Genotype-Specific Urinary Peptides as a Molecular Basis for Olfactory Recognition of Individuality and Genomic Relatedness in Mice

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

Dipl.-Biol. Theo Sturm

aus Pirna

Tübingen

2013

Tag der mündlichen Qualifikation:	30.07.2013
Dekan:	Prof. Dr. Wolfgang Rosenstiel
1. Berichterstatter:	Prof. Dr. Hans-Georg Rammensee
2. Berichterstatter:	Prof. Dr. Stefan Stevanović
3. Berichterstatter:	Prof. Dr. Michael Reth

Table of contents

1. Introduction

Everybody lives in a different world of consciousness and feelings. Which parts of the complex world become somebody`s own world depends on the power of senses and on the power of mind as well as on the external circumstances one lives in. Generally, it can be assumed that humans are able to realise more aspects of the real world than animals. In some areas, however, humans are easily outperformed by animals. This is perfectly true for olfaction. Mice and many other mammals can readily discriminate individuals by sniffing[1,2], an ability that humans nearly lack and rarely make use of[3,4]. Indeed, olfaction appears to be the dominant sense in most mammals apart from primates[2,5]. Mice are able to smell the sex[6-8], reproductive status[6], competitive status[9], health status[10-12] and stress status[13,14] of conspecifics[2,15], to mention just a few examples. Even more remarkably, mice can gather most of this information not only when encountering other mice but even when just sniffing their urine. Furthermore, in addition to genome-related and physiology-dependent information, mice gain insight into the temporal dimension - they can recognise the age of urine scent marks and thereby estimate the time at which a given conspecific stayed at a particular place[15-18]. The realm of information that mice can extract from the odour of urine and other body secretions largely determines their social behaviour including territorial scent marking[16], aggression[19], mate choice[20] and interspecies defensive behaviours[21]. In the following sections, I will introduce this fascinating topic in more detail. Because the present thesis deals with genotype-related individual differences in urinary peptides including MHC (major histocompatibility complex)-related ones, I will also cover some general aspects of urine, genome variation and the MHC. As humans, we have the great pleasure to be able to apply science to gain insights into worlds that are not amenable to our own senses.

1.1. Urine

1.1.1. The formation of urine

Although mouse urine is an important medium for scent marking and contains a multitude of olfactory cues relevant in social communication[4], the main reason for urine formation is the excretion of metabolic end products and the maintenance of the osmotic balance in the body. In the glomeruli of the kidney, the capillary pressure of the blood leads to the ultrafiltration of blood plasma. The filter is composed of four layers. The first part, situated at the blood side of the glomeruli, is formed by the endothelial cell surface layer, a special type of extracellular matrix. Numerous open fenestrae in the endothelial cell layer constitute the second component of the sieve. The core of the filter comprises the basal lamina, a specialised type of extracellular matrix which is also known as the glomerular basement membrane in the context of the kidney[22,23]. On the urine side of the glomeruli, the filter is

completed by a layer of podocytes with interdigitating processes that spare small slits (known as "slit membrane") through which the filtrate can pass. This complex glomerular filter selects blood components according to size and charge[22]. The molecular weight cut-off is about 70 kDa[24]. However, negatively charged molecules, like serum albumin at physiological pH, are retained even at smaller sizes. In contrast, positively charged molecules are more prone to pass the filter also at slightly higher molecular weights[22]. The originating filtrate of blood is called primary urine. It flows from the Bowman`s capsule, which surrounds each glomerulus, via the proximal tubule, the loop of Henle, the distal tubule and the collecting tubules into the bladder. Within the proximal tubule, abundant plasma proteins that escaped retention by the filter, e.g. serum albumin[25,26], immunoglobulin light chains[27], haemoglobin[28] and myoglobin[29], are reabsorbed by endocytic receptors like megalin and cubilin[30]. The brush border membrane of the proximal tubule also expresses the peptide transporter PEPT2 which transports virtually all kinds of di- and tripeptides from the primary urine back into the kidney tissue[31]. On the way to the bladder, many more valuable molecules including glucose and amino acids are reabsorbed nearly completely. A large proportion of the water, the Na^+ ions and the Cl^- ions is also regained[32]. Lactotransferrin, which is mostly just called lactoferrin, is excreted and later reabsorbed in the distal collecting ducts of the kidney medulla. This protein chelates free Fe^{3+} ions thereby regaining iron for metabolic use and depleting it from urine to inhibit microbial growth[33,34, (Dipl)].

Reabsorption is not the only process that converts primary urine into the final urine voided from the bladder. The kidney brush-border membrane of mice contains large quantities of an active metalloendopeptidase, namely meprin A (Enzyme Commission (EC) number 3.4.24.18, ref. [35]), which is expressed only at low amounts in humans. Mice possess a secreted form of active meprin A in urine constituting the major endopeptidase in mouse urine[36,37]. Because meprin A will have an important role in later sections of this thesis, I will introduce it more thoroughly in the following. Meprin A is encoded by two different genes, *mep1a* coding for meprin A α, and *mep1b* coding for meprin A β. Meprin A assembles into disulfide-linked dimers. Several dimers can associate noncovalently to form oligomers. In the kidney brush-border membrane, meprin A β_2 homodimers and meprin A $\alpha_2\beta_2$ as well as $\alpha_3\beta$ tetramers exist. Urine contains homo-oligomers composed of 10 to 100 subunits of meprin A α building the largest protease described so far[38]. Meprin A β prefers to cleave N- or C-terminal of acidic amino acid residues, whereas homo-oligomeric meprin A α avoids cleavages N-terminal of glutamic acid and is prone to cleave N-terminal of the sequence XP, where X represents any amino acid and P symbolises proline[39,40].

1.1.2. The composition of urine of *Mus musculus*

The most striking feature of mouse urine in comparison to human urine is its huge protein concentration. While human urine contains only about 0.05 g/l of protein[41], this value is 100-fold

higher in male mice[42]. The difference is caused by the excretion of major urinary proteins (MUPs) by *Mus musculus*. MUPs constitute between 65% [43] and more than 99% [44] of the total urinary protein in male mice. Female urine contains only a quarter to a third of the MUP concentration present in male urine[45]. The MUPs found in urine are produced in the liver and are excreted via the kidney. MUPs have a molecular weight of approximately 19 kDa and are members of the lipocalin family of proteins[15]. Their importance for recognition of genomic identity is discussed in section "1.6.2. Recognition of genomic identity based on non-MHC genes".

Another dominant protein of mammalian urine is uromodulin, also known as Tamm-Horsfall urinary glycoprotein. It is present in a filamentous form and acts against urinary tract infections[46], e.g. by reducing the adsorption of fimbriated uropathogenic *Escherichia coli* to the urothelium[47,48]. Uromodulin promotes the production of inflammatory molecules in monocytes[49]. It induces the maturation of dendritic cells via toll-like receptor 4 (TLR-4) dependent signalling thereby supporting a T_H1 immune response[50,51]. It has been proposed that the immunostimulatory capabilities of uromodulin might become relevant when fimbriated uropathogenic bacteria covered with uromodulin start to enter the urothelium[51]. Apart from uromodulin, there are several other proteins which are present in urine not for disposal but which are intentionally excreted to exert a special function. For example, the abundant epidermal growth factor (EGF) is secreted in an active form[52] from the renal tissue to the urine[53-55], probably to support the growth of the urothelium [(Dipl)].

A special feature of urine is the presence of exosomes. Exosomes are tiny vesicles with a diameter between 40 and 80 nm which are secreted from the renal epithelial cells into urine. Many proteins of urine can only be found in the exosomes[56, (Dipl)].

Urine contains a multitude of peptides[57-59] as analysed in the present study for *Mus musculus*. A realm of small molecules is also present, e.g. urea as an end product of amino acid metabolism and the typical yellow dye of urine, urobilin, which results from the degradation of the red haem prosthetic group of haemoglobin[60]. Urine is rich in salts, for example sodium chloride and calcium oxalate[42], and contains a plethora of small organic volatiles[61-63, (Dipl)].

Urine is not a static mixture, but its composition changes over time. Whereas the proteome of living cells and blood plasma must be tightly regulated to ensure proper function, most urinary proteins are present in urine just for waste disposal, not being worth energy investment for reabsorption. Therefore, the intra- and inter-individual variability of the proteome is expected to be higher in urine than in cells or blood plasma[64]. Some components vary with the circadian clock as has been demonstrated for MUP2 [65]. Others change according to androgen levels[15,66], social status[15,67], stress[13,14,68] or health status[69], to mention just a few factors. As many readers will have experienced with their own urine, urinary components dependent on certain foods can even be detected by the human nose, e.g. after consumption of asparagus or popcorn.

Particularly profound changes of urine composition occur during urinary tract infections. Urine is usually sterile before voiding from the bladder[70]. Upon infection, pathogens release several compounds into the urine; for example, bacteria can secrete siderophores, which are small organic chelators of Fe^{3+} ions, to promote their own supply with iron[71]. The host organism consequently releases a multitude of molecules into the urine to fight the infection. Among these is the neutrophil gelatinase associated lipocalin (NGAL) binding bacterial siderophores for iron sequestration as a mechanism of innate immune defence[70,72-74]. Urinary tract infection boosts the secretion of antimicrobial peptides into the urine, namely that of cathelicidin[75], certain β-defensins[70] and hepcidin[76].

1.2. Genomic diversity

Mutations can be considered as the "motor of evolution" and continuously occur in every cell. Accordingly, genetic differences are documented even for closely related substrains of laboratory inbred mice[77]. In wild populations, genetic diversity is essential to adapt to variable or new environmental conditions. Genomic diversity can be exploited in biological research to identify genes associated with a specific genotype. For example, the comparison of olfactory receptor (OR) genes between people with a specific anosmia and control persons capable of smelling the respective substance has revealed several odorant-OR-pairs[78].

According to the size of the DNA sequence affected, genomic diversity can be classified into three main groups. Sequence alterations of a size above 3 Mbp are defined as microscopic structural variants (SVs), whereas those of a size ranging from 100 bp or 1 kbp to 3 Mbp are termed submicroscopic SVs[79,80]. Variations affecting less than 100 bp or 1 kbp are grouped as small-scale[80,81].

A major group of SVs are the copy-number variants (CNVs) which arise from insertions, deletions or duplications of DNA segments longer than 1 kbp. Inversions and translocations are other important forms of SVs[79].

Small-scale variation is classified into three groups. First, there are the base pair substitutions. If a single base pair is exchanged, this is called single nucleotide variation (SNV). Sometimes, short sequences of base pairs get substituted by the mechanism of gene conversion[82]. Second, the DNA sequence can be changed by deletions, i.e. the removal of one or more nucleotides. The addition of one or more nucleotides to the DNA is called insertion and represents the third possibility[81].

The diverse types of genomic variation occur with very different frequencies. For *Mus musculus*, Keane *et al.*[80] identified 56.7 million unique SNVs, 8.8 million unique insertions and deletions (< 100 bp) and 280.000 SVs.

The proportion of the protein-coding sequence in the *Mus musculus* euchromatic DNA is only 1.27%[83]. Consequently, most genomic variation does not affect protein sequences. Nevertheless, mutations in non-protein-coding DNA can alter protein sequences in rare cases with splice site mutations being a prominent example. Variations in promoter, enhancer, silencer or polyadenylation signal sequences are instances that can affect protein abundance. On average, however, mutations in coding DNA sequences affect the phenotype more severely than those in non-coding regions, and they are therefore counterselected more intensively in evolution resulting in reduced genomic variability in coding versus non-coding sequences[81]. The extent of genomic variation also differs between different protein-coding genes. The most polymorphic genes of *Mus musculus* are the MHC genes[84] (see section "1.3. MHC molecules"). In contrast, genes encoding proteins with very basic cellular function, e.g. actins, histones and ribosomal proteins, are very conserved[81].

I will now focus on genomic variability affecting protein sequences. Within the protein-coding DNA, SNVs can be grouped as follows[81,85]:

- Synonymous: the mutation does not alter the amino acid sequence.
- Non-synonymous: the mutation alters the amino acid sequence. Non-synonymous SNVs occur in two types:
    - Missense: the mutation leads to a single amino acid variation (SAV).
    - Nonsense: the mutation leads to the generation of a stop codon. Such mutations are also called stop codon gain mutations.

Missense mutations play a special role for the present thesis (see section "3.2. The urinary peptidome comprises a respectable set of peptides with single amino acid variations (SAVs)"), and I will therefore introduce them in more detail in the following. Missense mutations can be subdivided into conservative and non-conservative. In a conservative missense mutation, a given amino acid is substituted by another amino acid having similar biochemical properties. An example is the leucine/isoleucine SAV. Conservative SAVs are favoured by the design of the genetic code. In non-conservative missense mutations, the biochemical properties of the exchanged amino acids differ profoundly. Non-conservative SAVs are typically discouraged by the design of the genetic code, and they are usually strongly counterselected in evolution. Due to its ability to form disulfide bonds, which are often crucial for the stabilisation of the global protein structure, the most conserved amino acid is cysteine[81,86].

Within protein-coding DNA, the insertion or deletion of a number of nucleotides that is unequal to three or a multiple of three leads to a frameshift mutation altering the reading frame used during translation. This usually results in a new stop codon upstream of the original one, and the incompletely translated mRNA is typically recognised and degraded in a process known as nonsense-mediated mRNA decay[85,87].

1.3. MHC molecules

Most major histocompatibility complex (MHC) proteins are encoded within a single DNA region comprising about $4 \cdot 10^6$ base pairs. This gene cluster is called MHC in all jawed vertebrates and is also known as the *H2* in *Mus musculus*. The *H2* region is situated on chromosome 17. MHC molecules can be divided into two major groups, MHC class I and MHC class II proteins[88,89]. Many MHC proteins can bind peptides with certain characteristic sequence motifs within their special peptide binding grove[90,91]. This MHC binding motif is less strict in the case of the classical MHC class II proteins of *Mus musculus*, H2-A and H2-E, as compared to classical MHC class I molecules. Furthermore, in contrast to MHC class I proteins, the length of bound peptides is highly variable even for the same allotypic MHC class II molecule complex[92-94]. Therefore, MHC class II peptide ligands provide only a blurred signature of an individual`s MHC class II genes and inherently appear far less suitable than MHC class I peptide ligands to signal the MHC type in behavioural situations. Accordingly, theoretical concepts[95] and experimental work[96-100] have focused on MHC class I peptides as the relevant molecules for signalling of MHC type in social interactions. Consequently, I only searched for MHC class I-dependent and not for MHC class II-dependent urinary peptides during my PhD thesis, and I will therefore only introduce MHC class I proteins, both classical and non-classical ones, in more detail in the following [(Dipl)].

1.3.1. Classical MHC class I molecules

Classical MHC class I molecules, also known as MHC class Ia molecules, comprise H2-K, H2-D and H2-L in *Mus musculus*. They are present on almost all nucleated cell types, but their expression is particularly high on T cells, neutrophils and professional antigen presenting cells (APCs, i.e. dendritic cells, B cells and macrophages)[101]. The main function of classical MHC class I molecules is the presentation of peptides derived from intracellular proteins on the cell surface. This enables $CD8^+$ cytotoxic T lymphocytes (CTLs) to continuously check the health status of the cells. If appropriately activated CTLs recognise an MHC class I peptide ligand as non-self or strange, e.g. because it is derived from a virus or a mutated tumour protein, they can kill the respective cell [(Dipl)].

H2-K, H2-D and H2-L each represent the α-chain of a respective MHC class I molecule complex. This polymorphic α-chain is noncovalently associated with a non-polymorphic 12 kDa subunit, the β_2-microglobulin (β_2m). The two N-terminal domains of the α-chain (α1 and α2) form a peptide binding groove. The walls of this furrow consist of two α helices, whereas the bottom is composed of a β pleated sheet (Figure 1-1). The groove is about 2.5 nm long and contains special binding pockets which interact with certain amino acid side chains of the bound peptide. Different MHC class Ia alleles usually encode different amino acids in the binding pockets defining their binding properties.

Furthermore, MHC class I molecules bind the N-terminus and the C-terminus of the peptide ligand, which is largely independent of its sequence[102]. Thereby, each allotypic α-chain binds peptides according to its characteristic MHC motif determining peptide length (usually 8-11 amino acids) and certain amino acids at two or three defined positions, called anchor residues[91-94,103]. Peptide amino acid residues at auxiliary anchor positions can further increase the stability of the interaction with the MHC[104], but MHC bound peptides are otherwise variable. If the peptide ligand is slightly longer than optimal for the corresponding MHC class I molecule, it can happen that it forms a small arch extending from the MHC binding groove[105,106], which still binds the N-terminus and the C-terminus as usual. Alternatively, it has been reported that overlong MHC class I peptide ligands overhang at one end of the groove[107, (Dipl), (*Nat Comm*, Theo)].

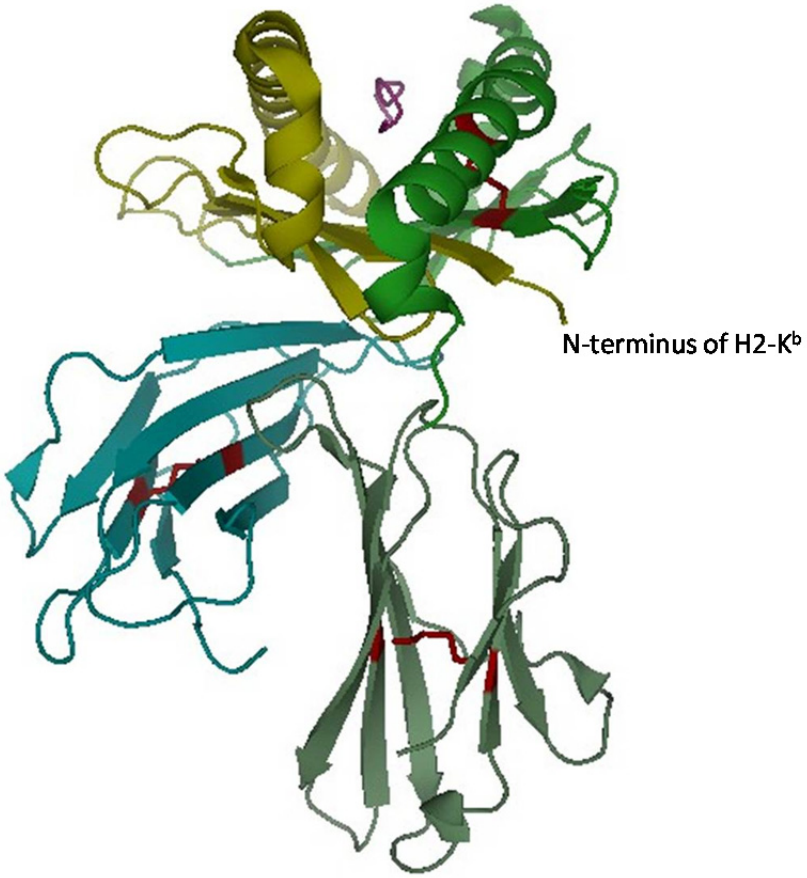


Figure 1-1 | Schematic representation of the crystal structure of H2-K^b in complex with the β_2m subunit and the peptide RGYVYQGL derived from the vesicular stomatitis virus (VSV). The H2-K^b chain contains an α1-domain (yellowish-green), an α2-domain (rich green) and an α3-domain (grey). *In vivo*, the α3-domain is connected to a transmembrane domain and a cytoplasmic tail comprising 23 and 41 amino acids respectively[108]. The transmembrane domain and the cytoplasmic tail are not contained in the resolved crystal structure and therefore not depicted. H2-K^b is noncovalently associated with the β_2m subunit (blue) and a peptide consisting of eight to nine amino acids (violet). The β_2m subunit, the α2-domain and the α3-domain each contain a disulfide bridge, which is labelled in red. I generated the figure during my diploma thesis using the software PyMOL[109] and structural data deposited in the Protein Data Base[110] (entry no. 1kpu, Rudolph and Wilson[111]) and the UniProtKB/SwissProt database[108] (entry no. P01901) [(Dipl)].

