferred is a method according to the present invention wherein the sequence identification is achieved by mass spectrometry in such a way that all peptide

sequences contained after the isolation step are identified comprehensively, quantitatively, at the highest sensitivity and with high confidence in the identified sequence match.

In the context of the present invention, one of the techniques and methods listed as follows may be preferably applied:

- a) The specific use or a combination of any number of different mass spectrometry machines and mass spectrometry fragmentation techniques (e.g., collision-induced dissociation (CID), surface-induced dissociation (SID), electron-capture dissociation (ECD), Higher-energy C-trap dissociation (HCD), electron-transfer dissociation (ETD), negative electron-transfer dissociation (NETD), electron-detachment dissociation (EDD), infrared multiphoton dissociation (IRMPD), blackbody infrared radiative dissociation (BIRD), electron-transfer/higher-energy collision dissociation (ETHCD), ultraviolet photodissociation (UVPD), electron-transfer and collision-induced dissociation (ETCID)) or activation energies to allow for better sequence coverage of the peptide tandem MS (MS/MS) spectra.
- b) Mass spectrometry experiments in data-dependent (DDA) as well as data-independent acquisition mode (DIA). This may further include specific acquisition strategies employing a predefined list of peptide sequences (targeted analysis) or other methods such as sequential window acquisition of all theoretical mass spectra (SWATH) analysis.
- c) A pre-separation of peptide mixtures for example by HPLC (e.g. nano-UHPLC run with a gradient of acetonitrile in water) before or directly coupled to the mass spectrometry analysis.
- d) Replicate measurements of the same peptide mixture in order to allow a more robust statistical evaluation.
- e) A search of MS/MS spectra using different search engines (e.g. MASCOT, Sequest, Andromeda, Comet, XTandem, MS-GF+) or software tools using one of these search engines as well as *de novo* sequence identification algorithms.

f) A search further employing computational tools for MS/MS spectra prediction as for example pDeep (Zhou *et al.*, 2017)

g) A search of MS/MS spectra against different protein sequence databases (e.g. UniProtKB, IPI) as well as custom sequence databases generated for specific purposes (e.g. protein sequences translated from mRNA sequences).

h) Mass spectrometry measurements of synthetic versions of the peptides in question to confirm their identity by comparing peptide specific characteristics such as their MS/MS spectra and their retention time for example on an HPLC column.

i) A search of MS/MS spectra involving a database of previously identified MS/MS spectra (e.g. spectral library or spectral archive). This database can be used for the peptide identification process by comparing newly recorded MS/MS spectra to respective MS/MS spectra of already identified peptide sequences.

j) Any database as described in i) that further stores information on the experimental or predicted retention time (RT) of already identified peptides to aid the identification process

j) A quantitative assessment of peptide signal areas on MS or MS/MS level for example by extraction and integration of MS features using appropriate algorithms (e.g. SuperHirn) (Mueller *et al.*, 2007).

Preferred is a method according to the present invention, further comprising the step of identifying the presentation of said peptide or peptides on cancerous and/or non-cancerous cells or tissues.

Such identification comprises both the identification of the binding motif as well as of the off-target peptide sequences identified as disclosed elsewhere herein in cancerous or non-cancerous tissue.

For this purpose, the peptide:MHC complexes displayed on cells from samples of cancerous and non-cancerous tissues can be analyzed.

One way to do this, is to isolate the entirety of peptide:MHC complexes (the “immunopeptidome”, e.g. by immunoaffinity capture. The displayed peptides can then be dissociated from the MHC by e.g. acid denaturation. The peptide cargo is then extracted and separated into fractions by HPLC, and the peptide sequences in these fractions are identified using LC–MS/MS. In such way, several thousand peptides can be identified from a wide variety of cell types, including cancerous and infected cells and those from tissues.

The peptide sequences thus obtained are then entered into a database, with annotation whether or not the respective peptides were found on cancerous and/or non-cancerous cells or tissues. Such database can then be scanned through for the presentation of peptides corresponding to binding motifs or superordinate motifs identified as described elsewhere herein.

Methods and protocols to accomplish this are disclosed, *inter alia*, in WO2005076009, WO2011128448 and WO2016107740, the contents of which are incorporated herein by reference, and all of which are assigned to the present applicant.

Thus, methods of the invention may further comprise evaluating identified peptides for their relevance to cause off-target toxicity by further investigation applying any of the following techniques, but not limited to:

- i) Analysis of gene expression profiles of the source gene of said off-target peptides on different normal or cancer tissues as well as cell lines;
- ii) Analysis of peptide presentation profiles of said off-targets on different normal or cancer tissues as well as cell lines; and
- iii) Analysis of peptide copy numbers per cell of said off-targets on different normal or cancer tissues as well as cell lines

In another important aspect of the method according to the present invention, said method further comprises the step of a computational analysis of said identification and/or off-target binding, in particular in preparation for further modification(s) of the

peptide:MHC binding polypeptides as disclosed herein. Respective programs are well known to the person of skill.

According to the present invention, if off-target peptides are identified by this method, peptide:MHC binding polypeptides can then be suitably altered in order to reduce binding to these off-target peptides. Such alterations comprise modifying the amino acid sequence of the peptide:MHC binding polypeptides, in particular in rounds of maturation, in order to improve the specificity of the peptide:MHC binding polypeptide. Thereby specificity of the newly generated molecules can be greatly improved as exemplary shown by a reduced killing of the target negative cell line T98G in Figure 2. Methods of a respective maturation of peptide:MHC binding polypeptides are known to the person of skill, and in particular include changes in the peptide:MHC binding polypeptide, such as the six complementary determining regions (CDRs) of the TCR. Similarly, the CDRs of an antibody can be modified accordingly (See, for example (Smith *et al.*, 2014; Stewart-Jones *et al.*, 2009); US 2014-0065111A1; WO 2017/174823A1; WO 2016/199141; and WO 2012/013913).

Preferred is then a method according to the present invention wherein said biological sample from which a peptide:MHC mixture is derived, is selected from one to several cancer cell lines, which can be analyzed separately or combined at the isolation step of said method. Figure 3 illustrates based on XPRESIDENT® data how a combination of several cell lines will increase the coverage of the normal tissue peptide space, which can be addressed directly with this method by selecting several cell lines for sample generation. A skilled person may further want to combine several cell lines with strong MHC expression or modify such cell lines in a way for example by transfection or viral transduction with a gene of interest or treatment with a substance or chemical compound (e.g. Interferon gamma) to increase or modify the MHC expression in such a way as to increase or modify the number of different peptides presented by MHC molecules in such cell lines. These genes could be but are not limited to specific MHC class I or class II genes (e.g. HLA-A*02, HLA-DRB3), genes involved in the MHC peptide processing and presentation (e.g. TAP1/2, LMP7), or transcription factors capable of inducing or modifying gene expression of cells (e.g. AIRE).

Further preferred is a method according to the present invention, wherein said biological sample from which the peptide:MHC mixture is derived is selected from one to several primary normal tissue samples or blood of healthy donors, as well as tumor tissues from cancer patients or infected tissue. Of special relevance are normal tissues or isolated cells derived from these tissues or specific body compartments with a high risk of fatal adverse events in the case of cross-recognition of a peptide on these tissues. Such normal tissues or cells isolated therefrom may be but are not restricted to: brain tissue, heart tissue, blood, lung tissue, spinal cord, nerve tissue or liver tissue.

All biological samples can be fresh or processed (e.g. frozen or prepared), as long as they are still suitable for the method according to the invention. In an aspect, biological samples may include tissues, organs, cells, proteins, or membrane extracts of cells, blood, or biological fluids, such as blood, serum, mucus, urine, ascites fluid or brain fluid obtained from a subject.

According to one embodiment of the method according to the invention, said method further comprises adding to said sample in step a) at least

- one peptide having a known sequence and/or
- one defined and/or preselected peptide:MHC complex the peptide of which has a known sequence,

preferably in a predetermined amount ("spiking").

According to one embodiment, the sequence of said peptide is altered or mutated relative to the sequence of the peptide ("target peptide") of a peptide:MHC complex to which the polypeptide molecule comprising at least one defined peptide binding domain ("pMHC binding polypeptide") binds.

According to one embodiment, a series of mutants of the peptide ("target peptide") of a peptide:MHC complex to which the polypeptide molecule comprising at least one defined peptide binding domain ("pMHC binding polypeptide") binds is created and added to the sample in step a). Therein, each mutant has, over its entire length or over at least a subfraction thereof, the amino acid residue at one position substituted for an alternative amino acid.

According to one embodiment, each mutant has, over its entire length or over at least a subfraction thereof, the amino acid residue at one position substituted for alanine or glycine, as e.g. illustrated in example 5.

As said, the mutant peptides thus obtained can be spiked into a sample (e.g. pHLA library or cellular lysate) in step a) of the presented method. This approach allows to even better characterize the superordinate binding motif and identify potential off-targets.

According to one embodiment, the synthetic peptides utilized for spiking are derived from naturally occurring peptides, e.g. based on mass spectrometric or genomic data and selected based on defined criteria such as chemical similarity to the target peptide, as illustrated in example 4.

In one embodiment, using methods as discussed herein, it is not necessary to include the target peptide sequence into the sample for the evaluation of these peptides. Further, in said embodiment, it is not required to calculate any ratios in reference to the target peptide sequence or other peptide sequences. The binding strength of the mutated peptides is only determined *via* their recoveries in binder vs. BB7.2 preparations.