1.3.2. Non-classical MHC class I molecules

The fundamental molecular structures and functions of non-classical MHC class I molecules, also known as MHC class Ib molecules, are much more diverse than is the case for classical MHC class I proteins. Not all MHC class Ib proteins associate with the β_2m subunit, some do not bind peptides but rather other molecules like fatty acids or immunoglobulin G, some are devoid of a permanently bound ligand[112, (Dipl)].

An example for an MHC class Ib molecule which is evolutionary relative closely related to MHC class Ia molecules is Qa-2. Like MHC class Ia molecules, Qa-2 associates with the β_2m subunit and binds peptides of a defined length (9 amino acids) according to clearly defined anchor residues[113]. Consequently, I also searched for urinary peptides with Qa-2 binding motif during the present study. Qa-2 molecules with identical peptide binding specificities are expressed in most tissues of *Mus musculus* regardless of whether the *H2*^{b} or the *H2*^{d} haplotype is present. About 50% of Qa-2 molecules are attached to the plasma membrane via a lipid anchor, whereas others occur in a soluble form[112,114]. Qa-2 presents peptides for T cells, and it also binds to an NK (natural killer) cell receptor[112, (Dipl)].

Another type of *Mus musculus* MHC class Ib molecules that is of particular interest with regard to olfaction are the members of the H2-M10 family and the H2-M1 family. They are associated with β_2m and are expressed in the basal, Gα_O positive vomeronasal sensory neurons (VSNs). H2-M10 molecules were not detected in tissues other than the vomeronasal organ (VNO), and they bind to certain V2R receptors (see section "1.5. The olfactory system of *Mus musculus*"). In fact, they are essential for the expression of these V2Rs at the cell surface[21]. Therefore, it is assumed that they act as coreceptors or chaperones for V2Rs[115]. It has been demonstrated that the MHC binding groove of M10.5 molecules is present in an open conformation, that it is empty and does not bind to MHC class Ia peptide ligands[116]. Notably, peptides are detected by basal VSNs which express V2Rs[96, (Dipl)].

1.3.3. MHC class I antigen processing

MHC antigen processing is a term for all processes that are necessary to present an MHC peptide ligand at the cell surface starting at the level of proteins destined for degradation. There are two major types of substrates for MHC class I antigen processing. First, these are proteins that have served their time and got polyubiquitylated. Second, newly synthesised, but misfolded proteins, the so-called defective ribosomal products (DRiPs)[117], are estimated to constitute up to 30% of the total input to the MHC class I processing pathway[89,118,119]. These proteins are first unfolded and then proteolysed by proteasomes in the cytosol. Proteasomes have three distinct cleavage specificities: a chymotryptic one, a tryptic one and a V8-like, the latter cleaving C-terminal of aspartate and glutamate. The peptide

products leaving the proteasome can be transported by the TAP (transporter associated with antigen processing) complex into the endoplasmic reticulum (ER). The *Mus musculus* TAP has a preference for peptides with a hydrophobic C-terminus paralleling the binding preferences of MHC class Ia molecules of this species. Within the ER, peptides can be shortened at their N-termini by the action of ERAAP (endoplasmic reticulum aminopeptidase associated with antigen processing)[101,120]. The ER membrane contains newly synthesised MHC class I molecules. After these have bound β_2m, they can associate with the TAP complex via the tapasin protein. The MHC-TAP-complex is stabilised by the chaperones calreticulin and ERp57 enabling loading of the empty MHC groove with peptides corresponding to its MHC binding motif. The proteins of the MHC class I peptide loading complex ensure that peptides with low affinity for the respective MHC class I molecule dissociate, so that better fitting peptides can bind instead. The trimeric complex of MHC α-chain, β_2m and peptide then leaves the ER and is transported via the Golgi apparatus to the cell surface[101, (Dipl)].

In general, any kind of protein which is translated within a cell can give rise to peptides presented on MHC class I molecules of that cell. Dendritic cells have the remarkable capability to additionally present peptides from proteins that they do not translate themselves on their MHC class I molecules. This process is known as cross-presentation and is accomplished by the uptake of proteins from the extracellular space. The ingested proteins can gain direct access to the ER[121]. Subsequently, they can be transported from the ER to the cytosol where they are degraded by the proteasome and embarked to the usual MHC class I processing pathway[122]. In an alternative, TAP-independent route of the cross-presentation process, MHC class I proteins are loaded with peptides derived from extracellular proteins within the acidic MHC class II compartment (MIIC)[123,124].

1.4. Olfactory cues in *Mus musculus*

Mice are able to smell a multitude of chemically diverse molecules ranging from small organic volatiles[4,14,66,125] to peptides[96-98,126,127] to proteins[8,19,128]. These olfactory cues can be functionally grouped into at least three categories: pheromones, signature mixtures and kairomones[1,5].

Pheromones

Many olfactory cues of mammals are often falsely referred to as “pheromones” being in conflict with the original definition of a pheromone as a molecule that is an “evolved signal”, which is “emitted by an individual and received by a second individual of the same species, in which it causes a specific reaction, for example, a stereotyped behaviour or a developmental process”[1,129]. In mice, like in other mammals, pheromones appear to be very rare with MUP20 being a prominent exception[1]. MUP20, also known as darcin, is a MUP occurring in male but not female mouse urine. It triggers the attraction of females towards urine of males as compared to that of females, and MUP20 is as

effective in doing so as whole male urine. Furthermore, MUP20 induces the formation of a memory for the individual male`s volatile odour in the female`s brain, so that the female will later remember that the particular volatile odour belongs to a male. Only after female mice have learned this association between the individual volatile odour and the male identity, they show a preference for the respective male volatile odour versus a female volatile odour[8]. A single exposure to MUP20 can also induce spatial learning, i.e. competitor males and females are triggered to spend more time at places where they encountered MUP20 even after removal of the scent[130].

Signature mixtures

Most olfactory cues of *Mus musculus* are not pheromones but components of complex "signature mixtures". As defined by Tristram Wyatt, signature mixtures represent "variable chemical mixtures (a subset of the molecules in an animal's chemical profile) learned by other conspecifics and used to recognize an animal as an individual or as a member of a particular social group such as a family, clan or colony"[1]. Each mouse produces several signature mixtures, for example the body odour is distinct from the odour of the urine[61].

Mammalian volatiles are the result of complex metabolic pathways, and therefore they reflect a multitude of physiological and environmental conditions. For example, socially subordinate male mice produce less farnesenes than dominant male territory owners[67] in order to be tolerated by the latter[15]. Another example are sulfated glucocorticoids whose urinary concentrations are elevated many-fold in stressed versus unstressed mice and which act as prominent olfactory cues in the VNO[14,131]. The metabolic pathways producing volatiles depend on many different genes and their complex interactions. Nevertheless, within a short period of time, volatiles can be used to identify an individual[8]. However, volatiles do not constitute a constant genetic signature.

In contrast, proteins and peptides longer than about six amino acids can usually be assigned to a single gene or a small set of homologous genes[132]. From a theoretical point of view, this makes them much more suitable than volatiles for signalling genetic identity in a relatively constant fashion. The signature mixture of urinary peptides will be closely investigated in this thesis. There are only two protein families that have been described to act as olfactory cues in mice as intact proteins, and both families show a mouse strain-specific expression pattern. First, these are the MUPs, which we already got to know as major constituents of mouse urine in section "1.1.2. The composition of urine of *Mus musculus*". Second, exocrine gland-secreting peptides (ESPs, which one should better refer to as "proteins" because their predicted molecular weight ranges from 5 to 15 kDa) also act as olfactory stimuli in the vomeronasal organ. I will introduce the importance of MUPs and ESPs for recognition of genomic identity in section "1.6.2. Recognition of genomic identity based on non-MHC genes".

Kairomones

Another functional type of olfactory cues in *Mus musculus* are kairomones, which are defined as "cues transmitted between species that selectively disadvantage the signaller and advantage the receiver"[5,128]. Remarkable examples of kairomones acting in favour of *Mus musculus* are the fear-evoking odours of mouse predators[133]. Snake skin, cat fur and rat urine all induce defensive behaviour even in naive laboratory mice that have never before encountered a predator. Two fear-inducing kairomones were molecularly identified as homologs of *Mus musculus* MUPs, namely rat MUP13 present in rat urine and MUP-Feld4 produced in the submandibular salivary glands of cats and naturally distributed via the saliva onto the fur[128].

1.5. The olfactory system of *Mus musculus*

Mus musculus has a very powerful sense of smell, which is based on several olfactory subsystems. The two most important subsystems are the main olfactory epithelium (MOE) and the vomeronasal organ (VNO). Furthermore, the olfactory system of *Mus musculus* comprises a septal organ of Masera[134] and a Grueneberg ganglion[135,136], but these two structures have not yet been described to recognise peptides, and they are therefore not further considered in this thesis[125,137, (Dipl)].

The MOE is the most important olfactory subsystem of mammals including mice and humans. It contains a few million olfactory sensory neurons (OSNs)[137] expressing about 1300 different types of olfactory receptors (ORs) in *Mus musculus*[138]. The MOE can be divided into several MOE-subsystems, e.g. the OR37 subsystem and the GC-D (guanylyl cyclase type D) cell system, the latter being characterised by a distinct signal transduction system lacking the G-protein G_{Olf} and the type III adenylyl cyclase typical for the other OSNs (see below)[137]. The axons of the MOE converge in the main olfactory bulb (MOB)[125, (Dipl)].

Like most mammals, but in contrast to humans, *Mus musculus* has a VNO[139]. It is situated at the anterior basis of the nasal septum and consists of two bilaterally symmetric, blind ended tubes. Each tube contains a lateral blood vessel whose contractions provide a pumping mechanism to suck up olfactory cues[125]. The VNO comprises about 300.000 vomeronasal sensory neurons (VSNs)[98]. The sensory epithelium of the VNO is separated into an apical and a basal layer. VSNs of the apical layer each express one out of 137 functional V1R receptors, and their axons project to the rostral part of the accessory olfactory bulb (AOB)[125,140]. In contrast, the axons of VSNs of the basal layer are connected with the caudal part of the AOB[125]. At least three different groups of proteins are known or assumed to directly bind olfactory cues in the basal VSNs of *Mus musculus*: family ABD V2R receptors (115 genes), family C V2R receptors (7 genes) and H2-M10 MHC class 1b molecules (9 genes). Members of these three groups are coexpressed in basal VSNs in a non-random manner[131]. Peptides,

MUPs and ESP1are detected by V2R expressing basal VSNs[19,96,141]. Interestingly, vomeronasal receptors (VRs, comprising V1Rs and V2Rs) are highly polymorphic within *Mus musculus*, and they differ even between laboratory inbred strains. Coding sequence variations were found to be more than twice as frequent in VR genes as compared to the genome average. This might explain some of the behavioural differences observed between laboratory inbred strains as well as between laboratory and wild mice[142]. Some VSNs express formyl peptide receptors (FPRs), which are supposed but have not yet been demonstrated to be involved in the detection of formyl methionine peptides by VSNs[126, (Dipl)].

The functions of the VNO and the MOE appear to be partly overlapping. There are both volatiles, e.g. 2-heptanone, and nonvolatile molecules, e.g. peptides, that can be detected by the VNO as well as by the MOE[96,97,125]. However, some nonvolatile substances, for example ESPs and MUPs, can only be detected by the VNO but not by the MOE[19,143]. On the other hand, there are many molecules, e.g. (methylthio)methanethiol (MTMT), that lack receptors in the VNO but are potent ligands for ORs in the MOE[125, (Dipl)].

In both OSNs and VSNs, the efficient binding of an olfactory cue to its chemosensory receptor(s) results in the depolarisation of the plasma membrane via the influx of cations into the cytosol. ORs, VRs and FPRs are G protein-coupled receptors (GPCRs) characterised by the typical seven transmembrane helices and the interaction with trimeric GTP-binding proteins (G proteins)[125,127,131]. Despite these basic similarities, the signal transduction mechanisms differ substantially between OSNs and VSNs even in response to the same ligand. In OSNs, triggering of ORs activates the G protein subunit $G\alpha_{Olf}$ which stimulates type III adenylate cyclase to convert adenosine triphosphate (ATP) into cyclic adenosine monophosphate (cAMP). cAMP results in the opening of cyclic-nucleotide-gated (CNG) cation channels in the plasma membrane mediating the depolarisation[125]. V1Rs and four VNO FPRs are associated with the $G\alpha_{i2}$ G protein subunit, whereas V2Rs and VNO FPR-rs1 interact with the $G\alpha_O$ G protein subunit [125,127]. Upon VR activation, $G\alpha_{i2}$ and $G\alpha_O$ G proteins stimulate phospholipase C-β (PLC-β) to cleave phosphatidylinositol-4,5-bisphosphate (PIP_2) into inositol-1,4,5-trisphosphate (IP_3) and diacylglycerol (DAG). DAG activates the transient receptor potential channel 2 (TRPC2) enabling influx of Na^+ and Ca^{2+} ions into the cell[125,131, (Dipl)].

1.6. Social behaviours depending on genomic identity

Many social behaviours of animals are crucially influenced by the genomic identity of the interacting individuals. Depending on the behavioural context, animals can interpret the signals of genomic identity either as related with a specific individual or merely use them as indicators of genomic relatedness[144]. In some situations, sensing of the degree of genomic relatedness would already be sufficient from a theoretical point of view. Such behaviours include cooperative interactions among

kin and between parents and their offspring as well as inbreeding avoidance. In other behaviours, the degree of genomic relatedness alone is not sufficient, but information about the genomic individuality is required instead[144]. For example, in territorial scent marking, a mouse should be able to discriminate territories of individual mice and not just territories of kin versus non-kin. One of the best studied paradigms of individual recognition in mice is the pregnancy block effect, which is also known as the Bruce effect in honour of Hilda Bruce who first described it in 1959[145]. The principle is as follows. A female mouse is mated with a male mouse and during the mating the female forms a memory for the individual odour of that male. When the inseminated female is exposed to a different male or its urine shortly after fertilisation, the pregnancy is aborted and the female returns to oestrus 4 to 5 days after the original copulation[139,146]. However, male mice need to be genetically different from the mating male in order to induce the pregnancy block, i.e. males of the same inbred strain are not discriminated by the pregnant female[96,99,146]. Consequently, the pregnancy block effect represents an excellent experimental setup for testing the recognition of genomic individuality in mice.

1.6.1. Recognition of genetic identity based on MHC type

The hypothesis that the MHC might serve as an olfactory signal of individuality has first been proposed by Lewis Thomas in 1975 based on theoretical considerations about the possible biological relevance of the extreme MHC polymorphism[84] in most jawed vertebrates[147]. Working with MHC congenic mouse strains, Thomas`s colleagues observed marked preferences of mice to mate with mice of specific MHC types[20]. In a systematic study involving three pairs of MHC congenic inbred strains (BALB/c ($H2^d$) versus BALB.B ($H2^b$), B6 ($H2^b$) versus B6-$H2^k$, B10 ($H2^b$) versus B10.A ($H2^a$)), Yamazaki *et al.*[20] tested six different mating constellations. Mice had the choice to mate either with animals of their own strain or with MHC dissimilar congenic individuals. In four out of the six comparisons, mice of a given strain preferred to mate with MHC dissimilar partners. This is called MHC disassortative mating. A strain preference to mate with MHC identical individuals, i.e. MHC assortative mating, was observed in only one constellation. In one comparison, the preference for mating with mice of self or non-self MHC type was only consistent for individual mice but inconsistent within the strain[20,148]. In the context of the pregnancy block test, male mice that are MHC congenic but MHC dissimilar with respect to the mating male are recognised as strange by the female leading to abortion[149].

At least 25 different species of mammals, birds, reptiles, amphibians and fish have been investigated for MHC related behaviours. While mating preferences have been reported to be MHC-disassortative in most of these species[144], MHC-assortative mating has been found to occur in the malagasy giant jumping rat (*Hypogeomys antimena*)[150] and the tiger salamander (*Ambystoma tigrinum*)[151]. Studies in three species including domestic sheep (*Ovis aries*)[152] have not observed any MHC preference during

mating[144]. *Mus musculus* is not the only species with conflicting reports about MHC mating preference. In humans, results of studies on MHC mating preferences vary depending on the population studied. MHC-disassortative mating has been reported from a Hutterite[153] and a Mormon[154] population. In contrast, no evidence for an MHC-dependent preference was detected in Amerindian tribes[155] and in the Yoruba[154] population.

Many social behaviours including cooperative behaviour among kin and between parents and their progeny, territorial scent marking, inbreeding avoidance and mating male recognition in the context of the pregnancy block have been specifically ascribed to the MHC[148] (reviewed in Ruff *et al.*[144]). However, the mentioned examples do *a priori* just require a signal of genomic relatedness or individuality but do not need to match the MHC. Note that some of these behaviours also occur in invertebrates, which lack the MHC. Mate choice of jawed vertebrates is a different case because in this setting, selection of partners with certain MHC alleles really makes sense from a theoretical point of view as detailed in the following.

MHC-disassortative mating inevitably leads to MHC heterozygous offspring. Consequently, the progeny have the possibility to take advantage of the immunological capabilities of more different MHC molecules than in a homozygous setting because MHC molecules are codominantly expressed. On the other hand, the expression of too many different MHC molecules increases the number of T cells that are negatively selected during T cell development thereby limiting immunocompetence[156]. Consequently, there is an optimum for the number of different MHC alleles in an individual, and this optimum is also the reason why mice have two (*H2-K* and *H2-D*, e.g. in $H2^b$ haplotype) or three (*H2-K*, *H2-D*, *H2-L*, e.g. in $H2^d$ haplotype) functional MHC class Ia genes and not one, four or more. In general, however, the increased number of different MHC alleles in MHC heterozygotes versus MHC homozygotes results in an immunological and therewith evolutionary advantage. Accordingly, due to the high polymorphism of MHC genes in most jawed vertebrates, MHC heterozygosity appears to be the rule in natural populations. Although MHC-heterozygosity is not superior to both corresponding homozygote constellations in every single infection, its benefit arises from challenges with multiple pathogens evolutionary optimised for different MHC alleles[144,157,158].

MHC-disassortative mating has been proposed to be evolutionary advantageous for an additional reason. Several pathogens rapidly mutate their CTL epitopes adapting to certain MHC alleles in order to escape MHC-mediated immunity, and this increases their pathogenicity in an MHC allele specific manner[159-161]. MHC-disassortative mating ensures that the combination of MHC alleles in the progeny differs from that of the parents as well as between mother and father. Therefore, a pathogen specifically adapted to the MHC type of one family member is probably less pathogenic in other family members decreasing the effectiveness of transmission within the family and reducing the probability that at all animals will be severely impaired at the same time. The diversity of the MHC types provides a so-called “moving target” for pathogens[144,162].