According to one embodiment, the anchoring positions of the peptide ("target peptide") of a peptide:MHC complex to which the polypeptide molecule comprising at least one defined peptide binding domain ("pMHC binding polypeptide") are not altered/mutated.

Note, in this context, that the so-called "anchoring residues" (abbreviated "AR" herein), mediate binding of the peptide to the peptide binding groove in the MHC. In HLA-A*02:01, these anchoring residues are most often leucine in P2, and leucine or valine in P9, and play only a minor role in the binding reaction between the binding polypeptide and the peptide:MHC complex.

In a further preferred aspect of the method according to the present invention, mixtures of peptide:MHC molecules can be generated artificially by biotechnological production. The latter might be achieved but is not restricted to transformation and expression of an MHC molecule in a prokaryotic (e.g. *E. coli*) system. Such MHC molecules might

be further modified to increase their solubility for example by replacing or altering the transmembrane part of such molecules. Peptide loading and reconstitution of such molecules could be achieved but is not limited to refolding of inclusion bodies of MHC components (e.g. protein heavy chain(s) and beta-2 microglobulin) in the presence of the peptide of interest as well as additional chemical substances (glutathione, arginine, etc.) which promote reconstitution of the desired molecule.

A skilled person might want to combine several of these artificially produced peptide:MHC molecules loaded with different peptides to generate a library of tens, to hundreds to thousands to ten-thousands or hundred-thousands of different peptide:MHC molecules, which can be tested with a peptide:MHC binding polypeptide according to the presented method. The skilled person might further want to spike these mixtures of artificially generated peptide:MHC molecules into another mixture of peptide:MHC molecules from a biological source. In another aspect a skilled person could reconstitute these artificially generated peptide:MHC molecules with a peptide containing one or several heavy stable isotope labels, such as but not limited to ^{13}C , ^{15}N , or ^2H . These peptide:MHC molecules could then be spiked into another mixture of peptide:MHC molecules for example from a biologic sample to gain additional information of the peptide of interest containing the heavy label and provide quantitative mass spectrometry data on the efficiency of the peptide isolation using the presented method.

In a further preferred aspect of the method according to the present invention, the method is combined with a computational analysis of said identification and/or off target binding.

This additional information might be constituted by but is not restricted to the calculation of scores for peptide similarity and peptide binding to MHC molecules such as PMBEC, Blosum62, Pam250, NetMHC, NetMHCpan, SYFPEITHI.

The term "similarity" of two peptides takes into consideration the relatedness of two amino acids at a given position (see, for example below Table 1a). Similar amino acid sequences e.g. an off-target peptide sequence similar to a target peptide sequence can be retrieved from protein databases using the BLAST algorithm, which performs a statistical analysis of the similarity between two sequences (see, e.g. (Karlin and

Altschul, 1993)). Preferred settings for such an alignment are: wordlength of 3, and expectation (E) of 10, and the use of the BLOSUM62 or PMBEC scoring matrix (Kim *et al.*, 2009), preferably the PMBEC scoring matrix is used in the determination of the similarity. These matrices quantify amino acid similarity based for example on evolutionary or functional similarity between amino acids, which correlate well with the similarity according to physicochemical parameters. For each substitution of an amino acid in a given target peptide sequence a score (decimal value) can be calculated by using these matrices, which indicates the similarity of the amino acid in the target peptide sequence with the substituted amino acid in the off-target peptide sequence. Multiple substitutions can be considered by summing up the effect (scores) of individual substitutions in the target peptide sequence. By definition, the maximum score which can be achieved for an off-target peptide is provided by the unsubstituted target peptide sequence, whereas any substitution leading to an off-target peptide will be penalized in the scoring matrix and ultimately lead to a lower score of an off-target peptide. This maximum score is however dependent on the length and amino acid sequence of the target peptide (i.e. different target peptide sequences will have different maximum scores). Typically, a longer amino acid sequences results in a higher score. However, the score of a target peptide depends on the score allotted to the amino acids it consists of. In order to be able to calculate and compare the similarity of an off-target peptide in reference to a target peptide without considering the difference of maximum scores of distinct target peptide sequences the respective decimal values are converted, which are the result of calculating the similarity of an off-target peptide in reference to a target peptide, into a percentage score wherein the maximum score of a target peptide sequence will therefore, always be 100%.

Table 1a: Amino acids and conservative and semi-conservative substitutions, respectively. Changing from A, F, H, I, L, M, P, V, W or Y to C is semi-conservative if the new cysteine remains as a free thiol. Furthermore, the skilled person will appreciate that glycines at sterically demanding positions should not be substituted.

Amino acid	Conservative substitution	Semi-conservative substitution
A	G; S; T N; V; C	
C	A; V; L	M; I; F; G

D	E; N; Q	A; S; T; K; R; H
E	D; Q; N	A; S; T; K; R;
F	W; Y; L; M; H	I; V; A
G	A	S; N; T; D; E; N; Q
H	Y; F; K; R	L; M; A
I	V; L; M; A	F; Y; W; G
K	R; H	D; E; N; Q; S; T; A
L	M; I; V; A	F; Y; W; H; C
M	L; I; V; A	F; Y; W; C;
N	Q	D; E; S; T; A; G; K; R
P	V; I	L; A; M; W; Y; S; T; C; F
Q	N	D; E; A; S; T; L; M; K; R
R	K; H	N; Q; S; T; D; E; A
S	A; T; G; N	D; E; R; K
T	A; S; G; N; V D; E; R; K; I	
V	A; L; I M; T; C; N	
W	F; Y; H	L; M; I; V; C
Y	F; W; H	L; M; I; V; C

The present invention will now be described with reference to the following examples, nevertheless, without being limited thereto. For the purposes of the present invention, all references as cited herein are incorporated by reference in their entireties.

Figure legends

Figure 1: Schematic overview of the experimental approach according to the present invention. A sample containing peptide:MHC molecules is provided for example by generating a lysate of peptide:MHC expressing cells derived from a tissue or cell line. Alternatively, the sample can be modified by addition of or constituted by mixtures of artificially produced peptide:MHC molecules. Specific peptide:MHC molecules are isolated from this sample for example by contacting it to a matrix, which has a peptide:MHC binding polypeptide attached to it. Glycine columns which do not contain

an attached polypeptide maybe used to deplete the sample from non-specifically isolated peptide:MHC molecules, which interact non-specifically with the matrix. Mass spectrometry can be used to identify the MHC bound peptides, which have been isolated from the sample to identify and sequence the peptide space recognized by the peptide:MHC binding polypeptide. Thereby previously unknown off-target risks can be elucidated, without the need of prediction tools. In a modification of this approach the same peptide:MHC molecule containing sample can be split and in parallel be subjected to a second affinity chromatography with HLA broad specific antibodies or TCRs that bind to these molecules irrespective of the bound peptide species. In the illustrated example the HLA-A*02 pan specific antibody BB7.2 is employed to isolate all peptides presented by HLA-A*02 irrespective of the peptide sequence, herein referred to as the HLA-A*02 immunopeptidome. The abundance of interesting (e.g. off-target) peptides in both isolations can be used to assess the binding affinity of the peptide:MHC binding polypeptide for said peptide in comparison to the employed broad specific antibody or TCR, in this example BB7.2.

Figure 2: Cytotoxicity experiments showing killing of a target positive (U2OS) and a target negative (T98G) cell line with two variants of the PRAME-004 peptide:MHC binding polypeptide (black rectangles: original variant, white dots: specificity improved variant following an additional round of maturation of the peptide:MHC specific binder directed against PRAME-004 using the identified off-target peptides as selection determinants). Killing of the target negative cell line T98G is strongly reduced when employing the specificity-improved variant of the peptide:MHC binding polypeptide, whereas killing of the target positive cell line U2OS is only slightly affected.

Figure 3: The combination of several cell-lines for sample generation can be used to achieve a high coverage of the HLA-A*02 presented immunopeptidome. Based on XPRESIDENT® immunopeptidome data for 60 cell lines it is shown that a combination of already 10 of these cell lines would enable a coverage of more than 60% of the HLA-A*02 immunopeptidome, if peptides are to be considered, for which at least 10 peptide identifications on normal tissues have been previously detected.

Figure 4: Cytotoxicity analysis of all peptides identified in the analysis using the PRAME-004 specific peptide:MHC binding polypeptide from Example 1. In brief T2 cell

loaded with 10 nM of respective peptides were co-incubated with human CD8⁺ T-cells in the presence of indicated concentrations of the PRAME-004 specific peptide:MHC binding polypeptide. After 48 h cytotoxicity was quantified by measuring LDH release.