Although MHC-disassortative mating appears beneficial from an evolutionary point of view, the crucial question is, whether there is a specific MHC-dependent olfactory cue that influences mating preferences, or whether MHC heterozygosity of the progeny is achieved by indirect mechanisms. I will discuss this in section "4.2. Questioning the role of the MHC as an olfactory signal in social behaviours". In my opinion, Jane Hurst`s statement that "direct evidence for MHC-determined disassortative mating is surprisingly limited"[15] aptly summarises the present situation. In MHC congenic mouse strains, there is evidence not only for MHC-disassortative but also for MHC-assortative mating preferences depending on the strain combination and sex investigated. Sometimes, an MHC-related mating preference was not observed at all[20,163]. However, inbred mice might not reflect natural mating preferences of wild animals because these mice were selected for their willingness to mate with autosomally identical relatives for many generations[15,164]. Addressing this problem, Potts *et al.*[165,166] crossed laboratory and wild mice to generate mice with a hybrid genotype but controlled laboratory-derived MHC alleles. When these hybrid mice were allowed to mate freely, a relative lack of MHC homozygous offspring was observed as one would expect from MHC-disassortative mating. Unfortunately, variations in non-MHC loci that might have randomly been associated with mice of a given MHC type, e.g. MUPs, were not controlled in this setting[15].

Studies on MHC related mating preferences have usually been based on statistical comparisons of MHC allele dissimilarities within mating couples (sometimes measured indirectly via the offspring) versus random partner selection. However, statistical association does not yet provide a conclusive functional mechanism and some problems of the hypothesis of MHC-mediated behaviour will be addressed in section "4.2. Questioning the role of the MHC as an olfactory signal in social behaviours".

Two major mechanisms for explaining the functional link between the MHC as defined by the MHC class I molecules and relevant olfactory cues have been proposed in mice. In the first one, the MHC has been considered to influence the volatile composition of urine as demonstrated by habituation, training and other tests with MHC congenic or mutant inbred strains[167-173]. Analysis of urinary volatiles of single inbred mice showed considerable variation in the abundance of certain volatiles. However, no qualitative and only negligible quantitative variations in a large number of different volatiles have been observed when comparing urine mixtures of groups of an inbred wild type and a congenic knock-out strain lacking functional MHC class I molecules. In contrast, more quantitative differences were observed between groups of the same inbred wild type strain derived from different breeders[62] (see review[174]). Hence, whereas variability in small airborne organic molecules may be important for recognition of individual scent owners[8,15], physical evidence for an influence of the MHC on urinary volatiles remains elusive (*Nat Comm*, joint).

The second mechanism calls attention to the immunological function of the MHC class I molecules as peptide presenting proteins. As detailed in section "1.3.1. Classical MHC class I molecules", co-

expression of several MHC alleles together with their respective sets of bound peptides provides an immunological signature for each individual. This immunological signature must be interpreted by the olfactory system. It is now well established that, in freely behaving mice, nonvolatile peptides gain access to sensory neurons of both the main olfactory system and the VNO during behavioural situations involving direct physical contact[96,97], that such peptides are powerful ligands for subsets of VSNs and OSNs[96-98,131,175], that they induce brain activity downstream from the sensory neurons *in vivo*[176] and that synthetic MHC peptide ligands can be discriminated in social preference tests[97]. Synthetic MHC class I peptide ligands of non-mating male but not mating male MHC type have been demonstrated to induce the pregnancy block in the presence[96,99] or absence[99] of urine when applied at concentrations of about 10^{-4} M. This issue will be discussed in detail in section "4.2.1. The role of MHC-dependent peptides in social behaviours" (*Nat Comm*, joint).

1.6.2. Recognition of genomic identity based on non-MHC genes

There is clear evidence that urinary volatiles are strongly regulated by non-MHC genes[177], and that qualitative and quantitative differences in these volatiles increase with decreasing evolutionary relatedness[62]. As volatiles originate from complex metabolic pathways, each volatile appears to integrate information from many different genes. However, I am not aware of a study that causatively links a defined variation of a single gene to the abundance of a particular mouse volatile. In contrast, two gene families coding for proteinaceous olfactory cues have been suggested as MHC-independent signals of genomic identity in mice: MUPs[178] and ESPs[143] (already introduced in sections "1.1.2. The composition of urine of *Mus musculus*" and "1.4. Olfactory cues in *Mus musculus*"). Humans lack both MUPs[179] and ESPs[143].

The importance of MUPs for signalling of genomic individuality in mice is well established. MUPs act as a signature mixture for recognition of individuality in territorial scent marking[16,178,180], and they can also be used by females to discriminate males in the context of the pregnancy block[181]. For several reasons detailed in the following, MUPs are ideal olfactory cues of genetic individuality in *Mus musculus*:

1.) MUPs are encoded by a highly polymorphic family of at least 19 genes in *Mus musculus*[179].

2.) The polymorphism of the MUP gene family is reflected by abundant MUPs in the urine of *Mus musculus*[182-184] (cf. section "1.1.2. The composition of urine of *Mus musculus*"; some MUPs are expressed in the lachrymal glands, salivary glands, nose or mammary glands[15], but, in contrast to urine, their role as signals of individuality in tears, saliva and milk is not yet clear.)

3.) MUPs are stable in scent marks for many weeks[15].

4.) MUPs can bind volatiles in the central cavity of their barrel-shaped lipocalin protein structure[185-187]. This mechanism functions to extend the duration of volatile release from scent marks[44,188,189]. Different MUPs bind certain volatiles with different affinities, and thereby MUPs are able to create several distinct individual scent profiles from essentially the same original volatile profile[190-193].

5.) Not only the MUP-bound volatiles can be detected by mouse sensory neurons, but MUPs themselves activate specific V2R receptors in the VNO.[19] Because the polymorphism of MUPs mostly affects amino acid residues at the protein surface[184], it has been speculated that V2Rs might discriminate different MUPs directly[15].

Consequently, MUPs might signal genomic individuality in two different principal ways suggested for the generation of unique signature mixtures[95]. They reflect genomic individuality indirectly by forming "polymorphic transmitters"[95] for volatiles, but they might additionally act as signature mixtures themselves. As established during this thesis, a third mechanism of MUP-related signalling of genetic individuality might be via their degradation products, the MUP peptides (see sections "3.3. Further genomic variations influence urinary peptides" and "4.3. Genotype specific peptides offer new perspectives for olfactory research").

ESPs are less well characterised than MUPs, especially with respect to their potential role as signals of genomic identity. At least 24 different ESP genes and 14 apparent ESP pseudogenes have been identified in *Mus musculus*. In mice, the ESP gene cluster is located in close proximity to the MHC (see Figure 4-1), which is however not the case in *Rattus norvegicus*[143]. Kimoto *et al.*[143] compared 15 expressed ESP genes between six laboratory mouse strains, and found ESP34 and ESP36 to display some strain-specific sequence variation, with ESP36 being 24 amino acids shorter in BALB/c mice as compared to B6 mice. Additional quantitative differences in ESP expression were observed between the analysed laboratory mouse strains[143]. I would expect much more qualitative and quantitative individual differences of ESPs in wild mice or wild-derived inbred mouse strains (cf. Table 3-6 and Cheetham *et al.*[194]), because the mouse strains used by Kimoto *et al.*[143], like all classical laboratory mouse strains, originate from the same extremely small gene pool, and they are all derived from one common female ancestor[15,195-197]. Notably, each of the 15 ESPs acted as olfactory stimulant in the VNO at concentrations of 10^{-7} M. Some ESPs are released into the tears or expressed in the submaxillary glands producing saliva, but ESPs have not yet been found to occur in urine[143]. Therefore, a potential role of ESPs for signalling of genetic identity seems to be limited to behaviours involving direct facial contact of mice, which is common for example during mate selection.

The suitability of urinary peptides for signalling of genomic relatedness was addressed in the present thesis, and I will discuss this in section "4.3. Genotype specific peptides offer new perspectives for olfactory research".

1.7. Objectives

Since its first proposal in 1975 [147], the hypothesis that the MHC might influence social behaviours of jawed vertebrates has stimulated extensive experimental and theoretical work (reviewed for example by Kwak *et al.*[174] and Ruff *et al.*[144]). The idea has also gained access to the common public press, and there are commercial offers[198] for MHC typing of humans promising to help in finding the perfect sexual partner based on matching of MHC types between individuals. Obviously, there is a general public interest in this hypothesis, in addition to the fascinating scientific challenge that it provides.

The best studied species with regard to MHC related behaviours is *Mus musculus*. The putative MHC related signal inducing the pregnancy block effect is present in urine[149], but its molecular identity has remained unclear. Synthetic MHC class I peptide ligands are recognised with extreme (up to 10^{-14} M) sensitivity by mouse chemosensory neurons[96,97], and they have been reported to induce the pregnancy block effect when applied at high (about 10^{-4} M) concentrations[96,99]. Therefore, the idea that MHC-dependent peptides corresponding to MHC ligands are present in urine and mediate the putative MHC related behaviours has become the most popular molecular model for the hypothesis of MHC mediated social interactions. Milinski *et al.* even extended this model to sticklebacks (*Gasterosteus aculeatus*)[100] and humans[199] with water and armpit excretions, respectively, being the proposed fluids to contain MHC-dependent peptides. However, the lack of knowledge about such peptides in any natural source accessible for nasal recognition has been a major barrier for this hypothesis. Indeed, there is disagreement in the literature about the presence of MHC-dependent peptides in mouse urine[6,96,139,144]. The primary aim of the present study is to clarify whether mouse urine contains MHC-dependent peptides and, in case of a positive finding, at which concentrations these peptides occur (*Nat Comm*, joint).

Our finding that MHC-independent MHC motif peptides dominate over MHC-dependent ones in urine questions the original logic of the MHC peptide ligand model. In the light of the fact that many social behaviours *a priori* just require information about genomic identity and not necessarily about the MHC, and in view of behavioural evidence of MHC-independent but genotype-dependent social interactions in mice[16,200], I broadened the original hypothesis. I speculated that most if not all urinary peptides of a suitable length (starting from 3 amino acids[98] and possibly involving peptides longer than MHC class I peptides) might act as olfactory cues in *Mus musculus* and that these peptides could transmit information about genetic identity. Hence, the second major objective of my PhD thesis is the demonstration of MHC-independent urinary peptides that vary within *Mus musculus* according to individual genetic differences (*Nat Comm*, joint).

Self-evident, the mere demonstration of urinary peptides that contain information about genomic individuality is not sufficient to establish their relevance in social behaviours. A prerequisite for a potential role in olfaction is the presence of chemosensory neurons which are able to recognise the

peptides in a sequence specific manner, i.e. in a way that enables discrimination between the differences originating from genomic diversity. To identify such neurons, we established a cooperation with the neurobiologists Trese Leinders-Zufall and Frank Zufall from the University of Saarland.

2. Materials and Methods

2.1. Mice

All mice were males aged 3-25 months (for details see Table 2-1 and Figure 3-5). By using only male mice, we excluded variations originating from the sex or the menstruation cycle in our experimental analyses. Mice were kept singly ruling out variations depending on social status. The animals obtained food and water *ad libitum*. C57BL/6J mice (abbreviated B6, stock #000664) and mice with a targeted mutation in the β_2-microglobulin gene (B6.129P2-*B2m*^{tm1Unc}J, abbreviated B6/β_2m$^{-/-}$, stock #002087) were obtained from the Jackson Laboratories, Bar Harbor, Maine. Mice homozygous for the *B2m*^{tm1Unc} mutation have little if any MHC class I protein expression on the cell surface and there are few CD8$^+$ cytotoxic T-cells. The mutation was originally produced by Koller et al.[201] in strain 129. Backcrossing for 11 generations to C57BL/6J mice at the Jackson Laboratories ensures that all loci are of C57BL/6J origin except those surrounding the β_2m gene, i.e. B6 and B6/β_2m$^{-/-}$ mice are congenic (*Nat Comm*, joint), (Dipl).

Mice for mass spectrometric analyses of urinary peptides

For MS experiments, mice of strains C57BL/10SnJ (abbreviated B10, stock #000666), B10.D2-*Hc*^{1} *H2*^{d} *H2-T18*^{c}/nSnJ (abbreviated B10.D2, stock #000463) and BALB/cJ (stock #000651) were used in addition to B6 and B6/β_2m$^{-/-}$ animals. All B6-derived mice as well as B10 mice carry the *H2*^{b} haplotype. BALB/c mice express the H2^d haplotype; the congenic line B10.D2 carries the *H2*^{d} complex on the background of B10 mice. B10, B6 and B6/β_2m$^{-/-}$ mice lack H2-L molecules, whereas B10.D2 and BALB/c mice express H2-L^d. For each of the five strains used in MS analyses, we pooled urine from 4 to 5 individuals in order to diminish influences of intra-strain variability. Mice for MS analyses were fed with a defined composition diet (C1000 Control Diet Rats/Mice, Altromin, Lage, Germany) containing bovine milk protein as only protein source to control for urinary peptide contaminants from food (*Nat Comm*, joint), (Dipl).

Mice for the examination of the MHC-dependence of SIINFEKL in urine

For the search for urinary SIINFEKL, we applied the strain C57BL/6-Tg(ACTB-OVA)916Jen/J (abbreviated B6/OVA$^+$, stock #005145). B6/OVA$^+$ is a transgenic line produced in B6 mice, which expresses chicken ovalbumin as a transmembrane protein on cell surfaces of all organs[202]. Beate Pömmerl expanded this strain by crossing breeders obtained from the Jackson Laboratories at the end of 2005. Subsequently, she produced the strain B6/OVA$^+$/β_2m$^{-/-}$ by mating B6/β_2m$^{-/-}$ and B6/OVA$^+$ mice; all F1 animals expressed ovalbumin. Beate Pömmerl typed all derived F2 animals for ovalbumin and the presence of CD8$^+$ T lymphocytes. OVA$^+$ mice lacking CD8$^+$ T-cells were designated

B6/OVA$^+$/β_2m$^{-/-}$ (Figure 2-1; experimental procedure see below). B6/OVA$^+$ mice are congenic in relation to both B6 and B6/OVA$^+$/β_2m$^{-/-}$ mice (*Nat Comm*, joint).

Typing of mice for expression of ovalbumin

B6/OVA$^+$ and B6/OVA$^+$/β_2m$^{-/-}$ mice were bred in our institute by Beate Pömmerl. All mice originating from this in-house breeding were typed for their expression of ovalbumin by Beate Pömmerl using flow cytometry (for examples see Supplementary Table 9-1 and Supplementary Table 9-2). Additional analyses of ovalbumin expression were jointly performed by Beate Pömmerl and me (e.g. Table 3-12). The polyclonal anti-ovalbumin antibody was applied at a concentration of 10 µg/ml corresponding to an estimated $4 \cdot 10^6$ antibody molecules per leucocyte. However, this concentration was not saturating as I determined when titrating the antibody in an additional experiment (also assisted by Beate Pömmerl; data not shown). This might be due to pronounced contributions of cross-reactivity at high concentrations, i.e. saturating concentrations are not reached because more and more non-ovalbumin molecules are stained or point to the presence of more than $4 \cdot 10^6$ ovalbumin molecules per leukocyte. Therefore, accurate relative comparisons of the amounts of ovalbumin are not possible even within the same experiment.

The experimental procedure for standard flow cytometric phenotyping of mice

All centrifugations were performed at 535 g and 4°C for 3 min. As soon as samples had been brought in contact with the fluorescent dyes (phycoerythrin, fluorescein isothiocyanate (FITC) or peridinin chlorophyll (PerCP)), they were kept in the dark whenever they were not handled for pipetting and washings.

Extraction and preparation of leucocytes

For each staining, Beate Pömmerl collected 75 µl of blood from retrobulbar blood vessels using heparin-coated collection tubes on ice. We (see the individual tables and figures for assignment of the respective experimenter) processed blood immediately after collection. The blood sample was mixed with 5 ml of cold (4°C) ACK buffer (150 mM NH_4Cl, 10 mM $KHCO_3$, 0.1 mM EDTA (ethylenediaminetetraacetic acid), adjusted to pH = 7.3 applying NaOH). We then incubated the sample at room temperature until its turbidity had changed to a clear solution (no longer than 7 min), which is indicative of complete bursting of erythrocytes, and centrifuged it. All following steps until flow cytometric measurement were performed on ice. The supernatant was decanted and discarded, and the cell pellet was resuspended by forcefully pulling the sample tube over a coarse grid. Cells were washed first with 5 ml of phosphate buffered saline (PBS) and then with 5 ml of FACS-buffer consisting of PBS with 2% fetal bovine serum, 0.02% NaN_3 and 2 mM EDTA. The resuspended cell pellet was transferred into the well of a round-bottom 96-well plate. The plate was centrifuged and flicked to remove the supernatant, and the pellet was resuspended with a multichannel pipette applying 200 µl FACS buffer − this is referred to as “washing” in the following. Afterwards, the plate was centrifuged and flicked again.

Immunochemical staining for ovalbumin

The pellet was resuspended in 50 µl of polyclonal rabbit anti-ovalbumin antibody (10 µg/ml in FACS buffer; biotin conjugated IgG, Rockland, Gilbertsville, PA, USA) and incubated for 20 − 30 min on ice. After two washings, the sample was resuspended and incubated with 50 µl of R-phycoerythrin conjugated streptavidin (Invitrogen, Carlsbad, CA, USA) at 2.5 µg/ml in FACS buffer for 20 − 30 min.

Immunochemical staining for CD8

For standard phenotyping of mice, Beate Pömmerl stained each sample both for ovalbumin and CD8. She applied the PerCP-conjugated anti-CD8 monoclonal antibody (PerCP Rat anti-Mouse CD8a, clone 53-6.7, BD Biosciences Pharmingen) at a concentration of 1 µg/ml using 50 µl/sample. To control for the specific depletion of $CD8^+$ T cells in $\beta_2m^{-/-}$-deficient mice, she also tested the presence of $CD4^+$ T cells applying a FITC-conjugated anti-CD4 monoclonal antibody (FITC Rat anti-Mouse CD4, clone GK1.5, BD Biosciences Pharmingen) at 2.5 µg/ml and 50 µl/sample.

Immunochemical staining for MHC molecules

MHC molecules on the surface of T cells were investigated in separate experiments unrelated to the ovalbumin and CD8 stainings. MHC expression analyses were performed by Beate Pömmerl and me during my diploma studies and are detailed in my diploma thesis[203]. In short, we applied the monoclonal IgG2b antibody Y-3[204,205] directed against H2-K^b and the monoclonal IgG2a antibody B22-249[206] targeting H2-D^b. IgG2a or IgG2b monoclonal antibodies specific for NP were used as isotype control antibodies. Y-3, B22-249 and the isotype control antibodies were applied at concentrations of 10 µg/ml, and they had been produced in-house by Claudia Falkenburger or Wiebke Ruschmeier using protein A affinity purification of the filtered supernatants of hybridoma cultures. MHC expression was measured only on T cells and not on all lymphocytes to avoid binding of the monoclonal antibodies to the Fc receptors present on NK cells. T cells were selected in the flow cytometric data analyses based on T cell receptor staining obtained by applying a FITC-conjugated anti-mouse CD3 antibody (BD Biosciences Pharmingen, diluted 1/100).