Figure 5: Analysis of the expression profile of the peptide encoding source exon of IFT17-003 in different normal as well as tumor tissues. Tumor (black dots) and normal (grey dots) samples are grouped according to organ of origin. Box-and-whisker plots represent median FPKM value, 25th and 75th percentile (box) plus whiskers that extend to the lowest data point still within 1.5 interquartile range (IQR) of the lower quartile and the highest data point still within 1.5 IQR of the upper quartile. Normal organs are ordered alphabetically. FPKM: fragments per kilobase per million mapped reads. Tissues (from left to right): Normal samples: adipose (adipose tissue); adrenal gl (adrenal gland); bile duct; bladder; blood cells; bloodvess (blood vessels); bone marrow; brain; breast; esoph (esophagus); eye; gall bl (gallbladder); head&neck; heart; intest. la (large intestine); intest. sm (small intestine); kidney; liver; lung; lymph node; nerve periph (peripheral nerve); ovary; pancreas; parathyr (parathyroid gland); perit (peritoneum); pituit (pituitary); placenta; pleura; prostate; skel. mus (skeletal muscle); skin; spinal cord; spleen; stomach; testis; thymus; thyroid; trachea; ureter; uterus. Tumor samples: AML (acute myeloid leukemia); BRCA (breast cancer); CCC (cholangiocellular carcinoma); CLL (chronic lymphocytic leukemia); CRC (colorectal cancer); GBC (gallbladder cancer); GBM (glioblastoma); GC (gastric cancer); HCC (hepatocellular carcinoma); HNSCC (head and neck squamous cell carcinoma); MEL (melanoma); NHL (non-hodgkinHodgkin lymphoma); NSCLCaden (non-small cell lung cancer adenocarcinoma); NSCLCother (NSCLC samples that could not unambiguously be assigned to NSCLCaden or NSCLCsquam); NSCLCsquam (squamous cell non-small cell lung cancer); OC (ovarian cancer); OSCAR (esophageal cancer); PACA (pancreatic cancer); PRCA (prostate cancer); RCC (renal cell carcinoma); SCLC (small cell lung cancer); UBC (urinary bladder carcinoma); UEC (uterine endometrial cancer).

Figure 6: Analysis of the peptide presentation of IFT17-003 on different normal as well as tumor tissues. Upper part: Median MS signal intensities from technical replicate measurements are plotted as dots for single HLA-A*02 positive normal (grey dots, left part of figure) and tumor samples (black dots, right part of figure) on which the peptide

was detected. Boxes display median, 25th and 75th percentile of normalized signal intensities, while whiskers extend to the lowest data point still within 1.5 interquartile range (IQR) of the lower quartile, and the highest data point still within 1.5 IQR of the upper quartile. Normal organs are ordered alphabetically. Lower part: The relative peptide detection frequency in every organ is shown as spine plot. Numbers below the panel indicate number of samples on which the peptide was detected out of the total number of samples analyzed for each organ (N = 592 for normal samples, N = 710 for tumor samples). If the peptide has been detected on a sample but could not be quantified for technical reasons, the sample is included in this representation of detection frequency, but no dot is shown in the upper part of the figure. Tissues (from left to right): Normal samples: adipose (adipose tissue); adrenal gl (adrenal gland); bile duct; bladder; blood cells; bloodvess (blood vessels); bone marrow; brain; breast; esoph (esophagus); eye; gall bl (gallbladder); head&neck; heart; intest. la (large intestine); intest. sm (small intestine); kidney; liver; lung; lymph node; nerve cent (central nerve); nerve periph (peripheral nerve); ovary; pancreas; parathyr (parathyroid gland); perit (peritoneum); pituit (pituitary); placenta; pleura; prostate; skel. mus (skeletal muscle); skin; spinal cord; spleen; stomach; testis; thymus; thyroid; trachea; ureter; uterus. Tumor samples: AML (acute myeloid leukemia); BRCA (breast cancer); CCC (cholangiocellular carcinoma); CLL (chronic lymphocytic leukemia); CRC (colorectal cancer); GBC (gallbladder cancer); GBM (glioblastoma); GC (gastric cancer); GEJC (gastro-esophageal junction cancer); HCC (hepatocellular carcinoma); HNSCC (head and neck squamous cell carcinoma); MEL (melanoma); NHL (non-hodgkinHodgkin lymphoma); NSCLCaden (non-small cell lung cancer adenocarcinoma); NSCLCother (NSCLC samples that could not unambiguously be assigned to NSCLCaden or NSCLCsquam); NSCLCsquam (squamous cell non-small cell lung cancer); OC (ovarian cancer); OSCAR (esophageal cancer); PACA (pancreatic cancer); PRCA (prostate cancer); RCC (renal cell carcinoma); SCLC (small cell lung cancer); UBC (urinary bladder carcinoma); UEC (uterine endometrial cancer).

Figure 7: Superordinate binding motif of the PRAME-004 directed peptide:MHC binding polypeptide determined using the described method, and displayed in the format of a sequence logo. The size of individual amino acids in selected positions reflects their abundance among the identified off-targets. This approach hence delivers

precise information of the binding motif of the PRAME-004 directed peptide:MHC binding polypeptide.

For example, within the identified off-targets histidine (H) in Position 5 of the peptide sequence was more frequent as compared to lysine (K), meaning that, in said position, histidine is preferred by the PRAME-004 binding polypeptide. In other words: A peptide having a histidine residue in position 5 will be bound by the respective PRAME-004 binding polypeptide with higher affinity than a peptide having a lysine residue in position 5. Any other amino acid than histidine or lysine in position 5 will reduce the affinity of the respective peptide even further.

As can be seen, the binding motif accepted by the respective PRAME-004 binding polypeptide is quite flexible in positions 1, 3 and 4, while on positions 5 - 8 a clear preference for selected amino acid residues can be determined, with following hierarchy as shown in Table 1b:

Table 1b: Hierarchy of preferences of amino acids at positions 5, 6, 7 and 8 of PRAME-004 by the respective PRAME-004 binding polypeptide

Position	Hierarchy of preference/tolerance
5	H>K
6	L>N>V>A>I>Q>T
7	I>L>M>K>P>V
8	T>G>S>K>A>D>L>P

Note, in this context, that the so-called “anchoring residues” (abbreviated “AR” herein), mediate binding of the peptide to the peptide binding groove in the MHC. In HLA-A*02:01, these anchoring residues are most often leucine in P2, and leucine or valine in P9, and play only a minor role in the binding interaction between the binding polypeptide and the peptide:MHC complex.

In the following, the full-length sequence of PRAME (Melanoma antigen preferentially expressed in tumors, UniProt: P78395) is shown, with the sequence of the PRAME-004 target peptide (SLLQHLIGL, SEQ ID NO: 1) being shown in bold underline:

>sp|P78395|PRAME_HUMAN Melanoma antigen preferentially expressed in tumors
OS=Homo sapiens OX=9606 GN=PRAME PE=1 SV=1

MERRRLWGSIQSRYISMSVWVTSPPRLVELAGQSLLKDEALAIAALELLPRELFPPLFMAA
FDGRHSQTLKAMVQAWPFTCLPLGVLMKGQHLHLETFKAVLDGLDVLLAQEVRRRWKLQ
VLDLRKNSHQDFWTVWSGNRASLYSFPEPEAAQPMTKKRKVDGLSTEAEQPFIPVEVLVD
LFLKEGACDELFSYLIKVKRKKNVLRCLCKKLLKIFAMPMDIKMILKMVQLDSIEDLEV
TCTWKLPTLAKFSPYLGQMINLRRLLSHIHASSYISPEKEEQYIAQFTSQFLSLQCLQA
LYVDSLFFLRGRLDQLLRHVMNPLETLSITNCRLSEGDMHLSQSPSVSQLSVLSLSGVM
LTDVSPEPLQALLERASATLQDLVFDECGITDDQLLALLPSLSHCSQLTTLSFYGNSISI
SALQSLQHLIGLSNLTHVLYPVPLESYEDIHGTLHLERLAYLHARLRELLCELGRPSMV
WLSANPCPHCGDRTFYDPEPILCPCFMPN

Hence, on the basis of the above data, it is possible to determine a position-based binding motif and a superordinate binding motif, as shown in the following Table 1c and 1d:

Table 1c: Position-based binding motif and superordinate binding motif of the respective PRAME-004 binding polypeptide.

Position	1	2	3	4	5	6	7	8	9
Sequence of PRAME-004	S	L	L	Q	H	L	I	G	L
Position-based binding motif	X	AR	X	X	H	L	I	G	AR
Superordinate binding motif	no preference	AR	no preference	no preference	H>K	L>N>V>A >I>Q>T	I>L>M> K>P>V	T>G>S>K> A>D>L>P	AR

With this information, one may then identify off-target peptides of the polypeptide molecule for peptide:MHC binding polypeptides, by merely shuffling the preferred residues, while placeholders in the residues that have no preferences. One will then end up with a list of potential off-target peptide sequences.

As a next step, one can then search in respective databases where peptides displayed in peptide:MHC complexes are archived, as described elsewhere herein, for peptides comprising such binding motifs.

Table 1d: Peptides identified by the presented method sharing the superordinate binding motif as determined above

Peptide code	SEQ ID NO		AR			superordinate binding motif				AR
PRAME-004	1	S	L	L	Q	H	L	I	G	L
IFT17-003	2	F	M	N	P	H	L	I	S	V
MCM5-006	3	M	L	A	K	H	V	I	T	L
IFIT1-001	4	V	L	L	H	H	Q	I	G	L
FADS2-001	5	L	L	L	A	H	I	I	A	L
CTBP1-001	6	A	L	M	Y	H	T	I	T	L
ITSN1-001	7	I	L	A	M	H	L	I	D	V
ATP1A1-001	8	F	L	P	I	H	L	L	G	L
MCMB-002	9	Y	L	I	L	H	L	I	S	T
EHD4-001	10	A	L	A	K	H	L	I	K	I
SF3B3-005	11	T	L	V	Y	H	V	V	G	V
EHD-001	12	A	L	A	N	H	L	I	K	V
FARSA-001	13	L	T	L	G	H	L	M	G	V
INTS7-002	14	I	L	G	T	H	N	I	T	V
MLXI-001	15	K	L	T	S	H	A	I	T	L
PPP4R1-003	16	H	L	M	P	H	L	L	T	L
RIF1-004	17	A	I	W	E	K	L	I	S	L
SFXN3-001	18	S	L	T	K	H	L	P	P	L
TBCK-002	19	A	L	S	P	H	N	I	L	L
TNRC6B-001	20	S	L	A	R	H	L	M	T	L
ZFYVE16-002	21	A	L	C	P	H	L	K	T	L

These peptides can then be synthesized and be tested for cross-reactivity with the respective PRAME-004 binding polypeptide.