Flow cytometric measurement

Three washings followed before the cells were transferred into small vials which are compatible with the FACSCalibur flow cytometer equipped with CellQuestPro software (Becton Dickinson). Lymphocytes were selected for data analysis by their characteristic forward and side scatter signals (*Nat Comm*, Theo).

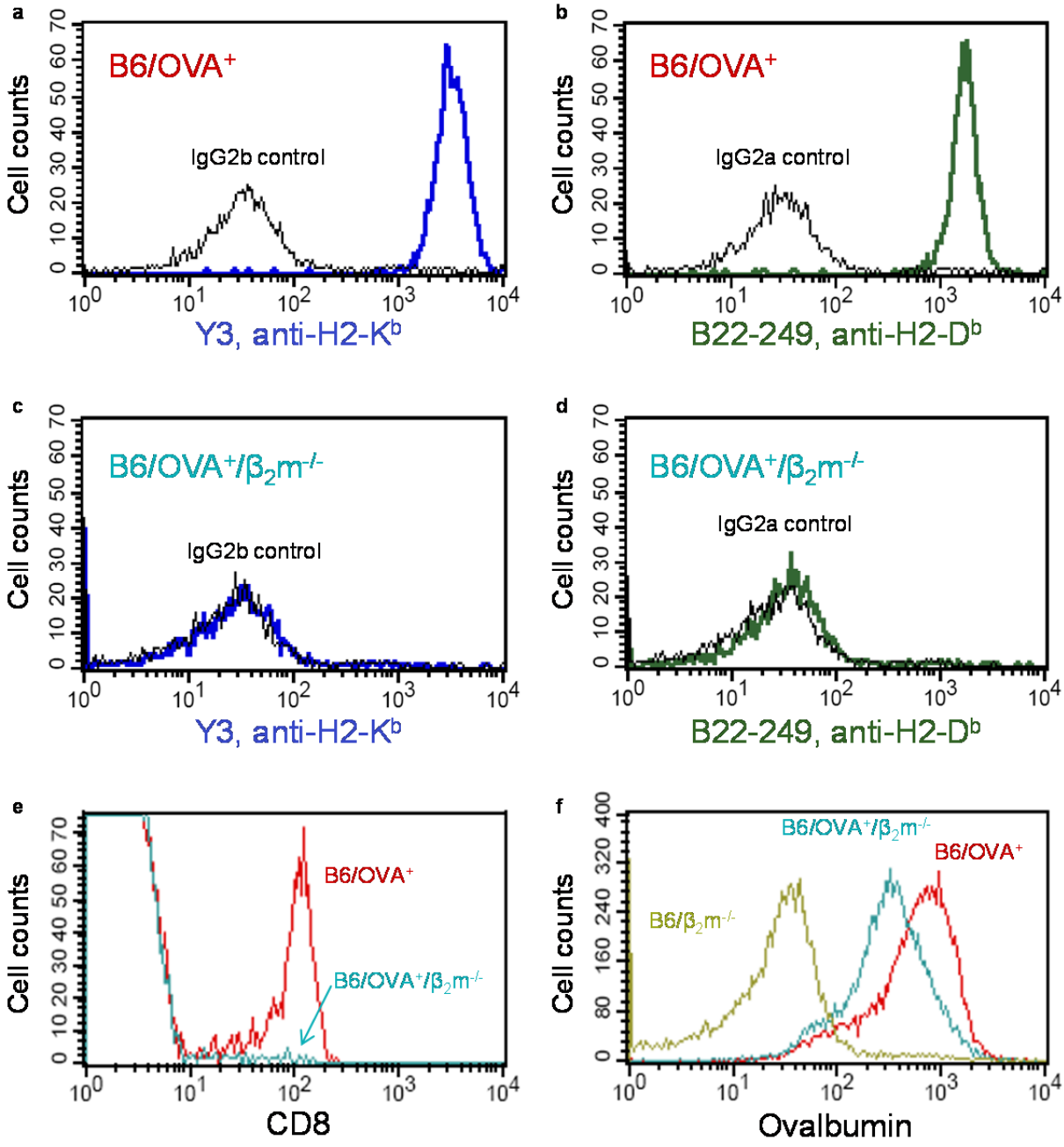


Figure 2-1 | Phenotypic characterisation of B6/OVA$^+$ and B6/OVA$^+$/β_2m$^{-/-}$ mice. Stained lymphocytes from peripheral blood were analysed by flow cytometry. Fluorescence intensities are depicted on abscissae. (**a-d**) Expression of MHC molecules on T cells of single mice. All analyses were repeated with a second mouse of each strain leading to comparable results. The blue graphs indicate staining with the monoclonal Y-3 IgG2b antibody[204,205] which binds to H2-K^b. Results obtained with the monoclonal B22-249 IgG2a antibody[206] specific for H2-D^b are depicted in dark green. Black graphs represent isotype-control antibodies directed against an epitope that is not present on the cells investigated. (**e**) Staining of lymphocytes with a monoclonal anti-CD8 antibody. The CD8$^+$ cell population is vastly diminished in B6/OVA$^+$/β_2m$^{-/-}$ mice as compared to B6/OVA$^+$ mice indicating the lack of CD8$^+$ T cells in the former. (**f**) Lymphocytes stained with polyclonal anti-ovalbumin antibodies. The experiment was done with 13 B6/OVA$^+$ and 10 B6/OVA$^+$/β_2m$^{-/-}$ mice. For both mouse strains a representative result close to the mean fluorescence intensity (MFI) of that strain is shown. B6/OVA$^+$ mice: MFI = 613, s.d. = 74; B6/OVA$^+$/β_2m$^{-/-}$ mice: MFI = 402, s.d. = 131. As an ovalbumin-negative control a B6/β_2m$^{-/-}$ mouse was included (MFI = 47). Data in **a-d** were jointly generated by Beate Pömmerl and me during my diploma thesis. Raw data in **e** and **f** were kindly provided by Beate Pömmerl (*Nat Comm*, Theo).

2.2. Isolation of urinary peptides

Collection of urine

Urine was collected individually by either holding the animals over small plastic petri dishes and freezing urine immediately in dry ice (for data in Figures 3-4 and 3-5 as well as Table 3-3 of this thesis and Figure 2 of Sturm *et al.*[207]) or with the help of metabolic cages (for data in Figures 2-2 and 3-2, Tables 3-1, 3-2, 3-4, 3-5, 3-6, 3-7 and 3-8, Supplementary Figures 9-1 to 9-40 and Supplementary Tables 9-3 to 9-7). Metabolic cages were from Techniplast (Buguggiate, VA, Italy) and allowed the separation of urine from faeces. Urine was stored at -80 °C (*Nat Comm*, joint).

Fractionation of urine and urinary peptides

After thawing, urine was supplied with protease inhibitor solution (final 1x, from Complete Protease Inhibitor Cocktail Tablets, Roche Diagnostics) and centrifuged at 3345 g and 4 °C through a 3 kDa cut off filter (Amicon Ultra Centrifugal Filters, Regenerated Cellulose, Millipore; in MS experiments no. 1 and 2, a 10 kDa cut off was used instead, see Table 2-1). In order to improve recovery of free peptides, the material retained by the filter was suspended in water containing protease inhibitors and centrifuged again. The combined filtrates were acidified with trifluoroacetic acid (TFA) to pH 2. A crude peptide fraction was isolated as follows[208] (setting for 5 ml urine per sample): A 100 mg Strata^{TM}-X Polymeric Reversed Phase column (Phenomenex) was conditioned using 5 ml methanol and equilibrated with 5 ml 0.1% TFA in water. Then the ultrafiltrate was applied. After washing the column with 5 ml 10% acetonitrile/0.1% TFA, peptides were eluted using 5 ml 80% acetonitrile and lyophilised. The samples were dissolved in 1 ml 0.1% TFA (solution A) by vortexing and ultrasonication, briefly centrifuged, and the clear supernatant was applied to a C_{18}-phase 250 x 8 mm HPLC column (Multospher 129 RP 18-5 µ, Ziemer Chromatographie, Langerwehe, Germany) previously washed with acetonitrile/0,08% TFA (solution B) and solution A. A blank run was performed before every sample application. Peptides were eluted with a 1 h gradient rising linearly from 10 to 60% solution B in 40 min and fractions were collected manually (*Nat Comm*, joint).

We processed urine of B6 mice supplemented with synthetic SIINFEKL in parallel with urine from B6, B6/OVA$^+$ and B6/OVA$^+$/β_2m$^{-/-}$ mice. Thereby we determined the mean yield (± s.d.) for the extraction of free SIINFEKL from urine to be 35% (± 19%, n = 5) using the SIINFEKL-H2-K^b-ELISA. This value takes into consideration, that approximately 8 to 13% (n = 2) of synthetic SIINFEKL did bind to carrier molecules (> 3 kDa) when added to urine before ultrafiltration (*Nat Comm*, Theo).

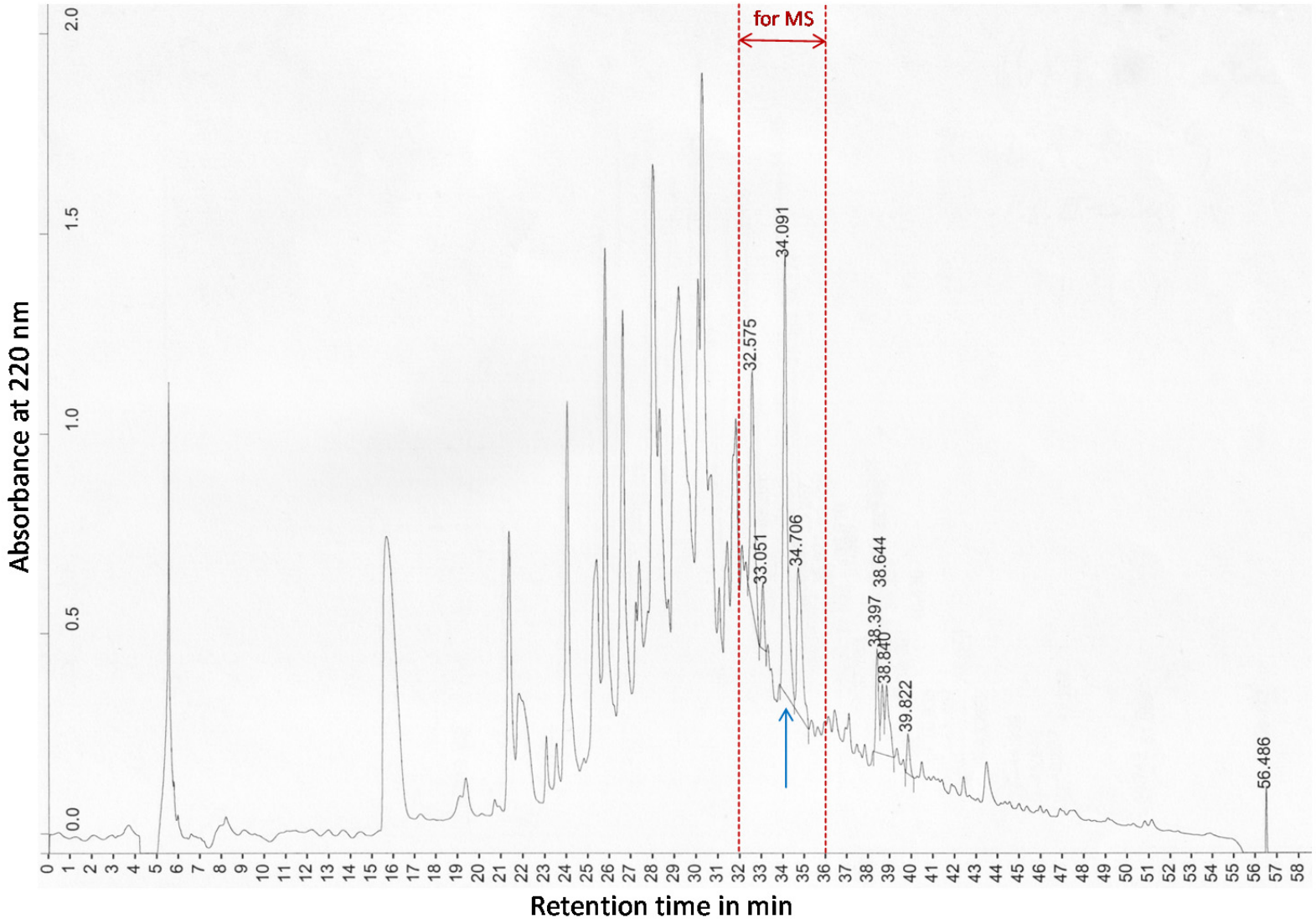


Figure 2-2 | HPLC elution diagram of the low molecular weight urine fraction obtained by 10 kDa ultrafiltration and solid phase extraction of peptides using Strata™-X Polymeric Reversed Phase columns. The sample applied to HPLC was derived from 5 ml of B6 mouse urine (exp. no. 2, cf. Table 2-1). Note that this sample had a brown colour and still contained many substances unrelated to peptides. Some of these are expected contribute to absorbance at 220 nm, in addition to the peptides. Therefore, it is not possible to estimate which proportion of all urinary peptides was subjected to MS. The fraction analysed by MS in experiments no. 1, 2 and 4 (cf. Table 2-1) is indicated by a red double arrow. The blue arrow indicates the elution time of synthetic SIINFEKL and SYFPEITHI in urine as determined in separate experiments by MS at a Q-TOF Ultima™ mass spectrometer (Micromass, Manchester, England) and / or by ELISA (Figure 3-4) (*Nat Comm*, Theo).

2.3. Mass spectrometric measurements

Today, mass spectrometry (MS) is the most widely used method for sequence analysis of peptides[209]. Within the last two decades, major improvements in sample fractionation, MS instrumentation and bioinformatic data processing have enabled the parallel analysis of hundreds to thousands of individual peptides in just one experiment with unprecedented velocity[210,211] and confidence of identification[212,213]. Further key advantages of liquid chromatography (LC)-MS over traditional Edman degradation[214,215] are its higher sensitivity and its ability to sequence peptides which are present in complex mixtures, i.e. individual peptides do not need to be applied in a pure form to LC-MS[216]. In contrast to Edman degradation, antibody-based methods can outperform LC-MS analyses in terms of sensitivity and dynamic range (ruggedness towards the presence of abundant other peptides; see section "3.4. Detection of an MHC-dependent peptide in mouse urine"). However, there are only very few antibodies recognising a particular peptide and their specificity for peptide sequence and length is usually not absolute (e.g. [217,218]). Moreover, antibody based assays are biased towards preselected peptides of interest not providing the objective discovery spirit immanent in most LC-MS experimental setups. Hence, LC-MS was the method of choice to start with urinary peptide analyses in the present study.

For MS experiments and for each mouse strain, we used mixtures of urine from 4 to 5 individuals in order to diminish influences of intra-strain variability. In experiment no. 4 (cf. Table 2-1), which was used for absolute mass spectrometric quantifications, synthetic peptides containing $^{13}C_5{}^{15}N_1$-valine (L-Valine-OH-$^{13}C_5{}^{15}N_1$, FMOC, 98% ^{13}C, 98% ^{15}N, CortecNet, Voisins-Le-Bretonneux, France) were spiked into manually collected urine directly after thawing and pooling. Our MS experimental setup was optimised for the detection of peptides having the approximate length of MHC class I peptide ligands. Very long and very short peptides were probably lost due to (i) the molecular weight cut off in urine ultrafiltration, (ii) the restrictions of the analysed 4 min HPLC fraction, (iii) the m/z range and precursor ion charge allowed for MS^2 spectra acquisition (cf. Table 2-1) as well as (iv) the applied target-decoy database search strategy in bioinformatic processing of our MS data, which is not suitable for the detection of peptides shorter than about 7 amino acids[132]. Although we detected only peptides of a length between 6 and 32 amino acids, the presence of shorter and longer peptides in mouse urine is likely (*Nat Comm*, Theo).

The measurements for MS experiments (exp.) no. 1 − 3 (see Table 2-1) were executed during my diploma thesis. However, they are included in this PhD thesis because they are the basis for both MS exp. no. 4 and the extensive data analyses regarding MHC-independent peptides (cf. sections "2.4.2. Prediction of SAV peptides" and "2.4.3. Reprocessing of MS data for SAV peptide identification") that I performed during my PhD thesis. The nano-HPLC and LTQ-Orbitrap XL systems were operated by Stephan Jung, Boris Maček, Johannes Madlung or Daniel Kowalewski using settings consulted with me.

Table 2-1 | Mass spectrometric experiments: setup and basic parameters. (*Nat Comm*, Theo)

Experiment no.	1	2	3	4
Used mouse strains The age of the mice [in months] at the time of urine collection is indicated with grey font in brackets	B6 (6) B6/$\beta_2m^{-/-}$ (6) BALB/c (6)	B6 (8) B6/$\beta_2m^{-/-}$ (8) BALB/c (8) B10 (8) B10.D2 (8)	B6 (16 – 18, obtained from Charles River) B6/OVA$^+$ (7 – 16)	B6 (6) B6/$\beta_2m^{-/-}$ (6) BALB/c (20) B10 (20) B10.D2 (20)
Collection of urine	metabolic cages	metabolic cages	metabolic cages	manually
Protease inhibitors	not used	present in collection tubes of metabolic cages	present in collection tubes of metabolic cages	added directly after thawing of urine
Acidification of urine before or after initial ultrafiltration	before	after	after	after
Molecular weigth cut off for urine ultrafiltration	10 kDa	10 kDa	3 kDa	3 kDa
Retention time (in minutes^{seconds}) of eluate from 250 x 8 mm C_{18} HPLC used for LC-MS	$32^{00} – 36^{00}$	$32^{00} – 36^{00}$	$34^{20} – 34^{30}$	$32^{00} – 36^{00}$
Amount of urine processed per mouse strain	10 ml	5 ml	15 ml	0.5 ml
Corresponding amount of urine used per LC-MS measurement (yield losses neglected)	250 µl	100 µl	2250 µl	100 µl
Number of LC-MS measurements per mouse strain	3 (B6: 2)	3	1	1
m/z range (in Th) measured in MS scans and allowed for fragmentation	300 – 2000	300 – 870	300 – 1000	300 – 2000
Charge states allowed for fragmentation	all	only 2^+ and 3^+	only 2^+ and 3^+	only 2^+
Fragmentation schedule	1st and 2nd run: LTQ-MS^2 of top 10 MS^1 signals; 3rd run: First LTQ-MS^2, then FT-MS^2 for each of top 3 MS^1 signals	1st and 2nd run: LTQ-MS^2 of top 10 MS^1 signals; 3rd run: First LTQ-MS^2, then FT-MS^2 for each of top 3 MS^1 signals	First LTQ-MS^2, then FT-MS^2 for each of top 3 MS^1 signals	LTQ-MS^2 of top 5 MS^1 signals
Resolution in FT-MS^2 scans	15,000	15,000	7,500	-
Number of identified peptides at FDR ≤ 1%*	469	267	75	inclusion list applied
Number of identified 8-, 9- and 10-mer peptides at FDR ≤ 1%*	102	88	17	inclusion list applied

Asterisks indicate that, in addition to the FDR ≤ 1%, peptides are only included if they were identified with Mascot score ≥ 25 in at least one MS² scan of the respective experiment. Contaminating peptides were mainly derived from the food of the mice (*Bos taurus* caseins) and are not contained in the numbers given.

2.3.1. Nano-HPLC for LC-MS coupling

In order to obtain high quality MS^2 spectra with good signal-to-noise ratios, it is essential that the respective peptides are present at high concentrations during the time of fragmentation. Furthermore, it is advantageous if the individual peptides reach the mass spectrometer in a sequential fashion to avoid cofragmentation of peptides with similar mass-to-charge ratios. Both goals can be promoted by the coupling of MS to high performance liquid chromatography (HPLC), which selectively enriches different peptides at sequential time points of the applied elution gradient. Reversed phase chromatography is especially suitable for such an LC-MS coupling because (e.g. in contrast to ion exchange chromatography) it does not rely on buffers containing salts. Salts do suppress electrospray ionisation (ESI) of analytes and often form adducts with peptides[174] thereby splitting the ion population of a given peptide species, decreasing sensitivity and confounding peptide identification. A description of the theoretical principles of LC with a focus on the separation of peptides and proteins can be found in chapter 11 of ref. [219, (Dipl)].

The extraction and coarse fractionation of urinary peptides has been described in section "2.2. Isolation of urinary peptides". For MS experiments, I concentrated the preparative HPLC fraction of interest, comprising 4 min (10 s in the case of MS exp. no. 3, cf. Table 2-1), using lyophilisation and / or vacuum centrifugation. I then purified the concentrate on a C_{18} StageTip[220] to support the electrospray ionisation process (cf. section "2.3.2. Electrospray ionisation") and to protect the nano-HPLC column from overload or clogging. Peptides were eluted with 80% acetonitrile/0.5% acetic acid. I concentrated the eluate from the C_{18} StageTip in a vacuum centrifuge and supplemented it with 1/10 volume 2% acetonitrile/1% TFA. Applying an autosampler, it was then loaded onto a 15 cm x 75 µm nano-HPLC column packed by Johannes Madlung with 3 µm C_{18} beads (Dr. Maisch GmbH). Peptides were eluted at a flow rate of 200 nl / min using a gradient of 16 to 40% acetonitrile in 0.5% acetic acid within 97 or 105 min. We applied a nanoLC-2D (Eksigent) or an Easy-LC (Proxeon Biosystems) nano-HPLC system coupled to an LTQ Orbitrap XL mass spectrometer (Thermo Fisher Scientific) via a nano-electrospray liquid chromatography-MS (LC-MS) interface from Proxeon Biosystems [(*Nat Comm*, Theo)].

Synthetic IVIYHTSAQSIL and FVIYHTSAQSIL (Supplementary Figures 9-30 and 9-32) were measured under slightly different C_{18} chromatographic conditions using an UltiMate 3000 Ultra-HPLC system, an additional 2 cm PepMap100 C_{18} Nano-Trap column, a 50 cm PepMap C_{18} nano-HPLC column (the latter three devices all from Dionex, part of Thermo Fisher Scientific) and 0.05% and 0.04% formic acid, respectively, instead of acetic acid in the LC solutions.

2.3.2. Electrospray ionisation

Electrospray ionisation (ESI)[221] is the most commonly used ionisation method for peptides[222]. Like matrix-assisted laser desorption ionisation (MALDI)[223] it enables the transfer of large, non-volatile biomolecules into the gas phase where the intact, ionised molecules can subsequently be analysed by MS. Due to the non-destroying nature of the ionisation process, ESI and MALDI are termed "soft" ionisation methods[216, (Dipl)].

A major advantage of ESI over MALDI is its suitability for easy on-line coupling of LC to MS (cf. section "2.3.1. Nano-HPLC for LC-MS coupling"). Whereas MALDI usually generates singly charged ions for peptides and proteins smaller than about 50 kDa[219], ions generated by ESI are usually multiply charged[221]. Even for one and the same molecular species, ESI generally results in several ion populations differing from one another by one charge[221]. In the positive ion mode applied for the present study, peptides have a positive net charge due to the acidic LC solutions. Their observed m/z values correspond to $(M + z \cdot H^+) / z$, where z corresponds to the number of protons (and thereby charges) per peptide. The number of charges per peptide raises with increasing peptide length and number of basic amino acid residues contained in the peptide[216]. This information can be integrated into the manual validation of MS^2 spectra (see section "2.4.4. Manual validation of MS2 spectra"). In the ESI-MS setting applied here, the doubly or triply charged ion population is typically the most dominant one for peptides of a length between 7 and 20 amino acids (cf. ref. [216] and Supplementary Tables 9-3 to 9-7; In the Supplementary tables, consider only values from MS exp. no. 1 for this purpose because ions with one, four or more charges were excluded from fragmentation in MS exp. no. 2 − 4).

For on-line coupled nano-LC-ESI-MS, the LC eluate flows into an ESI emitter, which is a very thin capillary tip that narrows at the end down to a diameter of a few micrometers. The nano-HPLC column and the ESI emitter were made out of a single part in the setting for exp. 1 − 4, leading to increased LC peak sharpness by avoiding post-LC volumes and associated diffusion. In the positive ion mode applied, the ESI emitter had a positive voltage of 1.5 − 2.0 kV relative to the negatively charged entrance of the mass spectrometer. This results in the electrochemical oxidation of negative ions (the oxidation can also generate new positive ions of special substances) in the emitter and electric attraction of positive ions towards the entrance orifice[224,225]. The interaction of this attraction with the repulsive Coulomb forces in the positively charged liquid stream and the surface tension of the eluate results in the formation of a Taylor cone[226] at the emitter tip. At the end of the Taylor cone, the eluate forms a linear stream which gets unstable after a short distance resulting in a plume of many small positively charged droplets[224]. On the flight to the mass spectrometer, the droplets in the spray continuously lose some of their solvent[227] removing thermical energy from the peptide ions and thereby suppressing peptide fragmentation in the ion source. The evaporation of the solvent reduces the diameter of the droplets leading to an increase of charge density and strong repulsive Coulomb

forces. According to the charged-residue model, the droplets explode when the latter forces exceed the surface tension (Rayleigh limit[228]) resulting in smaller droplets[224]. The "single ion in droplet theory" supposes that the processes of evaporation and explosion repeat until a single desolvated peptide ion remains[219,229,230]. In contrast, the "ion evaporation model" assumes that analyte ions are directly emitted from highly charged droplets that still contain many analyte ions[231,232]. The two theories are not mutually exclusive, and Coulomb fissions of droplets and ion emission might occur in parallel[216].

2.3.3. LTQ Orbitrap XL-MS

All MS experiments of this work (except tracing of synthetic SIINFEKL and SYFPEITHI in preparative HPLC fractions, see Figure 2-2) were performed at an LTQ-Orbitrap XL™ hybrid mass spectrometer (Thermo Electron, Bremen). This instrument consists of two separate mass analysers, a linear trap quadrupole (LTQ) and an orbitrap, both of which can operate in parallel. The orbitrap was commercially introduced in 2005, and the applied LTQ-Orbitrap XL mass spectrometer has been followed by more sensitive orbitrap instruments in the meantime, e.g. the Orbitrap Velos Pro[233,234], the Orbitrap Elite[235], the Q Exactive[210] and the Orbitrap Fusion[234]. Although we could have identified more peptides by applying one of the newer machines, the biological message of the work would probably not have changed. Except for the MHC-dependent peptides, we could validate every peptide group by several representatives (Figure 3-2, Supplementary Tables 9-3 to 9-7). If we had wanted to identify more peptides, we could just have analysed additional urine fractions from the preparative HPLC (cf. Figure 2-2).

Applying our LTQ Orbitrap XL, we were not able to reproducibly identify synthetic SIINFEKL spiked into urine at concentrations of 10^{-11} M even when applying inclusion lists or selected ion monitoring (SIM) scans. Besides the sensitivity, the major problem appeared to be the dynamic range, i.e. the presence of other, much more abundant peptides coeluting with SIINFEKL during nano-HPLC. It is conceivable that more advanced mass spectrometers in combination with targeted MS strategies like selected reaction monitoring (SRM[236]) would have been able to identify SIINFEKL in the urine of B6/OVA$^+$ mice, thereby replacing the need for the SIINFEKL-H2-K^b-ELISA. However, today`s most powerful mass spectrometers were not available at the time when I performed the SIINFEKL measurements, at least not in our institute.

The ion`s path to the orbitrap for acquisition of MS^1 spectra

The ions generated during ESI (cf. section "2.3.2. Electrospray ionisation") are transported via a heated ion transfer tube, a skimmer, two square quadrupoles and an radiofrequency (RF)-only octapole into the LTQ. Vacuum pumps connected to these elements gradually decrease the pressure from atmospheric levels at the entrance of the ion transfer tube to about $1.33 \cdot 10^{-3}$ Pa (10^{-5} Torr) in

the octapole. Thereby, peptide ions as well as charged and neutral molecules from the ambient air are drawn into the mass spectrometer. The orifice of the skimmer is offset from the line of sight defined by the ion transfer tube and efficient passage of all the elements requires repeated focusing of the ion beam by RF electromagnetic fields applied to the multipoles and several electromagnetic lens situated between them. Neutral molecules are not affected by the RF fields and are therefore prone to both collisions with the solid components and loss to the vacuum pumps. For an increased transmission of positive (peptide) ions, the pressure gradient is accompanied by a gradient of electric potentials generated by the multipoles and the electromagnetic lens, with increasing negative charges towards the octapole. The transport of anionic contaminants is actively discouraged by this electric gradient, and negative ions are also likely to collide with the positively charged walls of the ion transfer tube[174].

If the ions arriving in the LTQ linear ion trap[237] are destined for the measurement of an MS^1 spectrum in the orbitrap, the main purpose of the LTQ is to accumulate a defined number of ions via the Automatic Gain Control (AGC) mechanism (10^6 ions in MS exp. no 1 − 4, 500.000 ions for measurements of synthetic (I/F)VIYHTSAQSIL). This is achieved by collecting the incoming ion beam for a variable, scan-specific injection time (IT) calculated by the AGC system based on a very short prescan of 1 ms. AGC is of crucial importance with respect to extreme incoming ion amounts. If there are too little ions, the MS spectra will be characterised by low signal to noise ratios and bad quantitative reproducibility. This can be counteracted by longer ITs. On the other hand, if there are too many ions, mass resolution and mass accuracy will suffer due to space charge effects. This is avoided by shortening of ITs[174].

The LTQ consists of a quadrupole split into three parts: a front section, a centre section and a back section. To trap ions, an RF voltage applied to the quadrupole rods focuses ions in the radial direction, whereas additional DC voltages applied to the front and back sections of the quadrupole focus ions in the axial direction. Due to the injection process, ions have an initial momentum independent of the LTQ voltages. Even ions of the same m/z value differ largely in their kinetic energies hampering precise control of ions and lowering resolution of consecutive MS scans. To mitigate this problem, the LTQ is filled with Helium. Ions collide with Helium which causes them to lose kinetic energy ("collisional cooling") enabling better focusing in the centre of the trap. Note that ions do not fragment at this stage because their velocity is too low due to the lack of an additional collisional excitation voltage[174,237].

After accumulation in the LTQ, ions are axially ejected and transported via an RF-only octapole into the C-trap, a curved quadrupole mimicking the shape of the letter "C". At the distal end of the C-trap, ions are reflected by a trap electrode and consequently shuttle multiple times between the C-trap and the transfer octapole. To support the generation of a homogenous, focused ion population, the C-trap is filled with N_2. During the shuttling, ions loose energy due to collisions with the N_2 dampening gas ("collisional cooling"). By applying an electric potential to both ends of the C-trap, ions get axially

confined and form an arc paralleling the C-shape of the trap. Using pulses of direct current, the ions are then radially ejected via several electromagnetic lenses towards the centre of the circle adumbrated by the curvature of the trap. This centre point outside the C-trap forms the entrance of the orbitrap[174,238]. Using the curvature-based mode, ions can be injected into the orbitrap much more uniformly than it would be possible using axial ejection from a storage trap, especially for large number of ions. Thereby, mass accuracy and dynamic range of the m/z measurements in the orbitrap are crucially improved[239, (Dipl)].

Measuring m/z values in the orbitrap

The orbitrap consists of a spindle-shaped central electrode which is surrounded by two bell-shaped outer electrodes. The outer electrodes are symmetrically separated by a ceramic ring positioned around the middle of the central electrode. When ions are injected into the orbitrap, they start rotating around the central electrode. There is a balance between the centripetal forces, provided by an electric potential, and the centrifugal forces, originating from the rotational velocity. Despite the superior focusing capabilities of the C-trap, ions injected into the orbitrap have slightly different initial positions and kinetic energies even if they have the same m/z value. Therefore, such ions move with different radial (concerning the distance to the central electrode) and angular frequencies leading to ion clouds that form a thin ring around the central electrode. The ring oscillates in axial direction, and the corresponding frequency ω is independent from the variable injection parameters, i.e. initial energy, angles and positions[238,240]. Therefore, ω can be used for calculating the m/z value of the ion population according to the following formula[238].

$$\omega = \sqrt{\frac{z \cdot k}{m}}$$

ω	frequency of oscillations in axial direction
z	number of charges of the ion
m	mass of the ion
k	device-specific constant

Consequently, ions of different m/z values have different axial frequencies ω resulting in a separate ion ring cloud for every m/z ratio. The combined movement in circles around the central electrode and axial oscillation results in helical movements for every single ion. The moving ions of same and different m/z values cause a complex image current in the outer electrodes. Because the outer electrodes are separated only in axial direction, the non-informative circular component of the movements does not affect the recorded image current. Every ion ring cloud generates a sinusoidal image current informing about its axial movement. Due to the multitude of different ion ring clouds, the resulting complex image current must be deciphered via Fourier transformation (FT) to identify the sinusoidal function and its ω for every single ring. The MS spectrum is finally generated by converting the values for ω into m/z ratios and the amplitudes into intensities[238, (Dipl)].

We recorded all MS^1 spectra in the FT orbitrap mass analyser at a resolution of 60,000 at 400 Th (Thomson). The "lock mass" option[241] was enabled for FT-MS^1 and FT-MS^2 scans using the 445.120025 ion for real time internal calibrations. This, in combination with MaxQuant data processing (see section "2.4.1. Searching for peptides contained in standard protein databases"), resulted in very high mass accuracies of < 10 ppm. For identified MS^1 precursor ions, the average absolute mass error ± s.d. was 0.15 ± 0.23 ppm for exp. no. 1, 0.22 ± 0.33 ppm for exp. no. 2 and 0.34 ± 0.54 ppm for exp. no. 3. In comparison to MS^1 scans, the absolute mass error in orbitrap MS^2 spectra was higher (but in almost all cases < 10 ppm) due to the lower resolution applied for these scans (cf. Table 2-1), the usually smaller number of ions available for m/z measurements and the lack of the possibility to integrate the m/z measurements of multiple scans over the elution time (as is use in MaxQuant for MS^1 m/z values[242]) (*Nat Comm*, Theo).

Acquisition of MS^2 spectra

The acquisition of an MS^2 spectrum from an accumulated MS^1 ion population did comprise three basic steps in our setting. First, the parent ions of interest where isolated prior to fragmentation to ensure that most of the detected fragment (MS^2) ions really descend from the targeted precursor. The second step consists of the excitation of the isolated ion population to induce fragmentation. In the third step, m/z values of the MS^2 fragment ions are measured.

We applied two different methods to generate and measure MS^2 ions. One option was the use of the orbitrap for the determination of MS^2 m/z values. Thereby, high resolution, high mass accuracy MS^2 spectra are obtained (see preceding paragraph). Orbitrap MS^2 measurements were always combined with HCD (higher energy collisional dissociation) fragmentation[243] of the parent ions. For this fragmentation mode, accumulated parent ions are axially ejected from the LTQ and transferred via the RF-only octapole and the C-trap (cf. paragraph "The ion`s path to the orbitrap for acquisition of MS1 spectra" of this section) into a special octapole collision cell. On their way from the C-trap to the collision cell, ions are accelerated by electric fields resulting in their fragmentation when encountering the N_2 molecules in the collision cell. The relevant electric potentials are adapted according to mass and charge of ions as well as to the type of the collision gas, that is a normalised collision energy is applied. The generated fragment ions are transported back to the C-trap (ref. [244] and Supplementary Figure 1 of ref. [245]). Ion injection into the orbitrap and measurement of m/z values therein, follows the principles already described for MS^1 ions (see paragraph "Measuring m/z values in the orbitrap" of this section) (Dipl).

The second option that we applied comprises the measurement of MS^2 fragment ions in the LTQ. Ions radially ejected from the LTQ can be detected by two electron multipliers supported by a conversion dynode each. To direct the ejected ions towards this detection system, an additional resonance ejection alternating current (AC) voltage is applied to two of the four quadrupole rods. The ions will finally leave the trap through slits contained in these two rods[174]. Linear quadrupole ion traps can scan

through the entire m/z range or isolate particular ions of interest using two different principles. They either take advantage of resonant frequencies to selectively eject ions of corresponding m/z values, or they make use of the boundaries of the stability diagram of the ion motion in the quadrupole electric fields[237]. For the isolation of precursor ions, the LTQ Orbitrap XL takes advantage of the resonant frequency principle. The resonance ejection AC voltage is applied at multiple frequencies (spaced by 500 Hz each) corresponding to resonance frequencies of all m/z values except the one of the parent ions of interest. Thereby, all ions except the ones with the m/z value (± 2 Th for our LTQ-MS^2 setup) of the targeted precursor are ejected from the LTQ[174].

For LTQ-MS^2, we applied low energy collision-induced dissociation (CID) fragmentation. To achieve this, the resonance "ejection" AC voltage is now applied at the resonance frequency of the precursor ion of interest, i.e. the small frequency range that was left out during the isolation process. However, this time, the amplitude of the applied AC voltage is smaller. Therefore, the precursor ions cannot take up enough kinetic energy to get ejected from the LTQ, but they just speed up their movement within the trap. As during accumulation of ions (described in section "The ion`s path to the orbitrap for acquisition of MS1 spectra"), the trapped ions collide with the noble gas contained in the LTQ, but now with much higher velocities. Repeated collisions result in a considerable increase of the internal vibrational energy of the peptide ions leading to their fragmentation[174]. Especially peptides rich in serine, threonine, aspartate and glutamate[216,246] show a very dominant loss of H_2O from the parent ions during the MS^2 fragmentation reducing the amount of informative fragment ions in MS^2 spectra. To overcome this limitation, the WideBand Activation™ method was enabled, which is a special application of pseudo-MS^3 (Schroeder *et al.*[247]). The wideband excitation applies resonance frequencies not only for the precursor ions but also for all m/z values down to -20 Th below the isolated parent ions. Thereby, all ions in this m/z range including parent peptide ions with loss of H_2O are fragmented leading to the conversion of the non-informative H_2O-loss parent ions into more informative MS^2 ions[174,248]. As a side effect, signals of ions from this 20 Th excitation range are usually missing or of very low abundance in MS^2 spectra (see Supplementary Figures 9-17, 9-18, 9-25 and 9-26 for some prominent examples).