Figure 8: Identification of the position-based binding motif of the peptide:MHC binding polypeptide directed against PRAME-004 using positional scanning, replacing each amino acid in position 1-9 of the peptide sequence except for the anchor residues with alanine. The ratio of the KD of the target peptide PRAME-004 to the alanine substituted variant of the peptide sequence is presented for every peptide. A threshold of 50% (dashed line) of the KD ratio is applied to determine positions which are recognized by the binder. KD values were determined by bio-layer interferometry.

Compared to the method according to the invention, which delivers detailed information on the superordinate binding motif, as well as on off-target peptide

sequences, and hence allows exploring and/or identifying the cross-reactivity for said peptide:MHC binding polypeptide, the mere replacement of each amino acid by alanine only provides very limited information, namely the mere identification of positions where a given amino acid residue is necessary for recognition by the binder.

Figure 9: Complex binding motif determination using amino acid substitutions with all proteinogenic amino acids per position (except cysteine). Ratios of the KD of the target peptide PRAME-004 to the respective positional scanning variant are represented and greyscale-coded (showing low to high values colored from white to dark grey). KD values were determined by bio-layer interferometry.

Figure 10: Analysis of the expression profile of the peptide encoding source exon of MAGEA1 in different normal as well as tumor tissues. For a detailed figure description please refer to the legend of Figure 5.

Figure 11: Analysis of the peptide presentation of MAGEA1 on different normal as well as tumor tissues. For a detailed figure description please refer to the legend of Figure 6.

Figure 12: Potential off-target peptides for a PRAME-004 directed peptide:MHC binding polypeptide identified by the presented method. Depicted are the MS signal intensities obtained after affinity chromatography using the PRAME-004 directed peptide:MHC binding polypeptide against the MS signal intensities obtained after affinity chromatography using the HLA-A*02 specific ("MHC pan specific") antibody BB7.2. Other pan-HLA specific binders as e.g. shown in Table 1 can likewise be used. The ratio of recoveries of individual peptides between BB7.2 and the binder correlates with the binding affinity of the binder to these peptides (targets and off-targets).

The ratio of signal intensities corresponds to the distance to the diagonal (dashed line) with weakly binding peptides located in the upper left part and strong binding peptides located in the lower right part. The arrow indicates PRAME-004. The affinity chromatography for all binders is performed in such fashion that a large excess of all binders is used. The recovery of each peptide is therefore not limited by the number of available binding sites. Nearly quantitative precipitation for peptide:MHC binding polypeptide and HLA-A*02 specific antibody BB7.2 causes maximal values around

one, meaning that peptides which are strongly recognized by the binder appear close to the diagonal line. The grey scale represents the binding affinity fold reduction relative to the target determined by bio-layer interferometry using HIS1K biosensors, as described elsewhere herein. For peptides with binding below the detection limit of the bio-layer interferometry, the binding affinity fold reduction was set to 1000/0.1 %.

It can be seen that both approaches ((i)parallel binding experiment with pMHC binding polypeptide and MHC pan-specific binding molecule and (ii) bio-layer interferometry) deliver similar results regarding the affinities of off-target peptides. MS signal intensity of the MHC pan-specific antibody was not quantifiable for three peptides (e.g. IFIT1-001) due to the stochasticity of the mass spectrometric measurement at very low signal intensities and therefore indicated as "NA" in the figure. Therefore, no ratio could be calculated for these off-targets. However, these uniquely identified peptides after affinity chromatography also indicate a highly relevant off-target as confirmed by bio-layer interferometry measurements.

Thresholds can then be set based on binding affinity fold reduction to discriminate peptides with which the PRAME-004 directed peptide:MHC binding polypeptide is cross-reactive from peptides with which cross-reactivity is no longer considered too weak and therefore not relevant. In general, identified peptides can be classified in strong binding peptides (binding affinity fold reduction ≥ 0.1), medium binding peptides ($0.01 \geq$ binding affinity fold reduction > 0.1) and weakly binding peptides ($0.001 \geq$ binding affinity fold reduction > 0.01).

Figure 13: Correlation of the binding affinity fold reduction relative to the target determined by bio-layer interferometry with the ratio of signal areas [pMHC binding polypeptide/BB7.2] determined by the presented method. The correlation analysis (spearman rank correlation) resulted in a rho value of 0.84. An exponential regression was performed using the function $f(x) = a - \exp(b \cdot x - c)$. The 95% probability prediction bands are indicated as dashed lines. Three peptides were exclusively detected on the matrix coupled to the peptide:MHC binding polypeptide and therefore no ratio of signal intensities could be calculated. The value was set to the maximal detected value. For peptides with binding below the detection limit, the binding affinity fold reduction was set to 1000 for the exponential regression.

Figure 14: Demonstration of a synthetic peptide:MHC library used to identify relevant off-targets. Depicted are the MS signal intensities obtained after affinity chromatography using the PRAME-004 directed peptide:MHC binding polypeptide against the MS signal intensities obtained after affinity chromatography using the HLA-A*02 specific antibody BB7.2. The ratio of signal intensities corresponds to the distance to the diagonal (dashed line) with weakly binding peptides located in the upper left part and strong binding peptides located in the lower right part. Nearly quantitative precipitation for peptide:MHC binding polypeptide and HLA-A*02 specific antibody BB7.2 causes maximal values around one. For demonstration, five previously identified off-targets, the target and two non-binding peptides were isotopically labelled, loaded on a recombinantly produced and refolded MHC monomer of the HLA-A*02:01 allotype and used as synthetic spike-in.

Figure 15: Determination of the position-based binding motif of a PRAME-004 directed peptide:MHC binding polypeptide using the presented method with a synthetic peptide library based on positional mutation of individual amino acids of the target peptide. In this case, all positions 1 to 9 excluding the anchor residues (AR) were exchanged by alanine. The synthetic peptides were integrated in a pMHC complex and added to a cell lysate matrix. Depicted is the binding affinity fold reduction determined as described elsewhere for each amino acid exchange position. The maximum value was normalized to one. Error bars represent the standard deviation of two technical replicates. The obtained position-based binding motif is overall in line with the position-based motif determined in the comparative example using bio-layer interferometry (Figure 8). The presented method seems to have a higher sensitivity, recognizing also weakly binding mutant peptides, which can be seen from relatively high ratio in position 8. Although the dynamic range differs between both methods, the order of the individual weakly binding peptides is consistent with position 5 being the strongest recognized position, followed by position 7, 6 and 8. This is also in line with the determined superordinate binding motif (Figure 7) that shows for position 5 the highest selectivity by far, followed by position 7, 6, and 8. In line with the positional scanning approach, Figure 7 shows for position 8 as well reduced selectivity for glycine and a variety of allowed amino acids.

Examples

Example 1:

The targeted MHC peptide used in this example which is presented in the context of HLA-A*02 is derived from Melanoma antigen preferentially expressed in tumors (PRAME) and shows the sequence SLLQHLIGL ((SEQ ID NO: 1), herein also referred to as PRAME-004).

The peptide:MHC binding polypeptide used in these experiments was exemplified by a modified T-cell receptor molecule which has been engineered to be soluble and showed an enhanced affinity to the PRAME-004 peptide and additionally comprised a CD3-binding antibody moiety.

The peptide:MHC binding polypeptide used in these experiments, the method how it was obtained and its technical characteristics are disclosed in PCT/EP2020/050936, the content of which is incorporated herein by reference.

As a biological source of peptide:MHC mixtures, the human HLA-A*02 high expressing glioblastoma derived cell line T98G was used. This cell line had been previously tested in cytotoxicity experiments with the described peptide:MHC binding polypeptide directed against PRAME-004 and showed positive killing.

Five hundred million T98G cells were subjected to lysis in a CHAPS detergent-containing buffer and homogenized assisted by sonification.

The peptide:MHC binding polypeptide was coupled to a solid Sepharose® matrix at a pre-determined ratio using chemical coupling after BrCN activation. In parallel the same amount of Sepharose® was also activated for coupling using the same strategy, yet without addition of the peptide:MHC binding polypeptide. Instead a 0.1 M solution of the amino acid glycine was added to the Sepharose®, which instead coupled to the chemically activated groups. The T98G lysate containing the mixture of peptide:MHC molecules was then applied to two affinity chromatography columns loaded with 200 µl of the glycine coupled Sepharose® matrix or 200 µl of the Sepharose® matrix coupled

with the peptide:MHC binding polypeptide. The T98G derived lysate was thereby applied in such a fashion that it would first be run over the glycine coupled Sepharose® (referred to herein as glycine column) to remove or deplete any peptides, which would bind non-specifically to the column or the Sepharose® matrix before the isolation of peptides which bind to the peptide:MHC binding polypeptide (Figure 1). After washing of the affinity columns with PBS and ddH₂O the bound peptide:MHC complexes were eluted from the columns using Trifluoroacetic acid (TFA).