For measuring m/z values of MS^2 ions, the LTQ takes advantage of the boundaries of the stability diagram of the ion motion in the quadrupole electric fields. Ions of different m/z values have different stability trajectories in a quadrupole depending on the RF and DC offset voltages applied to the quadrupole rods[237]. Because the LTQ Orbitrap XL does not apply a DC offset, the stability of the ion motion can be described by the following simple formula[174].

$$q = \frac{k \cdot V}{m/z}$$

q stability parameter associated with the RF voltage
k device-specific constant, comprising the inner diameter of the quadrupole, the elementary charge and the constant frequency of the RF voltage
V amplitude of the RF voltage
m mass of the ion
z number of charges of the ion

All ions with q values between 0 and 0.9 are stable in the LTQ. Ions are ejected from the LTQ by increasing their q value above 0.9 which is achieved by ramping V from low to high values. As can be seen from the above formula, ions with smaller m/z values have higher q values than ions with larger m/z values under otherwise identical conditions. Therefore, ions are ejected from the LTQ from low to high m/z values enabling the recording of an MS^2 spectrum. However, in contrast to the HCD-orbitrap-FT-MS^2 measurements, very small fragment ions cannot be detected in the LTQ-MS^2 mode. The reason is outlined in the following. If fragmentation was performed at precursor ion q values close to zero, even extremely small fragment ions could be recorded because under the applied conditions their q value would still be lower than 0.9. However, if precursor ion q values are too low (e.g. 0.1), the range of q values that the different MS^1 peptide ions occupy decreases (all MS^1 ions with larger m/z values than the targeted precursor will have q values only between 0 and 0.1 in our theoretical example). Therefore, MS^1 precursor ions would crowd together during their movements in the LTQ leading to suboptimal fragmentation[174]. To balance the drawbacks of loss of small MS^2 ions and suboptimal fragmentation, the default value for fragmentation is set to $q = 0.25$. We always used this activation q. This implies that all MS^2 ions with m/z values smaller than about 28% of that of the precursor were lost, which is clearly visible in all of our LTQ-MS^2 spectra (see Supplementary Figures).

$$(m/z)_{\text{observable in MS}^2} > (m/z)_{\text{precursor}} \cdot q_{\text{activation}} / 0.9$$

$$(m/z)_{\text{observable in MS}^2} > (m/z)_{\text{precursor}} \cdot 0.28$$

Besides the low mass cut-off, major disadvantages of the LTQ are its limited resolution of about 1000 (see ref. [174]) and its mass accuracy of only about 0.5 Th. This reduces the confidence of annotations of single fragment ion peaks as compared to orbitrap FT-MS^2 measurements. However, the major advantage of the LTQ in comparison to the applied orbitrap is its superior sensitivity and speed[233,240]. Accordingly, most peptides of the present study could only be identified by LTQ-MS^2 but not by FT-MS^2 (cf. Supplementary Tables 9-3 to 9-7). However, an accurate comparison of the performance of LTQ-MS^2 versus FT-MS^2 would require the same number of runs and a separate FDR statistic for each method as well as the deactivation of the MaxQuant option “keep low scoring versions of identified peptides” during data processing (see section “2.4.1. Searching for peptides contained in standard protein databases”).

We excluded MS^1 ions of unidentified charge state from fragmentation. A normalised collision energy of 35% (100% correspond to 5 V[93,174]) was applied for collision-induced dissociation (CID) whereas 45% were used for higher-energy collisional dissociation (HCD). For MS exp. no. 1 − 4, the LTQ was filled with Helium, whereas Argon was applied instead for measuring synthetic (I/F)VIYHTSAQSIL. Activation time for all fragmentations was 30 ms. In the MS experiments for peptide discovery (no. 1, 2 and 3, cf. Table 2-1), the m/z values of fragmented parent ions were excluded from further sequencing for the following 90 s. During experiment no. 4, only predefined parent ions of an inclusion list containing the m/z values of the $^{13}C_5{}^{15}N_1$-valine synthetic peptides and their corresponding natural forms were allowed for fragmentation. More details about the settings used for MS experiments are listed in Table 2-1 (*Nat Comm*, Theo).

2.4. Mass spectrometric data analyses

MS data analyses essentially consisted of four parts which are detailed in the following subsections. First, Boris Maček and Stephan Jung performed a standard MS data processing using a protein database containing a single reference amino acid sequence for every protein (section "2.4.1. Searching for peptides contained in standard protein databases"). However, this standard protein database does not contain amino acid polymorphisms resulting from mouse strain-specific genomic variation. Therefore, Mathias Walzer and I predicted strain-specific urinary peptide sequences in a second step (section "2.4.2. Prediction of SAV peptides"). Afterwards, I integrated the predicted SAVs into the standard protein database applied in the first step and reprocessed the MS raw data against this extended database (section "2.4.3. Reprocessing of MS data for SAV peptide identification"). Finally, I confirmed the identity of selected key peptides by manual inspection of their respective MS^2 spectra (section "2.4.4. Manual validation of MS2 spectra").

2.4.1. Searching for peptides contained in standard protein databases

The initial MS data processing was performed with the MaxQuant software[242] version 1.0.14.3 utilising Mascot[249] (Matrix Science) as a search engine for peptide identifications. For database searches a mass tolerance of 7 ppm was used for MS^1 scans and FT-Orbitrap-MS^2 scans whereas 0.5 Da were accepted for LTQ-MS^2 scans. Oxidation of methionine to methionine sulfoxide and acetylation of protein amino termini were set as variable modifications. We utilised a concatenated database containing all entries of the IPI version 3.64 for *Mus musculus*, the ovalbumin sequence from *Gallus gallus*, 262 frequent contaminants and the four caseins of *Bos taurus* which were the dominant protein ingredients for the food of the mice used for MS. All amino acid sequences were not only contained in forward (true) but also in reversed (false) direction summing up to 114,088 entries. This

combination enabled decoy search[132] to control the false discovery rate (FDR). The cut-off posterior error probability (PEP) to achieve an FDR ≤ 1% was determined by MaxQuant. Because the number of peptides finally accepted by MaxQuant was only 112 to 509 per experiment, the FDR of ≤ 1% should be considered as a rough estimation. To further decrease the false positive rate, I confined results to those peptides that had been identified at least once with a Mascot score ≥ 25. Peptide identifications were transferred to unidentified parent ions in other runs of the same experiment if m/z values (within the individual mass error of the respective peptide), retention times (±2 min) and isotope patterns were identical. Therefore, we used the "Match between runs" option in MaxQuant. Peptide identifications with a score above the FDR threshold were also transferred to other runs of the same experiment where the corresponding peptide was identified as rank one candidate for at least one MS^2 spectrum but with a score below the FDR threshold. This was achieved by enabling the option "keep low scoring versions of identified peptides" in MaxQuant data processing (*Nat Comm*, Theo).

During my diploma thesis[203], I searched the peptide lists obtained from MaxQuant for MHC motif peptides using the filter option of Excel. I looked for peptides matching at least one of the following motifs:

H2-K^b: 8-mers: xxxx**F/Y**xx**L/M/I/V**; 9-mers: xxxx**F/Y**xxx**L/M/I/V** or xxxxx**F/Y**xx**L/M/I/V**
H2-D^b: 9-mers: xxxx**N**xxx**M/I/L**; 10-mers: xxxx**N**xxxx**M/I/L**
H2-K^d: 9-mers: x**Y/F**xxxxxx**I/L/V**; 10-mers: x**Y/F**xxxxxxx**I/L/V**
H2-D^d: 9-mers: x**GP**xxxxx**I/L/F**; 10-mers: x**GP**xxxxxx**I/L/F**
H2-L^d: 8-mers: **P/S**xxxxxx**F/L/M**; 9-mers: x**P/S**xxxxxx**F/L/M**
Qa-2: 9-mers: xxxxxx**H**x**L/I/F**

For absolute quantification of peptides synthesised with $^{13}C_5{}^{15}N_1$-valine, I first integrated MS-intensities within its elution time span taking advantage of the "base peak" option of the QualBrowser (Thermo Fisher Scientific). As elution times of isotope labelled and natural peptide forms were always identical, MS-intensity ratios of the two forms could be calculated from exactly the same integration step enabling direct conclusions about the concentration of the natural peptide in the original urine sample (*Nat Comm*, Theo).

2.4.2. Prediction of SAV peptides

The identification of SAV peptides within our urinary peptide data set required multiple rounds of bioinformatic data processing. To predict which peptides are affected by SAVs, we took advantage of the dbSNP database[250,251]. This public resource contains single nucleotide variations (SNVs) of the *Mus musculus* genome relative to the NCBI reference sequence[252] which is based on C57BL/6J mice (SNVs are also known as single nucleotide polymorphisms (SNPs), especially if the SNV is highly

polymorphic within a defined population). First, Mathias Walzer searched for all NCBI reference sequence proteins of *Mus musculus* that contain a peptide sequence identified in urine and retrieved their protein identification numbers (refseq_ids with NP prefix). Therefor he used the sequence file of RefSeq-release 50.11082011, refseq_prot version. The obtained refseq_ids were queried against the UCSC GenomeBrowser Database[253,254] to obtain the respective chromosomal positions encoding the protein. These confined chromosomal regions were in turn used to search the dbSNP XML files (available from ftp://ftp.ncbi.nlm.nih.gov/snp/; Revision 14 April 2010) for dbSNP 'missense' and 'stop-gain' entries (reference SNP (rs) accession numbers). Mathias Walzer then mapped both the retrieved SNVs and the urinary peptides to the respective amino acid positions in the protein. Thereby, he was able to discard SNVs altering those parts of the protein sequence not detected in urine. Consequently, the output only contains SAVs predicted to affect the identified urinary peptides. Relevant stop-gain SNVs did not occur (*Nat Comm*, joint).

When I checked the SNV output of Mathias Walzer`s bioinformatic approach manually via the dbSNP web interface[255] (rs entries, field "Integrated Maps", viewed in early spring and on 13 October 2012), it turned out that there are some contradictions in the gene assignment of the SNVs between the different versions and formats of dbSNP used, i.e. several SNVs that were assigned to genes encoding identified urinary peptides in the XML files were not assigned to these genes in the dbSNP web interface. Such discrepancies may result from different rounds of NCBI`s bioinformatic mapping of SNVs to diverse mouse genome assemblies and are particularly likely if the submitted SNV flanking nucleotide sequence is short. In any case, these conflicts point to SNV gene-annotations of low confidence. To reduce the number of false positive SAVs indicated in our peptide set, I therefore rejected such SNVs from further analyses and do not report them anywhere, shortening the SAV peptide list (see Tables 3-4 and 3-5) by 15 peptides (*Nat Comm*, Theo).

19 urinary peptides could not be screened for SAVs with the bioinformatic approach described above, because they had no corresponding protein in the NCBI reference sequence. Upon closer investigation, I encountered discrepancies in single amino acids between the NCBI reference sequence and the IPI sequence employed for MS peptide identifications (these discrepancies comprise all the SAVs reported in Table 3-7). I utilised the graphical sequence view option of the NCBI reference sequence (NM numbers) to search for SNVs that cause these SAVs. Thereby, the inter-database sequence discrepancies in 18 out of the 19 critical peptides could be explained (*Nat Comm*, Theo).

I subsequently queried the list of SNVs (all reported in Tables 3-4 and 3-5) obtained with our automated and manual approach against the Mouse Phenome Database (MPD)[256,257] to obtain information on the occurrence of the SNVs in laboratory and wild-derived inbred strains of *Mus musculus* and to discover those missense SNVs (by definition those that result in SAVs) that differ between B6, B10 and BALB/c mice (*Nat Comm*, Theo).

2.4.3. Reprocessing of MS data for SAV peptide identification

In database dependent MS approaches, one can only identify peptides whose sequence is contained in the protein database used for MS data processing. Because these databases usually contain only one or a few amino acid sequences (in general the most common ones) for each protein, all peptides containing the "non-standard" version of a SAV will be missed in MS analyses. However, our investigation of SNVs affecting urinary peptides predicted some SAVs differing between B6 and BALB/c mice and revealed that the IPI protein database applied in the first round of MS data analyses (see above) only contained the BALB/c but not the B6 versions of the SAV peptides listed in Table 3-7. Hence, I added the missing SAV forms to our MS search database in all cases with predicted or unknown differences between B6 and BALB/c mice (compare Tables 3-4 and 3-5). In doing so, I also considered the diverse combinations of multiple SAVs situated close to each other in the protein (*Nat Comm*, Theo).

Having complemented the MS search database, I then reprocessed our MS data using the same basic parameters as in the first round of MS analyses but applying the newer MaxQuant version 1.2.2.5 which employs the search engine Andromeda[243] instead of Mascot[249]. The results of this second MS data processing were only used for the identification of SAV peptides differing between B6 and BALB/c mice. To keep the number of false positive identifications low, I only accepted peptides that obtained an Andromeda score of ≥ 80 and for whom related forms (N- and / or C-terminally extended or shortened forms as well as complementary SAV forms) had already been identified in the first round of MS data analyses (compare Supplementary Table 9-4; an Andromeda score of 80 roughly corresponds to a Mascot score of 25 (ref. [243]) and is also used as a cut-off for trustworthy peptides in the software lock mass feature[258] of MaxQuant). Peptides identified in the second but not the first round of MS data processing are incorporated in Tables 3-4, 3-5 and 3-7 as well as Supplementary Table 9-4 but are neither included in the total number of the 639 identified peptides nor in the number of identified SAV peptides (at least 47, all contained in Table 3-4) to avoid double counting of SAV peptide pairs and a distortion of the calculated proportion of SAV peptides among all urinary peptides (*Nat Comm*, Theo).

A special issue arose when I realised that I could not identify the predicted B6-form of the SAV C391H resulting from a double SNV in the gene *Serpina3k* (Table 3-4), although we had identified eleven peptides containing the corresponding BALB/c-form with histidine in urine of BALB/c mice (Table 3-7). I reasoned that this might be due to a modification of the highly reactive cysteine residue of the B6-form. Therefore, I manually searched those 109 entries (12 March 2012) of the Unimod database[219,259] that contain observed modifications of cysteine and looked for cysteine modifications known or expected to occur *in vivo* in healthy mammals. I considered the following cysteine modifications in further analyses (ΔM refers to the monoisotopic mass difference between the modified and the unmodified form of the cysteine residue) (*Nat Comm*, Theo):

Oxidations to:

- cysteine sulfenic acid: Unimod no. 35; ΔM = 15.994915 Da; [260,261]
- cysteine sulfinic acid: Unimod no. 425; ΔM = 31.989829 Da; [260-262]
- cysteine sulfonic acid (cysteic acid): Unimod no. 345; ΔM = 47.984744 Da; [260,261,263]
- *S*-nitrosothiol cysteine: Unimod no. 275; ΔM = 28.990164 Da; [260]

Formation of disulfide bonds:

- *S*-cysteinylcysteine: Unimod no. 312; ΔM = 119.004099 Da; [264-266]
- *S*-homocysteinylcysteine: Unimod no. 1271; ΔM = 133.019749 Da; [267]
- *S*-glutathionylcysteine: Unimod no. 55; ΔM = 305.068156 Da; [264,266]
- *S*-proteinylcysteine: variable ΔM due to variable kinds of bound proteins; Therefore, this modification could not be considered in our MS analyses. [268-270]
- cysteine persulfide: Unimod no. 421; ΔM = 31.972071; [271]
- *S*-cysteaminylcysteine: hypothetical modification based on the well-known occurrence of cystamine (composed of two cysteamins linked via a disulfide bridge) at micromolar concentrations in mammalian blood plasma[265,272]; ΔM = 75.014270 Da
- *S*-γ-glutamylcysteinylcysteine: hypothetical modification based on the well-known occurrence of γ-glutamylcysteine at micromolar concentrations in mammalian blood plasma[265,272]; ΔM = 248.046692 Da

Acylations:

- *S*-myristoylcysteine: ΔM = 210.198365 Da; [273]
- *S*-palmitoylcysteine: Unimod no. 47; ΔM = 238.229666 Da; [273-276]
- *S*-stearoylcysteine: ΔM = 266.260966 Da; [274,275]

Other cysteine modifications:

- nitroalkylation by nitro oleic acid via Michael addition at the cysteine thiol group: Unimod no. 686; ΔM = 327.240959 Da; [277]
- *S*-(2-succinyl) cysteine: Unimod no. 957; ΔM = 116.010959 Da; [278]
- *S*-(4-hydroxynonenal) cysteine: Unimod no. 53; ΔM = 156.115030 Da; [279,280]
- *S*-(cGMP) cysteine: Unimod no. 849; ΔM = 344.039610 Da[219] or 343.031784 Da (manually calculated value based on the molecular formula of cGMP ($C_{10}H_{12}N_5O_7P$) considering that the resulting mass shift should correspond to $C_{10}H_{10}N_5O_7P$ because one H atom must be subtracted for the cysteine thiol group and one H atom at the bound cGMP); both theoretical ΔM values were considered in our preliminary MS searches for cysteine modification explorations; [281]

In preliminary analyses with MaxQuant version 1.2.2.5, I then included all the above listed cysteine modifications as variable modifications and processed our MS data against a very small database containing only the amino acid sequences relevant for the SAV analyses and 16 proteins known as dominant precursors for the identified urinary peptides. Andromeda scores ≥ 80 were obtained for several peptides containing the cysteinylated form of cysteine 391 (i.e. an *S*-cysteinylcysteine) of the serine protease inhibitor A3K but not for any other cysteine modification. Our finding is in accordance with previous observations[265,282] reporting an *S*-cysteinylcysteine residue for a homologous protein of *Mus musculus* serine protease inhibitor A3K, the human alpha-1-antitrypsin

(serpin A1). I confirmed the cysteinylation of the cysteine residue in the urinary peptides by reprocessing our MS data with MaxQuant version 1.2.2.5 applying our *full* MS search database (see above) and using the cysteinylation of cysteine as a variable modification. Oxidation of methionine to methionine sulfoxide and acetylation of protein amino termini were set as variable modifications in all MS data processing. The cysteinylation of *Serpina3k*-derived urinary peptides was also validated by manual investigation of MS^2 spectra (see Supplementary Figures 9-25 and 9-27) (*Nat Comm*, Theo).

2.4.4. Manual validation of MS^2 spectra

Many key peptides of our study were subjected to an additional verification by manual inspection of their MS^2 spectra taking into account alternative peptide candidates suggested by Mascot[249] version 2.2. These manually validated peptides include (*Nat Comm*, Theo):

(1) the 17 MHC motif peptides as well as three extended forms thereof (LNSVFDQLGSY, IDQTRVLNLGPI and SGNFIDQTRVLNLGPITR; Tables 3-1, 3-2 and 3-3, Figure 3 of Sturm *et al.*[207], Supplementary Table 9-3 and Supplementary Figures 9-1 to 9-20)

(2) at least one representative peptide for each SAV that affects a urinary peptide encoded by only one gene (Tables 3-4 and 3-7, Supplementary Tables 9-4 and 9-5, Supplementary Figures 9-21 to 9-40)

Furthermore, the manually validated peptides were also identified by rank 1 when I processed the raw-files using the ProteomeDiscoverer version 1.1 or 1.3 (Thermo Fisher Scientific; employing Mascot) instead of MaxQuant. We could obtain final evidence for the seven peptides synthesised with $^{13}C_5{}^{15}N_1$-valine (cf. Supplementary Table 9-3) by comparing MS^2 spectra as well as nano-HPLC retention times between the isotope-labelled and natural forms. The model SAV peptide pair IVIYHTSAQSIL / FVIYHTSAQSIL used for VSN stimulations (Figure 4 of Sturm *et al.*[207]) was definitely proven by measuring the MS^2 spectra of the $^{15}N_1$-valine labelled synthetic counterparts (Supplementary Figures 9-29 to 9-32). MS reliability parameters for sequence identities are given in Supplementary Tables 9-3 to 9-7. Supplementary Figures 9-1 to 9-40 show MS^2 spectra of manually validated peptides (*Nat Comm*, Theo).