During this step, MHC bound peptides are also released from the MHC moiety and can be separated from higher molecular weight molecules by ultrafiltration using specified devices with a molecular weight cutoff of less than 10 kDa.

The isolated peptide mixtures were then finally subjected to liquid chromatography coupled mass spectrometry (LC-MS) using a nanoACQUITY UPLC system (Waters) followed by an Orbitrap fusion™ Tribrid™ mass spectrometer (Thermo Scientific).

Mass spectrometry instruments were operated in data-dependent mode (DDA) utilizing different fragmentation techniques (in this example, CID and HCD fragmentation) as well as MS/MS spectra readout in two different analyzers (in this example, IonTrap and Orbitrap analyzers).

Peptide fragment spectra were searched against the human proteome using a modified version of the International protein index (IPI v.378) and the Universal protein resource (UniProt) sequence database with the search engine SEQUEST. All peptides eluted and identified from the glycine column were excluded from further analysis as these represent non-specific binding peptides. Furthermore, known contaminants according to in-house databases and algorithms for their identification were removed from the analysis.

In total 20 peptides were identified after isolation and processing which are shown in Table 2. For reference, the target peptide PRAME-004 is shown as well, which has however not been identified among the isolated peptides and was not expected to be identified.

In order to confirm their relevance and analyze the binding strength in comparison to the target peptide, all peptides were subjected to bio-layer interferometry. Measurements were performed with bio-layer interferometry (in this case, an Octet RED384 system using settings recommended by the manufacturer). Briefly, binding kinetics were measured at 30°C and 1000 rpm shake speed using PBS, 0.05% Tween-20, 0.1% BSA as buffer. Peptide:MHC complexes were loaded onto biosensors (HIS1K) prior to analyzing serial dilutions of the peptide:MHC specific binder. The ratio of equilibrium dissociation constants (KD) in comparison to PRAME-004 is presented in the last column of Table 2.

A selection of these peptides was further tested in cytotoxicity experiments. In brief T2 cell (10,000 cells/well) loaded with 10 nM of respective peptides were co-incubated with human CD8+ T cells (50,000 cells/well) in the presence of indicated concentrations of the PRAME-004 specific peptide:MHC binding polypeptide (Figure 4). After 48 h cytotoxicity was quantified by measuring LDH release using CytoTox 96 Non-Radioactive Cytotoxicity Assay Kits (PROMEGA). Corresponding EC50 values of tested peptides are also listed in Table 2. The main off-target emanating from this analysis was IFT17-003 which showed similar KD and EC50 values for the peptide:MHC binding polypeptide as compared to the target peptide PRAME-004.

XPRESIDENT® peptide presentation and gene expression data can be used to evaluate the potential safety risk of off-target peptides by differentiating relevant off-targets from less relevant off-targets, which are only presented/expressed in the context of other tumor tissues. In this example, IFT17-003 is considered a highly relevant off-target due to ubiquitous expression (Figure 5) and presentation (Figure 6) of the peptide on different normal tissues. Combination of the presented data from this example with additional large-scale peptide presentation or expression data is therefore of additional value for the off-target risk assessment.

In order to improve the specificity of the peptide:MHC binding polypeptide an additional round of maturation has been performed using the identified peptides as selection determinants. Thereby specificity of the newly generated molecules could be greatly improved shown by a reduced killing of the target negative cell line T98G in Figure 2.

Killing assays were essentially performed as described above. LDH release of target positive or target negative cells (10,000 cells/well) was quantified after co-incubation with human PBMCs (100,000 cells/well) and indicated concentrations of the peptide:MHC binding polypeptide for 48 h. The original peptide:MHC binding polypeptide molecule used for peptide isolation is shown as filled squares, whereas specificity improved variants are shown as open circles. Control peptide:MHC binding polypeptide molecule (square with asterisk) and control without bispecific molecule (circle with asterisk) do not induce target cell killing.

Table 2: Overview of the identified peptide:MHC binding polypeptide specific peptides. Indicated on top is the target PRAME-004. The EC50 values of cytotoxicity experiments using peptide loaded T2 cells is specified as well as the binding affinities determined by bio-layer interferometry using HIS1K biosensors.

Peptide code	Peptide sequence	EC50 [pM]	Binding affinity fold reduction
PRAME-004	SLLQHLIGL (SEQ ID NO: 1)	1.2	1
IFT17-003	FMNPHLISV (SEQ ID NO: 2)	1.6	1
MCM5-006	MLAKHVITL (SEQ ID NO: 3)	16.1	3
IFIT1-001	VLLHHQIGL (SEQ ID NO: 4)	38.4	8
FADS2-001	LLLAHIIAL (SEQ ID NO: 5)	83.7	13
CTBP1-001	ALMYHTITL (SEQ ID NO: 6)	79.31	13
ITSN1-001	ILAMHLIDV (SEQ ID NO: 7)	1024	36
ATP1A1-001	FLPIHLLGL (SEQ ID NO: 8)	196	106
MCMB-002	YLILHLIST (SEQ ID NO: 9)	n.a.	127
EHD4-001	ALAKHLIKI (SEQ ID NO: 10)	n.a.	61
SF3B3-005	TLVYHVVG (SEQ ID NO: 11)	n.a.	152
EHD-001	ALANHLIKV (SEQ ID NO: 12)	n.a.	159
FARSA-001	LTGLHLMGV (SEQ ID NO: 13)	n.a.	38

INTS7-002	ILGTHNITV (SEQ ID NO: 14)	n.a.	57
MLXI-001	KLTSHAITL (SEQ ID NO: 15)	n.a.	12
PPP4R1-003	HLMPHLLTL (SEQ ID NO: 16)	n.a.	16
RIF1-004	AIWEKLISL (SEQ ID NO: 17)	n.a.	156
SFXN3-001	SLTKHLPPL (SEQ ID NO: 18)	n.a.	60
TBCK-002	ALSPHNILL (SEQ ID NO: 19)	n.a.	142
TNRC6B-001	SLARHMLTL (SEQ ID NO: 20)	n.a.	4
ZFYVE16-002	ALCPHLKTL (SEQ ID NO: 21)	n.a.	33

Identification of a binding motif

The identified peptides can be further used to infer a binding motif for the peptide:MHC binding polypeptide, which provides information on which of the amino acids in the peptide sequence are of relevance for the binding of said polypeptide.

Moreover, additional information on the binding motif can be deduced from the amino acids within the relevant positions. Based on the identified set of peptides only a subset of amino acids is tolerated in positions 1-9 of the amino acid sequence (see Figure 7 and Table 3).

Table 3: Overview of the tolerated amino acids for each position identified by the presented method.

<i>Position</i>	<i>Tolerated amino acid residues</i>
1	A, S, F, I, L, H, K, M, T, V, Y
2	L, I, M, T
3	A, L, M, T, C, G, I, N, P, S, V, W
4	P, K, Y, A, E, G, H, I, L, M, N, Q, R, S, T
5	H, K
6	L, N, V, A, I, Q, T
7	I, L, M, K, P, V

8	T, G, S, K, A, D, L, P
9	L, V, I, T

In contrast to common amino acid scanning approaches, in which amino acids are replaced at individual positions by mutations and subsequently tested in *in vitro* assays also multiple substitutions with different potentially opposite effects on the overall binding strength can be elucidated. For example, if a substitution in position 6 of the natural amino acid sequence leads to a decrease in the overall binding affinity this might be rescued by a similar substitution in position 8 which can lead to a strong increase in the binding affinity of the peptide:MHC binding polypeptide to the peptide:MHC molecule.

The thus generated binding motif was used to search different protein sequence databases (e.g. UniProt, IPI) to find additional off-target peptides which reflect and fit to the restrictions imposed by the binding motif (i.e. defined sets of amino acids which are tolerated in relevant positions of the binding motif).

Comparative example 1

The following experiments show how currently available methods in the art would not identify the most relevant off-target peptides identified in example 1 and are therefore not able to predict unwanted side effects of peptide:MHC binding polypeptides intended for administration *in vivo*.

Identification of the binding motif using positional scanning:

Variants of the native PRAME-004 sequence, in which each amino acid is subsequently replaced with the amino acid alanine were tested for their potential to bind to the peptide:MHC binding polypeptide using bio-layer interferometry.

A L L Q H L I G L (SEQ ID NO. 38)
 S A L Q H L I G L (SEQ ID NO. 39)
 S L A Q H L I G L (SEQ ID NO. 40)
 S L L A H L I G L (SEQ ID NO. 41)
 S L L Q A L I G L (SEQ ID NO. 42)
 S L L Q H A I G L (SEQ ID NO. 43)
 S L L Q H L A G L (SEQ ID NO. 44)
 S L L Q H L I A L (SEQ ID NO. 45)

S L L Q H L I G A (SEQ ID NO. 46)

Figure 8 shows the results of these experiments. Five of the alanine-substituted peptides lead to a 50% or greater decrease in binding affinity (or 2-fold or greater increase of the K_D , respectively) as compared to the wild-type sequence and were therefore considered as essential for binding. Based on these results the binding motif would result in XXXXHLIGL (SEQ ID NO. 22), wherein X represents any amino acid.

In an extended variant of the positional scanning approach the PRAME-004 sequence was substituted at each position by any of the naturally occurring amino acids in a similar manner as described before. The only proteinogenic amino acid, which was not used for substitution of PRAME-004 was cysteine as this amino acid is known to rapidly undergo several chemical modifications which can lead to false interpretations regarding recognition of peptides during testing. So, in total $9 \times 18 = 162$ peptides were investigated.

Each peptide was again tested for its binding affinity using bio-layer interferometry (Figure 9). Peptides which lead to a 50% or greater decrease in binding affinity (or 2-fold or greater increase of the K_D , respectively) as compared to the wild-type sequence were considered as not tolerated or detrimental for peptide binding. This resulted in a complex binding motif with a set of different amino acids being tolerated or accepted in position 1-9 of the amino acid sequence:

$X_1X_2X_3X_4HX_5IX_6X_7$

wherein X_1 is selected from any of ADEFGHIKLMNPQRSTVWY; X_2 is selected from any of AFGILMQSTVY; X_3 is selected from any of ADGIKLMNQSTVW; X_4 is selected from any of AFGHIKLMNPQRSTVWY; X_5 is selected from any of ILM; X_6 is selected from any of GST; and X_7 is selected from any of EFHIKLMPQTVY.

Similarity search based on Ala-Scan derived binding motif:

An in-house software tool was used to search different protein sequence databases (IPI v. 3.78, Ensembl Version 77 GrCH38 including SNVs, NCBI non-redundant protein database) for human proteins which contain the identified motif sequence (X-X-X-X-X-H-L-I-G-L) (SEQ ID NO. 22), in which X could be constituted by any amino acid. Eight

unique peptides were identified: the target itself, PRAME-004 and seven peptides originating from different human proteins: VEZT (Vezatin), HTR2C (5-hydroxytryptamine receptor 2C), HEPHL1 (Hephaestin-like protein 1), COL28A1 (Collagen alpha-1(XXVIII) chain), SLC2A1 (Solute carrier family 2, facilitated glucose transporter member 1), SLC44A3 (Choline transporter-like protein 3), PIEZO2 (Piezo-type mechanosensitive ion channel component 2).

The amino acid sequences of these peptides are shown below.

Protein (Uniprot accession number)	Sequence									
PRAME (P78395) (SEQ ID NO. 1)	S	L	L	Q	H	L	I	G	L	
VEZT (Q9HBM0) (SEQ ID NO. 23)	H	P	S	Q	H	L	I	G	L	
HTR2C (P28335) (SEQ ID NO. 32)	S	F	L	V	H	L	I	G	L	
HEPHL1 (Q6MZM0) (SEQ ID NO. 33)	R	V	S	W	H	L	I	G	L	
COL28A1 (Q2UY09) (SEQ ID NO. 34)	I	N	E	S	H	L	I	G	L	
SLC2A1 (P11166) (SEQ ID NO. 35)	R	R	T	L	H	L	I	G	L	
SLC44A3 (Q8N4M1) (SEQ ID NO. 36)	M	W	S	Y	H	L	I	G	L	
PIEZO2 (Q9H5I5) (SEQ ID NO. 37)	F	T	A	G	H	L	I	G	L	

In-house XPRESIDENT® immunopeptidome data from 592 normal tissue samples and 710 tumor tissue samples, all derived from HLA-A*02 typed individuals, showed that none of the 7 predicted off-target peptides, has ever been identified to be presented in the context of HLA-A*02 on any of the analyzed samples. Notably, the VEZT and SLC2A1 derived peptide have been previously identified by XPRESIDENT® on tissue samples of non-A*02 positive individuals suggesting that they are presented by different HLA allotypes (HLA-B*07 in case of the VEZT derived peptide and HLA-B*27 in case of the SLC2A1 derived peptide) and are therefore not likely to produce an off-target risk in the context of an HLA-A*02 restricted peptide:MHC binding polypeptide. The positional scanning and prediction approach failed however to identify any of the relevant off-target peptides which could be identified with the superior method described in this application.

Similarity search based on the superordinate binding motif:

The same in-house software tool was also used to predict peptides derived from the human proteome which fulfill the criteria of the complex binding motif. As cysteine was excluded during substitution, this amino acid was hence additionally allowed for every position in the amino acid sequence resulting in the following motif:

$X_8X_9X_{10}X_{11}X_{12}X_{13}X_{14}X_{15}X_{16}$

wherein X_8 is selected from any of ACDEFGHIKLMNPQRSTVWY; X_9 is selected from any of ACFGILMQSTVY; X_{10} is selected from any of ACDGIKLMNQSTVW; X_{11} is selected from any of ACFGHIKLMNPQRSTVWY; X_{12} is selected from any of CH; X_{13} is selected from any of CILM; X_{14} is selected from any of CI; X_{15} is selected from any of CGST; and X_{16} is selected from any of CEFHIKLMNPQTVY.

The search resulted in a total list of 888 different peptides fulfilling the binding motif criteria. Only two peptides (IFT17-003 and ATP1A1-001) were overlapping with the list of relevant off-targets identified by the superior method described in this application in example 1, whereas the rest would not have been identified in the prediction-based approach, even if all 888 peptides would have been tested afterwards in downstream *in vitro* analyses.

Example 2:

The targeted MHC peptide used in this example which is presented in the context of HLA-A*02 is derived from the melanoma associated antigens A4 and A8 (MAGEA4/A8) and shows the sequence KVLEHVVRV (SEQ ID NO. 24), herein also referred to as MAGEA4/8.

The peptide:MHC binding polypeptide is constituted by a modified T-cell receptor molecule which has been engineered to be soluble and shows an enhanced affinity to the MAGEA4/A8 derived peptide and additionally comprised a CD3 binding antibody moiety. As a biological source of peptide:MHC mixtures the human HLA-A*02 high expressing and MAGEA4/8 positive lung adenocarcinoma derived cell line NCI-H1755 has been employed. Five hundred million cells of this cell line were subjected to lysis in a CHAPS detergent-containing buffer and homogenized assisted by sonification.

Coupling of the peptide:MHC binding polypeptide and affinity chromatography were carried out as described in Example 1. Before applying the NCI-H1755 lysate containing the mixture of peptide:MHC molecules to the glycine coupled and peptide:MHC binder coupled Sepharose® the volume was split half and half. The second half of the volume was run in parallel over a different glycine coupled Sepharose® matrix followed by a Sepharose® matrix coupled with the HLA-A*02 specific antibody BB7.2. The latter is aimed to isolate the complete spectrum of peptides presented by HLA-A*02 in this cell line (see also Figure 1).

Peptides were eluted from all columns and subjected to mass spectrometry analysis as outlined in Example 1. Peptides eluted from glycine columns as well as known contaminants were again excluded from further analysis. In addition, all peptide precursor signals were quantified over all different runs using SuperHirn algorithm (Mueller *et al.*, 2007). Features were extracted and quantified over all mass spectrometry experiments using a fixed retention time window of ± 3 min, and a mass accuracy of ± 5 ppm.

Ratios of the resulting area of individual peptide precursor signals from the MAGEA4/8 specific peptide:MHC binding polypeptide to the same precursor signals from the BB7.2 preparation were calculated. These ratios reflect the isolation efficiency of the MAGEA4/8 peptide:MHC binding polypeptide in comparison to the HLA-A*02 specific antibody BB7.2. Due to the high affinity of the MAGEA4/8 specific peptide:MHC binding polypeptide for their target as well as for potential off-targets peptides bound to the HLA-A*02 molecule, the isolation efficiencies for these peptides are much higher as compared to BB7.2 which has an affinity in the lower nanomolar range towards HLA-A*02 largely independent of the bound peptide species (Parham and Brodsky, 1981). Analysis of the mass spectrometry data identified 10 peptides including the target peptide MAGEA4/8 (see Table 4). Ranking of these peptides according to the ratio of areas of the peptide:MHC binding polypeptide and BB7.2 enables a determination of the isolation efficiency of these peptides in comparison to BB7.2 which correlates with the binding affinity employing bio-layer interferometry as described in Example 1. Thereby the risk for off-target toxicities and potential therapeutic windows between target and off-target peptides can be directly deduced from the quantitative data of the mass spectrometry experiments.

In the presented example in Table 4 the ratio of areas for MAGEA1 is smaller as compared to the target peptide MAGEA4/8 (~11 as compared to ~10.6), which translates into a very small reduction of binding affinity of 4.1. In contrast, for the peptide HEAT5RA the large decrease in the ratio of areas, around 800-fold lower as compared to MAGEA4/8, is also reflected in a largely reduced binding affinity of 238 as compared to MAGEA4/8.

Deeper analysis of peptide presentation and gene expression data in XPRESIDENT® show that MAGEA1 does not present a relevant off-target risk, as it is exclusively presented on cancer tissues (Figure 10) and shows a cancer-testis like expression pattern (Figure 11).

Table 4: Overview of the identified peptide:MHC binding polypeptide specific peptides. Indicated is the ratio of peptides eluted from the MAGEA4/8 binding polypeptide to the HLA-A*02 specific binding peptide BB7.2 by mass spectrometry. In the top row the target peptide MAGEA4/A8 is presented showing the highest ratio of signal areas. The PMBEC score is a measure for peptide similarity to the target sequence. The binding affinities were determined by bio-layer interferometry using HIS1K biosensors.