Self-evident, manual validation is more subjective than bioinformatic sequence assignment via mere scoring of the possible candidates. However, some erroneous automatic sequence identifications can be discovered when thoroughly checking the MS^2 spectra applying biochemical expertise. For example, peptides differing only by a leucine versus isoleucine exchange (cf. Figure 3-2) obtain exactly the same score by the bioinformatic search engines. In this case, one of the two top candidate peptides obtains rank 1 and is reported as identified (possibly even with very high score), whereas the

other leucine / isoleucine variant obtains rank 2 and is not reported in the output file although, *a priori*, both peptides have the same probability to be correct identifications.

Fragmentation of peptides − interpreting MS^2 spectra

To be able to manually validate MS^2 spectra, one must be familiar with the chemical rules of peptide fragmentation. Not all of these rules are considered by MS data processing software usually because calculations and thereby data processing times must be kept concise. The cleavage of weak chemical bonds is strongly preferred during low energy CID and HCD yielding characteristic types of fragment ions. Peptide bonds are particularly labile and therefore the main breakpoints in low energy CID of peptides. This results in b-ions containing the N-terminus of the original peptide and y-ions incorporating the corresponding C-terminus. The loss of CO (-27.9949 Da) from b-ions gives rise to a-ions[283,284]. Prominent satellite ions can originate from b-, y- and a-ions by the loss of H_2O (-18.0106 Da) and / or NH_3 (-17.0265 Da) if the peptide contains serine, threonine, aspartate and / or glutamate (loss of H_2O) and arginine, lysine, asparagine and / or glutamine (loss of NH_3) respectively[216,246]. Multiple losses of H_2O and / or NH_3 are likely to occur if the peptide is rich in the mentioned amino acids. If the same peptide molecule fragments at two of its peptide bonds, internal fragment ions originate[285]. Breakage N-terminal of proline residues is energetically favoured, whereas breakage C-terminal of proline is rarely observed[216]. Consequently, if a peptide contains proline apart from its N-terminal position, its y-ion starting with proline and its largest b-ion not containing proline should be particularly abundant. Correspondingly, internal fragment ions with an N-terminal proline are usually more abundant than the other internal fragment ions[285,286].

The charge state of a peptide and its fragment ions is also crucial in manual validation of MS data. In the positive ion mode applied for the present study, the number of charges raises with increasing peptide and fragment length respectively and with the number of contained basic amino acid residues. For example, in our setting, it is very unlikely that a peptide candidate having a molecular mass of less than about 1500 Da and containing no basic amino acid residues is correctly identified if it was not detected in its doubly or triply charged form but only with four or more charges (cf. ref. [216]). In different experimental conditions (concerning e.g. ionisation method, flow rate or solvents), one and the same peptide can have different charge state distributions[219,236]. MS^2 spectra of different charge states of the same peptide usually look very different (cf. Supplementary Figures 9-23 and 9-24). In the positive ion mode, basic amino acid residues (arginine, histidine and lysine) fundamentally determine the relative intensities of the fragment ions. If these amino acids are situated at the N-terminus of the peptide, b-ions are favoured. If they are present at the C-terminus, y-ions will be particularly prominent. Basic amino acid residues in the centre of a peptide support the formation of observable internal fragment ions[285,287].

2.5. RMA-S assay for binding of peptides to MHC molecules

To test MHC binding of urinary MHC motif peptides, we took advantage of the RMA-S cell system. The T-lymphoma cell line RMA-S is of the *H2*^{b} haplotype[288] and due to a lack of the transporter associated with antigen processing (TAP) its endogenous MHC peptide loading is impaired[289] leading to some peptide-free MHC molecules at the cell surface, especially at lower temperatures[290]. These peptide-free MHC molecules are unstable at 37°C unless they have bound a peptide supplied in the medium. Thus, this assay is a measure for the ability of a given peptide to bind to MHC molecules via the increase of MHC molecules at the cell surface (*Nat Comm*, Theo).

I grew RMA-S cells at 37°C in suspension in RPMI 1640 containing 25 mM HEPES and GlutaMAX^{TM}-I (Gibco, Invitrogen; referred to as "basic medium") supplemented with 10% fetal bovine serum (ultra low endotoxin), penicillin, streptomycin, 1 mM sodium pyruvate and non essential amino acids ("complete medium"). The cell suspension was diluted 2- to 4-fold to $5 \cdot 10^5$ cells/ml with complete medium and kept at 26°C for 12-14 h to enrich MHC molecules at the cell surface. Cells were washed with basic medium at 22°C and for each sample $2 \cdot 10^6$ cells were incubated with 200 µl basic medium supplemented with 10^{-5} M of the respective synthetic peptide. Peptide incubations were done with slight movements of 60 rpm at 37°C for 2-3 h and performed in triplicate. All following steps took place on ice and washings as well as reagent dilutions were done using FACS-buffer consisting of phosphate buffered saline (PBS) with 2% fetal bovine serum, 0.02% NaN_3 and 2 mM EDTA (ethylenediaminetetraacetic acid). Peptide-dependent enrichment of MHC molecules was examined with the IgG2b-antibody Y-3 directed against H2-K^{b} [204,205] or the IgG2a-antibody B22-249 targeting H2-D^{b} [206]. Additional samples were stained with the isotype-control antibodies GAP A3 (IgG2a, anti-human leukocyte antigen A3 (HLA-A3))[291] or 1-5C (IgG2b, anti-SYFPEITHI)[217]. Isotype-control stainings showed that there was no or only a negligible increase of unspecific antibody binding due to incubation with peptides. All monoclonal antibodies were produced by Claudia Falkenburger using protein A affinity purification of the filtered supernatants of hybridoma cultures. For staining of RMA-S cells, I applied the respective monoclonal antibody at 10 µg/ml in 50 µl/sample for 30 min. After two washings, each sample was incubated with 50 µl of a 1/200-dilution of FITC-conjugated Affini Pure goat anti-mouse antibody ($F(ab')_2$ fragment specific, Jackson) for 30 min. Three washings followed. I supplemented samples with propidium iodide at 1.7 µg/ml a few seconds before their respective measurement at a FACSCalibur flow cytometer equipped with CellQuestPro software (Becton Dickinson). Between 70,000 and 90,000 living cells per sample were selected for data analysis by their characteristic forward and side scatter signals as well as their low propidium iodide fluorescence. The flow cytometer was jointly operated by Karoline Laske and me (*Nat Comm*, Theo).

2.6. SIINFEKL-H2-K^b-enzyme-linked immunosorbent assay (ELISA)

2.6.1. Principles of the SIINFEKL-H2-K^b-ELISA

One of the most commonly applied immunological assays is the enzyme-linked immunosorbent assay (ELISA). To obtain optimal specificity, it is preferable to combine to different affinity matrices for the detection of the target molecule. Usually, two target-specific antibodies are combined in a so-called sandwich ELISA, where one antibody serves to specifically capture the antigen on the ELISA plate, whereas the second antibody is used as the first reagent for detection.[89]. Because MHC class I peptide ligands are relatively small, it is very difficult to generate two different antibodies that can bind simultaneously to the same MHC peptide ligand molecule, i.e. without mutually hindering each other`s binding. Consequently, to the best of my knowledge, the αSIINFEKL-H2-K^b antibody 25-D1.16 (Porgardor *et al.*[218]) is the only monoclonal antibody that recognizes a mouse-processed MHC class I peptide and that can be combined with a second specific affinity matrix (in this case H2-K^b) for detection by sandwich ELISA. This wonderful option has been realised by Hans-Georg Rammensee already before my diploma thesis and led him to the invention of the concept of the SIINFEKL-H2-K^b-ELISA, which is schematically depicted in Figure 2-3. Peter Overath first put this assay into practise, and I optimised its sensitivity during my diploma thesis[203].

The SIINFEKL-H2-K^b-ELISA relies on H2-K^b molecules, and these were produced by Jeanette Hauger and / or me before my PhD time as detailed in my diploma thesis[203]. The principle is as follows. Two *Escherichia coli* strains are transfected with the genes coding for H2-K^b and the human β_2m MHC subunit respectively. These genes had been optimised for codon usage of *E. coli*. The bacteria accumulate the recombinantly expressed proteins in the form of water-insoluble aggregates, called inclusion bodies, which can be separated by centrifugation and dissolved in 8 M urea. When combining dissolved H2-K^b and β_2m inclusion bodies with a chemically synthesised H2-K^b binding peptide (we applied TEYG**F**LN**L**) at appropriate salt concentrations and at a correctly adjusted redox potential, properly folded MHC peptide complexes will form *in vitro*. We utilised H2-K^b molecules lacking the cytoplasmic domain and the transmembrane domain. Instead, the transgenic H2-K^b protein possesses an additional recognition sequence for biotinylation by the biotin-[acetyl-CoA-carboxylase] ligase (EC 6.3.4.15) BirA. In an overnight *in vitro* reaction, this enzyme attaches a biotin residue to the ε-amino group of a lysine residue within the recognition sequence[292, (Dipl)].

As detailed in section “2.6.3. SIINFEKL-H2-Kb-ELISA in a 5 µl miniaturised format (IMAPlates)” and illustrated in Figure 2-3, the biotinylated MHC peptide complexes can be bound in the wells of a streptavidin coated ELISA plate. When the ELISA plate is incubated with a sample containing H2-K^b binding peptides like SIIN**F**EK**L**, these will replace some of the synthetic peptide (TEYG**F**LN**L** in our case) bound to the MHC molecules. Chen *et al.*[293] suggested that the exchange of MHC peptide ligands is promoted by the removal of the originally bound peptide(s) at pH = 12.5 prior to loading of

the test sample containing the MHC peptide ligands of interest. However, during my diploma thesis, I had found this alkaline stripping of H2-K^b molecules to reduce the sensitivity of the SIINFEKL-H2-K^b-ELISA by about three-fold (see my diploma thesis[203] or Supplementary Figure S5b of Sturm *et al.*[207]). I have therefore not applied this additional experimental step during my PhD thesis. When the H2-K^b binding peptides from the sample have been bound in the ELISA plate, the 25-D1.16 antibody can be used to detect SIINFEKL-H2-K^b complexes, a measure for the SIINFEKL concentration in the solution of interest. Peroxidase coupled to a secondary antibody directed against 25-D1.16 can then be used to catalyse a colour reaction for readout.

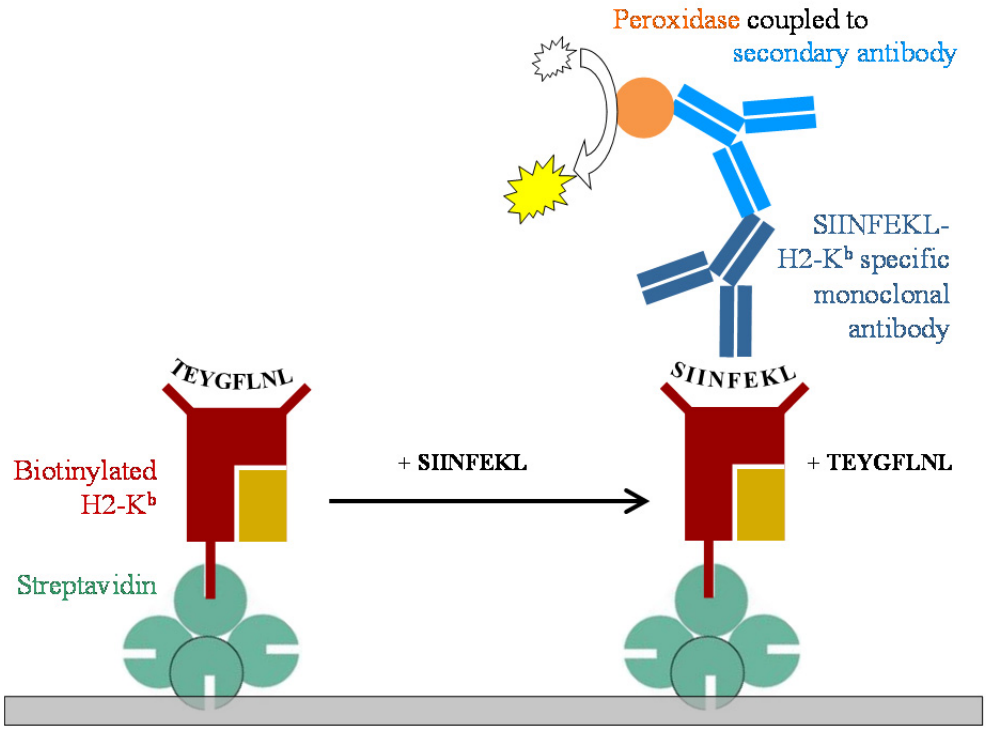


Figure 2-3 | Enzyme-linked immunosorbent assay (ELISA) for the detection of SIINFEKL. Recombinant H2-K^b complexes carrying the peptide TEYGFLNL in their peptide binding pocket are attached via their artificial biotin-residue near the carboxy-terminus to streptavidin on the solid support. TEYGFLNL can be directly displaced by SIINFEKL from the sample, and the resulting complex is detected by the monoclonal antibody 25-D1.16 and secondary reagents (*Nat Comm*, Theo).

2.6.2. SIINFEKL-H2-K^b-ELISA in 50 µl standard format

At the beginning of my PhD thesis (solely relevant for section "3.4.1. Challenging the results of my diploma thesis"), I performed the SIINFEKL-H2-K^b-ELISA in standard format plates with 400 µl wells (F96 MaxiSorp, Nunc^{TM}, Thermo Fisher Scientific). The procedure was exactly the same as described in my diploma thesis[203]. In comparison to the 5 µl miniaturised IMAPlate format described in section "2.6.3. SIINFEKL-H2-Kb-ELISA in a 5 µl miniaturised format (IMAPlates)", the ELISA procedure for the 50 µl standard format differed as follows. ELISA measurements were performed in

duplicate. Volumes of 200 µl/well were used for washings and blocking, whereas coating with streptavidin (1 h) and H2-K^b complexes as well as incubations with SIINFEKL standards (1-2 h) and antibodies were done with 50 µl/well. I applied 100 µl/well of the peroxidase substrate mixture and stopped the colour reaction by addition of 100 µl/well 1 M H_3PO_4. Absorbance at 450 nm was not corrected for absorbance at 650 nm (*Nat Comm*, Theo).

2.6.3. SIINFEKL-H2-K^b-ELISA in a 5 µl miniaturised format (IMAPlates)

The common 400 µl well ELISA format was not sensitive enough to detect SIINFEKL in HPLC fractions derived from 5 ml urine of individual B6/OVA^+ mice. Therefore, we employed a 5 µl miniaturised ELISA based on capillary forces (IMAPlate 5RC96, NCL New Concept Lab, Moehlin, Switzerland[294]). This enabled me to take up lyophilised HPLC fractions in a final volume as small as 17 µl RPMI 1640/1% BSA containing $(NH_4)_2CO_3$ to reach pH ≈ 7.5 for triplicate measurements. In comparison to the 50 µl standard format, the reduced volume improved the limit of detection (LOD) by about factor 6 with respect to the amount of SIINFEKL (Table 3-13) (*Nat Comm*, Theo).

All incubations, except blocking, were done in a moist chamber by contact-free insertion of the IMAPlate into an empty, flat-bottomed, 400 µl well ELISA plate. For all washings, I applied 50 µl/well phosphate-buffered saline/1% bovine serum albumin/0.05% Tween 20 (PBS/BSA/Tween) from the top of the plate. Coating was performed by dipping the IMAPlate into 36 ml 0.05 M Na_2CO_3/$NaHCO_3$ pH 9.6 containing 5 µg/ml streptavidin in a reagent dish (cover of an ELISA plate; incubation 30 min at 37°C). After two washings, the IMAPlate was blocked by putting it into a reagent dish with 70 ml PBS/BSA/Tween for 1 h at room temperature (RT). From the top of the plate, I then loaded 15 µl/well biotinylated H2-K^b complex (6 µg/ml in PBS/BSA/Tween, produced using the H2-K^b peptide TEYGFLNL for refolding[295]) and incubated at 7°C over night. The plate was washed twice and 5 µl of SIINFEKL standards in RPMI 1640/1% BSA or HPLC fractions were pipette-loaded into individual wells from the bottom side of the plate and incubated at RT for approximately 2 h. After two washings, the SIINFEKL-H2-K^b-complex was detected using the monoclonal antibody 25-D1.16 (hybridoma supernatant, produced by Claudia Falkenburger, 1/5 in PBS/BSA/Tween). I either pipette-loaded 5 µl/well from the bottom side or 15 µl from the top side of the plate, incubated 1 h at RT and washed twice. Horseradish peroxidase-coupled goat anti-mouse IgG1 antibody (SouthernBiotech) was dissolved 1/5000 in 12 ml PBS/BSA/Tween in a reagent dish and loaded by dipping. Incubation on ice for 30-60 min and three washings followed. 12 ml each of TMB Peroxidase Substrate and Peroxidase Substrate Solution B (KPL, Gaithersburg, USA) were mixed in a reagent dish and applied to the IMAPlate by dipping. The colour reaction was stopped by dipping the plate briefly in 85% H_3PO_4 and by subsequent inversion until the acid had distributed in

the wells. Absorbance was measured at 450 nm with a SpectraMax 340 ELISA reader (Molecular Devices, Munich) and corrected for absorbance at 650 nm (*Nat Comm*, joint).

2.7. Synthetic peptides

Nicole Zuschke, Patricia Hrstić and Stefan Stevanović synthesised peptides using automated peptide synthesisers 433A (Applied Biosystems, Weiterstadt, Germany) or EPS 221 (Abimed, Langenfeld, Germany) following standard 9-fluorenylmethyloxycarbonyl/*tert*-butyl (Fmoc/tBu) strategy. After removal from the resin by treatment with trifluoroacetic acid (TFA)/triisopropylsilane/water (95/2.5/2.5 by vol.) for 1 h or 3 h (arginine-containing peptides), peptides were precipitated from diethyl ether, washed three times with diethyl ether and resuspended in water prior to lyophilisation. Nicole Zuschke analysed the purity of synthesis products by C_{18}-HPLC (Varian Star, Zinsser, Frankfurt, Germany). Their identity was controlled by Nicole Zuschke and me using electrospray MS (Q-ToF Ultima, Waters, Eschborn, Germany). Purity of synthetic peptides was > 53% for $^{13}C_5{}^{15}N_1$-valine peptides and > 76% for peptides used for VSN experiments. Hubert Kalbacher and I purified synthetic SIINFEKL for IMAPlate ELISA experiments to 98% employing C_8-HPLC. I enriched the SAV peptides IVIYHTSAQSIL and FVIYHTSAQSIL to > 96% purity by C_{18}-HPLC fractionation. This pair had been synthesised with $^{15}N_1$-valine to avoid false positive identifications in future MS experiments resulting from contamination with the synthetic forms (*Nat Comm*, joint).