Peptide code	Peptide sequence	Ratio of signal areas [MAGEA4/8 binder/BB7.2]	PMBEC	Binding affinity fold reduction
MAGEA4/A8	KVLEHVVRV (SEQ ID NO. 24)	11	1.85568	1
MAGEA1	KVLEYVIKV (SEQ ID NO. 25)	10.615	1.42928	4.1
	KVLEFLAKV (SEQ ID NO. 26)	5.218	1.24828	
	KIIDLLPKV (SEQ ID NO. 27)	4.009	0.94598	
	KLQEFLQTL (SEQ ID NO. 28)	0.032	0.45866	
HEAT5RA	KVLETLVTV (SEQ ID NO. 29)	0.014	1.06691	237.7
FAM115A	KLGSVPVTV (SEQ ID NO. 30)	0.006	0.50865	503.1
	KIADFGWSV (SEQ ID NO. 31)	0.002	0.53985	

Example 3:

The targeted MHC peptide used in this example is the same that was described in Example 1 (PRAME-004, SEQ ID NO: 1). The peptide:MHC binding polypeptide was likewise exemplified by a modified T-cell receptor molecule which has been engineered to be soluble and showed an enhanced affinity to the PRAME-004 peptide and additionally comprised a CD3-binding antibody moiety, but a different variant compared to Example 1 leading to changes in the off-target recognition.

As a biological source of peptide:MHC mixtures, a pool of ten distinct cell lines was utilized. The cell lines were selected to maximize the HLA-A*02:01 peptide amount and diversity. Five hundred million cells per cell line were subjected to lysis in a CHAPS detergent-containing buffer and homogenized and pooled. For individual analyses, aliquots with cell numbers in the order of magnitude of five hundred million were used.

Coupling and affinity chromatography of the peptide:MHC binding polypeptide and HLA-A*02 specific antibody BB7.2 were carried out as described in Example 2, but to increase the contact time of lysate and matrix coupled peptide:MHC binding polypeptide, the incubation was performed in a reaction tube shaking for 16 hours.

In total 79 peptides were identified after isolation and processing. The ten most relevant off-targets are shown in Table 5 including the target peptide PRAME-004. 32 of the 79 peptides showed low ratio of signal intensities [pMHC binding polypeptide/BB7.2] (below 0.005) and were therefore expected to be low- or non-binding to the peptide:MHC binding polypeptide. These peptides were included in the further analysis as reference. The selected peptides are displayed in Figure 12 according to the obtained MS signal intensities after purification using the peptide:MHC binding polypeptide compared to the MS signal intensities after purification using the HLA-A*02 specific antibody BB7.2. The distance to the diagonal represents the ratio of signal intensities. For the target and strong off-targets a nearly quantitative affinity chromatography is expected leading to a ratio of signal intensities [pMHC binding polypeptide /BB7.2] around one (dashed line). The names of the ten strongest binding off-targets are displayed in the graph, all lying close to the diagonal as expected. Peptides with very low ratio of signal intensities [pMHC binding polypeptide /BB7.2] did

show extremely large or non-quantifiable binding affinity fold reduction using bio-layer interferometry as reference.

Table 5: Overview of the top ten most relevant off-targets based on their binding affinity fold reduction determined by bio-layer interferometry using HIS1K biosensors. Indicated is the ratio of peptides eluted from the PRAME-004 binding polypeptide to the HLA-A*02 specific binding peptide BB7.2 by mass spectrometry. For values marked with asterisk, no ratio could be calculated due to low abundance on the peptides eluted from the HLA-A*02 specific binding peptide BB7.2 and the maximal measured value was assumed.

Peptide code	Peptide sequence	Ratio of signal areas [PRAME binder/BB7.2]	Binding affinity fold reduction
PRAME-004	SLLQHLIGL (SEQ ID NO: 1)	0.9	1.0
IFT17-003	FMNPHLISV (SEQ ID NO: 2)	0.8	1.0
MCM5-006	MLAKHVITL (SEQ ID NO: 3)	3.1	2.4
IFIT1-001	VLLHHQIGL (SEQ ID NO: 4)	3.1*	8.1
MLXI-001	KLTSHAITL (SEQ ID NO: 15)	3.1*	9.2
CTBP1-001	ALMYHTITL (SEQ ID NO: 6)	1.0	10.6
PPP4R1-003	HLMPHLLTL (SEQ ID NO: 16)	1.4	13.0
TUB-006	YMNNDLIGL (SEQ ID NO: 47)	2.0	13.1
FADS2-001	LLLAHIIAL (SEQ ID NO: 5)	0.3	13.6
CTBP2-001	AMMYHTITL (SEQ ID NO: 48)	1.4	15.8

In order to confirm their relevance and analyze the binding strength in comparison to the target peptide, all peptides were subjected to bio-layer interferometry as described in Example 1. The ratio of equilibrium dissociation constants (KD) in comparison to PRAME-004 was determined and correlated with the ratio of signal intensities [pMHC

binding polypeptide/BB7.2] determined by the presented method. The determined binding affinity fold reduction is depicted as greyscale in Figure 12. The correlation of binding affinity fold reduction with ratio of signal intensities [pMHC binding polypeptide /BB7.2] determined by the presented method is demonstrated in Figure 13. As discussed above, peptides with strong binding to the peptide:MHC binding polypeptide could be identified with high ratio of signal intensities [pMHC binding polypeptide /BB7.2], while non-binding peptides showed ratios of signal intensities [pMHC binding polypeptide/BB7.2] below 0.001. In order to obtain a maximal sensitivity of the method, the peptide:MHC binding polypeptide and the HLA-A*02 specific antibody BB7.2 are applied in excess leading to a nearly quantitative extraction of off-targets in both cases and to a maximal ratio of signal areas [pMHC binding polypeptide/BB7.2] around one. To take this saturation effect into account an exponential regression was performed (Figure 13).

The data demonstrates that binding affinities can be estimated using the ratio of signal intensities of a single experiment applying the presented method.

Example 4:

The targeted MHC peptide used in this example is the same that was described in Example 1 (PRAME-004, SEQ ID NO: 1). The peptide:MHC binding polypeptide was likewise exemplified by a modified T-cell receptor molecule which has been engineered to be soluble and showed an enhanced affinity to the PRAME-004 peptide and additionally comprised a CD3-binding antibody moiety, but a different variant compared to Example 1 and 3.

To demonstrate the capacity of the presented method to identify potential off-targets from a synthetic peptide library, a mixture of eight isotopically labelled peptide:MHC complexes were spiked in a biological matrix derived from a renal cell carcinoma tissue sample. For the experiment 0.3 g tissue were lysed as described above and 2 pmol of the peptide:MHC complexes were added. The lysate was split into equal parts and subjected to affinity chromatography using the peptide:MHC binding polypeptide and HLA-A*02 specific antibody BB7.2. The isotopically labelled peptides were consisted of the target, five known cross-reactive peptides and two non-binding peptides as negative control (Table 6).

The signal areas of the PRAME binding polypeptide versus the HLA-A*02 specific antibody BB7.2 are displayed in Figure 14. As expected, the target and the already known off-targets show a high ratio of signal areas [pMHC binding polypeptide /BB7.2] in the order of magnitude around one and therefore would be classified as highly relevant by the presented method. Likewise, both negative control peptides show very low ratio of signal areas [pMHC binding polypeptide /BB7.2] and therefore would be considered as less relevant based on the presented method.

Table 6: Overview of the spiked isotopically labelled peptide:MHC complexes. The isotopically labelled amino acid is indicated by an asterisk. Indicated is the ratio of peptides eluted from the PRAME binding polypeptide to the HLA-A*02 specific binding peptide BB7.2 by mass spectrometry. In the top row the target peptide PRAME-004 is presented, followed by five already known off-targets and two non-binding negative control peptides. The PMBEC score is a measure for peptide similarity to the target sequence. The binding affinities were determined by bio-layer interferometry using HIS1K biosensors.

Peptide code	Peptide sequence	Ratio of signal areas [PRAME binder/BB7.2]	PMBEC	
PRAME-004*	SLL*QHLIGL (SEQ ID NO. 1)	0.5422	1.60	Target
IFIT1-001*	VL*LHHQIGL (SEQ ID NO. 4)	1.4524	1.08	off-target
ATP-009*	HLLMHLI*GS (SEQ ID NO. 49)	2.3194	1.05	off-target
FADS2-001*	L*LLAHIIAL (SEQ ID NO. 5)	0.5232	0.99	off-target
MCM5-006*	ML*AKHVITL (SEQ ID NO. 3)	1.7379	0.76	off-target
IFT17-003*	FMNPHL*ISV (SEQ ID NO. 2)	0.9949	0.75	off-target
GAL-001*	SL*DPSSPQV (SEQ ID NO. 50)	0.0015	0.31	neg. control
KLHL14-001*	VLDDSIYL*V (SEQ ID NO. 51)	0.0016	0.21	neg. control

Example 5:

The targeted MHC peptide used in this example is the same that was described in Example 4 (PRAME-004, SEQ ID NO: 1). The peptide:MHC binding polypeptide was likewise exemplified by a modified T-cell receptor molecule which has been engineered to be soluble and showed an enhanced affinity to the PRAME-004 peptide and additionally comprised a CD3-binding antibody moiety.

To demonstrate an additional approach of the presented method to retrieve information on the binding motif of the peptide:MHC binding polypeptide, synthetic peptide variants of the target peptide were utilized as spike-in library. In this approach, each amino acid position relevant for recognition by a peptide:MHC binding polypeptide was exchanged ("scanned") individually by alanine (SEQ ID NO. 38 to 46). Relevant positions were in this example all positions (1 to 9) without the anchor amino acids in position two and nine that are crucial for peptide-MHC binding. The synthetic peptides were loaded on recombinantly produced and refolded MHC molecules of the HLA-A*02:01 allotype and spiked in a biological matrix derived from a renal cell carcinoma tissue sample. For the experiment 0.1 g tissue were lysed as described above and 140 fmol of the peptide:MHC complexes were added. The lysate was split into equal parts and subjected to affinity chromatography using the peptide:MHC binding polypeptide and HLA-A*02 specific antibody BB7.2. Peptide quantification was performed using a targeted mass spectrometric measurement.

As described earlier, the signal intensity ratio of the pMHC binding polypeptide targeting PRAME versus the HLA-A*02 specific antibody BB7.2 were calculated as estimate for binding affinity. The obtained MS signal intensity ratios are depicted in Figure 15. The ratios were normalized to the maximal value for better comparison to the comparative example in Figure 8 using bio-layer interferometry measurements. In lined with the comparative example for position-based binding motif determination (Figure 8), positions 1,3 and 4 are not relevant for binding to the pMHC complex. The presented method seems to have a higher sensitivity to recognize also weakly binding mutant peptides which can be seen e.g. from the relatively high ratio in position 8 indicating a different dynamic range of the presented method. However, the order of the individual peptides is consistent with position 5 being the strongest recognized position, followed by position 7, 6 and 8. According to the data, position 8 would not be considered relevant for the binding, although the MS signal intensity ratio is markedly

reduced. This is also in line with the determined superordinate binding motif (Figure 7) that shows in position 8 a reduced selectivity for glycine and a variety of other allowed amino acids. The comparison to Figure 7 shows as well that position 5 is the most relevant position for binding by far, which is also reflected very well by the presented position-based binding motif.

An extension of the spike-in library to include also other replacement amino acids than alanine could increase the information obtained about the binding motif in a single experiment. Moreover, synthetic libraries of hundreds to thousands of naturally occurring peptides could be applied as synthetic library. One main advantage of the demonstrated aspect of the presented method is that the respective binding affinities could be determined in parallel.

Abbreviations

AR	Anchoring residues
APC	Antigen presenting cells
BIRD	Blackbody infrared radiative dissociation
BiTE	Bispecific T-cell engager
CAR	Chimeric antigen receptors
CDR	Complementarity determining regions
CID	Collision-induced dissociation
DART	Dual-affinity re-targeting antibody
DDA	Data-dependent acquisition
DIA	Data-independent acquisition
DRIP	Defective ribosomal particles
ECD	Electron-capture dissociation
EDD	Electron-detachment dissociation
ETCID	Electron-transfer and collision-induced dissociation
ETD	Electron-transfer dissociation
ETHCD	Electron-Transfer/Higher-Energy Collision Dissociation
HCD	Higher-energy collisional dissociation
IRMPD	Infrared multiphoton dissociation

IQR	Interquartile range
KD	Dissociation constant
NETD	Negative electron-transfer dissociation
LDH	Lactate dehydrogenase
PBMC	Peripheral blood mononuclear cell
SID	Surface-induced dissociation
SMITE	Simultaneous multiple interaction T-cell engaging
TandAb	Tandem antibody
TCR	T-cell receptor
TFA	Trifluoroacetic acid
TIL	Tumor-infiltrating lymphocytes

References

Barnstable CJ, Bodmer WF, Brown G, Galfre G, Milstein C, Williams AF, Ziegler A (1978). Production of monoclonal antibodies to group A erythrocytes, HLA and other human cell surface antigens-new tools for genetic analysis. *Cell* 14, 9-20.

Berger AE, Davis JE, Cresswell P (1982). Monoclonal antibody to HLA-A3. *Hybridoma* 1, 87-90.

Bijen HM, van der Steen DM, Hagedoorn RS, Wouters AK, Wooldridge L, Falkenburg JHF, Heemskerk MHM (2018). Preclinical Strategies to Identify Off-Target Toxicity of High-Affinity TCRs. *Mol Ther* 26, 1206-1214.

Birnbaum ME, Mendoza JL, Sethi DK, Dong S, Glanville J, Dobbins J, Ozkan E, Davis MM, Wucherpfennig KW, Garcia KC (2014). Deconstructing the peptide-MHC specificity of T cell recognition. *Cell* 157, 1073-1087.

Ekeruche-Makinde J, Miles JJ, van den Berg HA, Skowera A, Cole DK, Dolton G, Schauenburg AJ, Tan MP, Pentier JM, Llewellyn-Lacey S, *et al.* (2013). Peptide length determines the outcome of TCR/peptide-MHCI engagement. *Blood* 121, 1112-1123.

Jahn L, van der Steen DM, Hagedoorn RS, Hombrink P, Kester MG, Schoonakker MP, de Ridder D, van Veelen PA, Falkenburg JH, Heemskerk MH (2016). Generation of CD20-specific TCRs for TCR gene therapy of CD20low B-cell malignancies insusceptible to CD20-targeting antibodies. *Oncotarget* 7, 77021-77037.

Karlin S, Altschul SF (1993). Applications and statistics for multiple high-scoring segments in molecular sequences. *Proc Natl Acad Sci U S A* 90, 5873-5877.

Kim Y, Sidney J, Pinilla C, Sette A, Peters B (2009). Derivation of an amino acid similarity matrix for peptide: MHC binding and its application as a Bayesian prior. *BMC Bioinformatics* 10, 394.

Kolstad A, Hansen T, Hannestad K (1987). A human-human hybridoma antibody (TrB12) defining subgroups of HLA-DQw1 and -DQw3. *Hum Immunol* 20, 219-231.

Lampson LA, Levy R (1980). Two populations of Ia-like molecules on a human B cell line. *J Immunol* 125, 293-299.

Linette GP, Stadtmauer EA, Maus MV, Rapoport AP, Levine BL, Emery L, Litzky L, Bagg A, Carreno BM, Cimino PJ, *et al.* (2013). Cardiovascular toxicity and titin cross-reactivity of affinity-enhanced T cells in myeloma and melanoma. *Blood* 122, 863-871.

Maeda H, Hirata R (1984). Separation of four class II molecules from DR2 and DRw6 homozygous B lymphoid cell lines. *Immunogenetics* 20, 639-647.

Morgan RA, Chinnasamy N, Abate-Daga D, Gros A, Robbins PF, Zheng Z, Dudley ME, Feldman SA, Yang JC, Sherry RM, *et al.* (2013). Cancer regression and neurological toxicity following anti-MAGE-A3 TCR gene therapy. *J Immunother* 36, 133-151.

Mueller LN, Rinner O, Schmidt A, Letarte S, Bodenmiller B, Brusniak MY, Vitek O, Aebersold R, Muller M (2007). SuperHirn - a novel tool for high resolution LC-MS-based peptide/protein profiling. *Proteomics* 7, 3470-3480.

Parham P, Brodsky FM (1981). Partial purification and some properties of BB7.2. A cytotoxic monoclonal antibody with specificity for HLA-A2 and a variant of HLA-A28. *Hum Immunol* 3, 277-299.

Raman MC, Rizkallah PJ, Simmons R, Donnellan Z, Dukes J, Bossi G, Le Provost GS, Todorov P, Baston E, Hickman E, *et al.* (2016). Direct molecular mimicry enables off-target cardiovascular toxicity by an enhanced affinity TCR designed for cancer immunotherapy. *Sci Rep* 6, 18851.

Rebai N, Mercier P, Kristensen T, Devaux C, Malissen B, Mawas C, Pierres M (1983). Murine H-2Dd-reactive monoclonal antibodies recognize shared antigenic determinant(s) on human HLA-B7 or HLA-B27 molecules or both. *Immunogenetics* 17, 357-370.

Smith SN, Wang Y, Baylon JL, Singh NK, Baker BM, Tajkhorshid E, Kranz DM (2014). Changing the peptide specificity of a human T-cell receptor by directed evolution. *Nat Commun* 5, 5223.

Spits H, Keizer G, Borst J, Terhorst C, Hekman A, de Vries JE (1983). Characterization of monoclonal antibodies against cell surface molecules associated with cytotoxic activity of natural and activated killer cells and cloned CTL lines. *Hybridoma* 2, 423-437.

Stewart-Jones G, Wadle A, Hombach A, Shenderov E, Held G, Fischer E, Kleber S, Nuber N, Stenner-Liewen F, Bauer S, *et al.* (2009). Rational development of high-affinity T-cell receptor-like antibodies. *Proc Natl Acad Sci U S A* 106, 5784-5788.

Stone JD, Kranz DM (2013). Role of T cell receptor affinity in the efficacy and specificity of adoptive T cell therapies. *Front Immunol* 4, 244.

van den Berg JH, Gomez-Eerland R, van de Wiel B, Hulshoff L, van den Broek D, Bins A, Tan HL, Harper JV, Hassan NJ, Jakobsen BK, *et al.* (2015). Case report of a fatal Serious Adverse Event upon administration of T cells transduced with a MART-1 specific T cell receptor. *Mol Ther*.

Zhou XX, Zeng WF, Chi H, Luo C, Liu C, Zhan J, He SM, Zhang Z (2017). pDeep: Predicting MS/MS Spectra of Peptides with Deep Learning. *Anal Chem* 89, 12690-12697.