3. Results

To search for genotype-specific differences in the urinary peptidome, we isolated peptides from urine of different inbred mouse strains and fractionated them by high performance liquid chromatography (HPLC). After this preparative fractionation, we took a dual analytical approach by performing a partial characterisation of the urinary peptidome by MS (discussed in sections "3.1. Abundant peptides with MHC class I binding motif occur in mouse urine in an MHC-independent manner" to "3.3. Further genomic variations influence urinary peptide") and by searching specifically for the prototypic MHC peptide ligand SIIN**F**EK**L** using an MHC-based ELISA with superior sensitivity (section "3.4. Detection of an MHC-dependent peptide in mouse urine"). For the MS analyses, we used an HPLC fraction expected to be particularly rich in MHC ligands, as calibrated with the synthetic peptides SIIN**F**EK**L** and S**Y**FPEITH**I** (anchor residues in bold; see Figure 2-2 and Table 2-1). In an analysis with five mouse strains of the *H2*^{b} or *H2*^{d} haplotype (B10, B10.D2, B6, B6/β_2m$^{-/-}$ and BALB/c, see section "2.1. Mice" for details), we identified 639 peptides with a length between 7 and 32 amino acids. Selected groups of peptides with relevance for genotype-related individual differences will be categorised in the following paragraphs (*Nat Comm*, joint).

3.1. Abundant peptides with MHC class I binding motif occur in mouse urine in an MHC-independent manner

MHC peptide ligands occurring in urine in an MHC-dependent manner are expected to contain an MHC binding motif. To discover such peptides, we first scrutinised the ensemble of the 639 peptides for the presence of MHC binding motifs and discovered 17 peptides fitting to H2-K^{b}, -D^{b}, -K^{d} or -L^{d} motifs (Tables 3-1 and 3-2). However, the urinary occurrence of these 17 peptides did not correlate with the MHC type of the mice, suggesting an MHC-independent production (Table 3-2). This conclusion was confirmed by a quantitative analysis of five peptides that are marked in Table 3-1. Examples listed in Table 3-3 show that the peptide LNSVFDQL with an H2-K^{b} motif (xxxx**F**/**Y**x(x)x**L**/**M**/**I**/**V**) was present at a concentration of 3.5 – 10 · 10^{-7} M, the peptide TRVL**N**LGP**I** with an H2-D^{b} motif (xxxx**N**xx(x)x**M**/**I**/**L**) occurred at 1 – 4 · 10^{-8} M. MHC motif peptides, like other urinary peptides, were derived from proteins of different locations in the body, most prominently from extracellular and membrane proteins typical for urine, blood and kidney (Table 3-1). We also found extended forms of MHC motif peptides, e.g. for TRVL**N**LGP**I** the longer versions IDQTRVL**N**LGP**I,** TRVL**N**LGP**I**TR and SGNFIDQTRVL**N**LGP**I**TR occurred, pointing to a gradual proteolytic generation. A multitude of further examples for series of N- and / or C-terminally extended MHC motif peptides can be found in my diploma thesis[203], (*Nat Comm*, joint).

Table 3-1 | Urinary peptides with MHC class I binding motif (*Nat Comm*, Theo).

H2 allele	Peptide	Origin of peptide	Location of protein[108]
K^b	NKQE**F**GW**I**	Carboxylesterase 1B, 1C, 1D and / or 1E	e, i
	*LNSV**F**DQ**L**	Kidney androgen-regulated protein	e
	LAPQP**F**LR**V**	Leucine-rich alpha-2-glycoprotein	e, mem (?)
D^b	FNIQ**N**REPL**I**	ATP-binding cassette, sub-family A, member 13	mem
	KELQ**N**SIID**L**	Kidney androgen-regulated protein	e
	*TRVL**N**LGP**I**	Uromodulin (Tamm-Horsfall urinary glycoprotein)	e, mem
K^d	*V**Y**RPDQVSI**L**	Deoxyribonuclease-1	e, i
	L**Y**WVDVERQ**V**	Pro-epidermal growth factor	mem
	L**F**KDSAFG**L**	Serotransferrin	e
L^d	Y**S**MPPIVR**F**	Alpha-1-antitrypsin 1-3 and / or 1-4	e
	***P**AVRGFS**L**	LDL-receptor-related protein 2 (Megalin)	mem
	SSDIKER**F**	Major urinary protein (No. 1 – 8, 11 and / or 12)	e
	L**S**SDIKER**F**	II	e
	PSFVPLSK**F**	Napsin-A (Kidney-derived aspartic protease-like protein)	lysosomal
	***P**VESKIY**F**	Pro-epidermal growth factor	mem
	L**S**RQMGMV**F**	II	mem
	G**P**VQGTIH**F**	Superoxide dismutase [Cu-Zn]	i, e

Urine of five mouse strains (B10, B10.D2, B6, B6/$\beta_2m^{-/-}$ and BALB/c) was collected in metabolic cages over night. For each strain, urine mixtures of 4 to 5 mice were subjected to MS analysis. The MHC anchor amino acids are marked in bold. Asterisks indicate peptides that were chosen for quantification (see Table 3-3 for two examples). e, extracellular; i, intracellular; mem, plasma membrane. Data of this table were generated during my diploma thesis but are displayed as a basis for newer data.

Two of these peptides, LNSV**F**DQ**L** and TRVL**N**LGP**I**, were tested for their ability to bind to MHC class I molecules using an assay based on the peptide-induced stabilisation and thereby increase of MHC molecules on the surface of RMA-S cells (*H2*^{b} haplotype)[296,297]. As expected, LNSV**F**DQ**L** showed binding to H2-K^b and TRVL**N**LGP**I** did assemble with H2-D^b molecules (Figure 3-1; see section "2.5. RMA-S assay for binding of peptides to MHC molecules" for a description of the method). Together, these results provide clear evidence for the existence of a pool of MHC motif peptides in urine that is independent of MHC expression and, therefore, these peptides cannot transmit information on the MHC type of an individual. For reasons detailed below, the detection of *bona fide* MHC-dependent peptides by MS is not straightforward (*Nat Comm*, joint).

Table 3-2 | Occurrence of peptides with MHC binding motifs of the *H2*^{b} and *H2*^{d} haplotypes in urine of five inbred mouse strains (*Nat Comm*, Theo).

			Mean of normalised MS intensities, in 1000					Standard deviation of normalised MS intensities, in 1000				
		Mouse strain	B10	B10.D2	B6	B6/β_2m$^{-/-}$	BALB/c	B10	B10.D2	B6	B6/β_2m$^{-/-}$	BALB/c
H2 allele	Peptide	Exp. no.	*b*	*d*	*b*	*(b)*	*d*	*b*	*d*	*b*	*(b)*	*d*
K^b	NKQE**F**GW**I**	3			9185							
	LNSV**F**DQ**L**	1			12860	2900	474			1464	2680	820
	"	2	1674	827	3995	1246	1033	700	199	1518	1112	1789
	LAPQP**F**LR**V**	1			82	229	1118			115	97	470
D^b	FNIQ**N**REPL**I**	2	0	4	0	21	0	0	4	0	28	0
	KELQ**N**SIID**L**	1			398	144	41			79	249	72
	"	2	32	97	89	0	141	56	107	155	0	56
	TRVL**N**LGP**I**	1			8523	17291	6335			15	7756	2468
		2	3466	2144	3662	4863	1737	2152	2130	431	3039	1531
K^d	V**Y**RPDQVSI**L**	1			122	74	495			44	64	133
	"	2	76	87	12	53	98	26	37	10	28	44
	L**Y**WVDVERQ**V**	1			676	167	18			195	150	30
	L**F**KDSAFG**L**	1			52	99	752			74	44	204
	"	2	0	0	0	0	27	0	0	0	0	46
L^d	Y**S**MPPIVR**F**	2	80	74	0	102	0	28	76	0	134	0
	PAVRGFS**L**	2	60	97	0	97	135	5	48	0	137	119
	SSDIKER**F**	1			0	209	42			0	141	38
	L**S**SDIKER**F**	1			23	132	17			10	79	29
	PSFVPLSK**F**	1			17	32	107			25	5	25
	"	2	42	32	0	0	0	23	15	0	0	0
	"	3			3697							
	PVESKIY**F**	2	69	133	34	239	56	23	79	30	196	20
	L**S**RQMGMV**F**	1			1148	2099	329			539	606	237
	"	2	119	205	504	628	22	27	120	123	421	20
	G**P**VQGTIH**F**	1			0	24	170			0	9	72

Fractions of urine mixtures of 4-5 mice per mouse strain were analysed by MS. In exp. no. 1 and 2 MS intensities were normalised so that the mean of the sum of intensities of all identified peptides was equal for all mouse strains. MS intensities correlate only very loosely with peptide concentrations and should therefore be interpreted more qualitatively than quantitatively. An HPLC fraction was measured in triplicate for exp. no. 1 and 2 whereas we performed only a single LC-MS run of an HPLC fraction for exp. no. 3. Peptides with an H2-D^d or Qa-2 motif were not found. MS reliability parameters for peptides in this table are specified in Supplementary Table 9-3. Data of this table were generated during my diploma thesis but are displayed for ease of comparison with newer data (e.g. Table 3-3).

Table 3-3 | Urinary concentrations of two peptides with MHC class I binding motif and their extended versions (*Nat Comm*, Theo).

		Urinary concentration in 10^{-8} M				
	Mouse strain	B10	B10.D2	B6	B6/β_2m$^{-/-}$	BALB/c
Peptide	*MHC allele*	*b*	*d*	*b*	*(b)*	*d*
LNSV**F**DQ**L**	*H2-K^b*	97	100	44	68	35
LNSV**F**DQ**L**GSY	-	99	154	212	265	194
Ratio [LNSV**F**DQ**L**] / [LNSV**F**DQ**L**GSY]		*0.98*	*0.65*	*0.21*	*0.26*	*0.18*
TRVL**N**LGP**I**	*H2-D^b*	4.1	3.0	0.9	1.4	4.1
IDQTRVL**N**LGP**I**	-	11.9	15.6	14.4	20.0	14.1
Ratio [TRVL**N**LGP**I**] / [IDQTRVL**N**LGP**I**]		*0.35*	*0.19*	*0.06*	*0.07*	*0.29*

The MS-based quantification was done with the help of corresponding synthetic $^{13}C_5{}^{15}N_1$-valine-containing peptides spiked into manually collected urine of five mouse strains. Ratios for urinary concentrations of the MHC motif peptides and their extended forms were calculated in order to correct for possible strain specific differences in protein expression.

The MHC-independent MHC motif peptides have exactly the same principal structural features as *bona fide* MHC peptide ligands concerning peptide length and anchor amino acids at certain positions. These properties had been described to be crucial for stimulation of both VSNs[96] and OSNs[97]. Furthermore, at least the tested examples of urinary MHC motif peptides are indistinguishable from *bona fide* MHC peptide ligands with regard to their ability to bind to MHC molecules according to the *in vitro* assay that I applied (Figure 3-1). Consequently, I reasoned that they should also be detected by VSNs and OSNs. We synthesised the H2-D^b motif peptide TRVL**N**LGP**I** and two of its longer forms, IDQTRVLNLGPI and SGNFIDQTRVNLGPITR. Testing these peptides individually at a concentration of 10^{-11} M, Trese Leinders-Zufall could readily identify VSNs responding selectively to TRVL**N**LGP**I**, and some of the neurons also responded to the extended versions (Fig. 3 in Sturm *et al.*[207]).

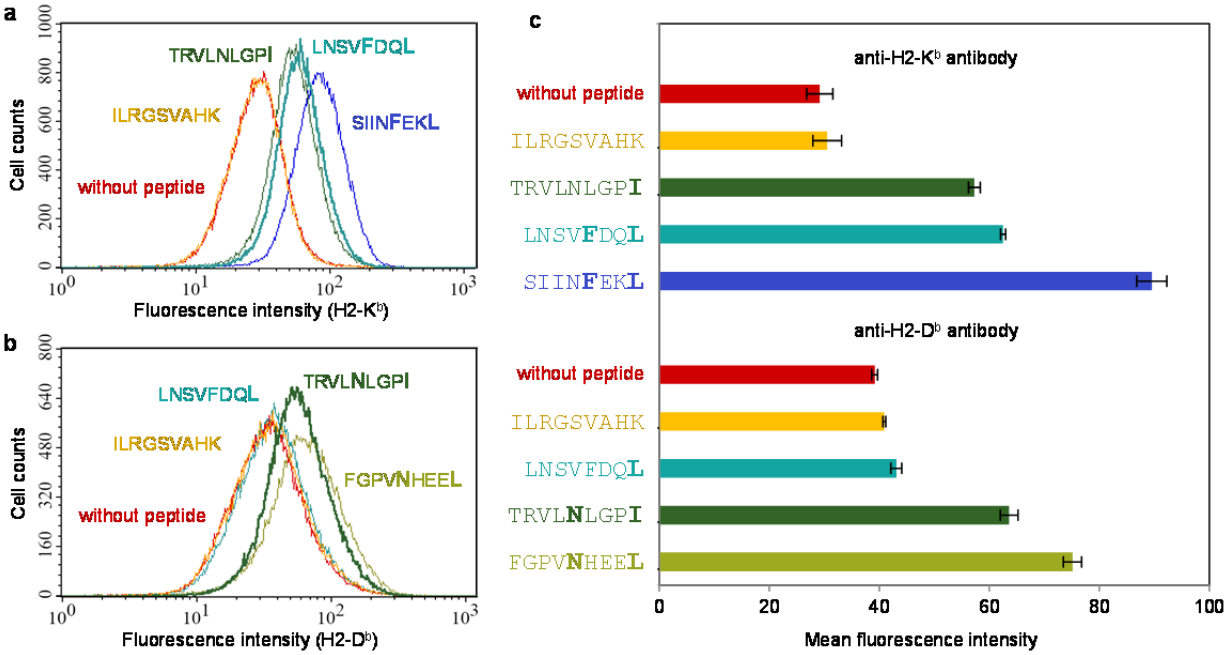


Figure 3-1 | The MHC motif peptides LNSVFDQL and TRVLNLGPI present in urine of mice bind to MHC molecules on RMA-S cells. All peptide incubations were done with synthetic peptides at 10^{-5} M and performed in triplicate. Peptide-dependent enrichment of H2-K^b and H2-D^b molecules at the cell surface was measured by flow cytometry using the monoclonal antibodies Y-3 and B22-249, respectively. From each triplicate, the sample with the median MFI (mean fluorescence intensity) is depicted in **a** or **b**. The mean MFI of each triplicate measurement is represented by the columns in **c** with the corresponding standard deviations illustrated by the error bars. Amino acids that represent classical anchor residues for the MHC allele tested are printed in bold. The MHC-peptide ILRGSVAHK, a natural ligand of the human leukocyte antigen A3 (HLA-A3), serves as a negative control lacking anchor residues typical for peptides binding to H2-K^b and -D^b. SIINFEKL is used as a positive control for binding to H2-K^b, whereas the natural H2-D^b ligand FGPVNHEEL[298] fulfills this purpose for H2-D^b. The H2-K^b motif peptide LNSVFDQL clearly binds to H2-K^b (**a** and **c**) but not to H2-D^b (**b** and **c**). TRVLNLGPI containing an H2-D^b motif binds to H2-D^b but also to H2-K^b although positions 5 or 6 are not occupied by the classical H2-K^b anchors (phenylalanine or tyrosine). However, its carboxy-terminal isoleucine is an anchor residue for H2-D^b as well as H2-K^b and at least one natural H2-K^b peptide, SLVELTSL, with leucine at the anchor position 5 has been described[299, (*Nat Comm*, Theo)].

3.2. The urinary peptidome comprises a respectable set of peptides with single amino acid variations (SAVs)

Given the new observation that VSNs can detect MHC-independent MHC motif peptides as well as extended forms thereof, we asked for the biological information that mice could possibly extract from this. My first trivial thought was that such peptides might be used for discrimination of species because the proteome clearly differs between different taxa, and this will necessarily be reflected in the urinary peptidome. However, the more interesting question is whether urinary peptides also differ within *Mus musculus* establishing the possibility that these molecules function as signals of individuality in a multitude of social behaviours. After I had performed some initial investigations on mouse strain-specific urinary MUP peptides (see section "3.3. Further genomic variations influence urinary peptide"), Hans-Georg Rammensee came up with the great idea to exploit genomic databases to identify SNVs resulting in SAVs of urinary peptides.

Accordingly, I asked our bioinformatician, Mathias Walzer, to search the dbSNP database[250] for non-synonymous single nucleotide variations (SNVs) affecting the sequence of urinary peptides by the exchange of single amino acids. Out of the 639 peptides, at least 47 (> 7%) turned out to be encoded only in some but not all laboratory or wild-derived inbred mouse strains (Tables 3-4 and 3-5). Predicted peptides with single amino acid variations (SAVs) occur even within related inbred strains of mice, e.g. within the C57 lineage, but are especially frequent between mice of different *Mus musculus* subspecies (Table 3-6). Importantly, at least four SAVs of urinary peptides were also anticipated to differ between B6 and BALB/c mice. Including these amino acid sequence variations in the bioinformatic processing of our MS spectra, I could identify several B6-BALB/c discriminating SAV peptide pairs (Table 3-7; also see sections "2.4.2. Prediction of SAV peptides" and "2.4.3. Reprocessing of MS data for SAV peptide identification"), and their distribution corresponded exactly to the nucleotide variations encoded in these mice, with B10 mice being more related to B6 than to BALB/c mice. For example, at position 398 of serine protease inhibitor A3K, a blood plasma protein, B6J and B10J mice encode an isoleucine, whereas BALB/cJ mice encode phenylalanine. Accordingly, we detected the resulting urinary peptide IVIYHTSAQSIL only in mice of B6 or B10 background and the corresponding SAV form FVIYHTSAQSIL only in BALB/c mice (Tables 3-4, 3-5, 3-6 and 3-7). The urinary concentration of SAV peptides is in the same order of magnitude ($\leq 10^{-6}$ M) as the MHC motif peptides described above. This is demonstrated by the fact that they are readily detected by MS and, more accurately, by the quantification of LNSV**F**DQ**L** (Table 3-3), an MHC motif peptide of both B6 and BALB/c mice, that is encoded as LNSV**F**DR**L** in *Mus musculus musculus* (*Kap* gene coding for kidney androgen-regulated protein, dbSNP accession number rs31887216, see Tables 3-4 and 3-6) (*Nat Comm*, Theo).

Table 3-4 | Identified urinary SAV peptides encoded by no more than one gene, which is the only gene fitting to the underlying SNV (*Nat Comm*, Theo).

Gene	Protein name	NCBI Reference Sequence	Accession number of SNV in dbSNP	Number of ss entries for SNV	SAV resulting from SNV[a]	Difference between B6 and BALB/c mice	SAV peptides detected in urine
Abca13	ATP-binding cassette sub-family A member 13	NP_839990.2	rs50794155	3	**S**747F	no[(MPD)]	LSNPSGSFPAPEFDNL
Alb	Serum albumin	NP_033784.2	rs13468473	1	**Y**34C	no[(MS: B6 + BALB/c)]	EAHKSEIAHRY EAHKSEIAHRYND EAHKSEIAHRYNDL EAHKSEIAHRYNDLGEQHFKGLVL IAHRYNDLGEQHFKGLVL AHRYNDLGEQHFKGLVL RYNDLGEQHFKGLVL YNDLGEQHFKGLVL
			rs13468471	1	**E**39V	no[(MS: B6 + BALB/c)]	EAHKSEIAHRYNDLGEQHFKGLVL IAHRYNDLGEQHFKGLVL AHRYNDLGEQHFKGLVL RYNDLGEQHFKGLVL YNDLGEQHFKGLVL NDLGEQHFKGLVL GEQHFKGLVL
			rs13468469	1	**K**561E	no[(MS: BALB/c)]	LVKHKPKATAEQLKTVMDDF
			rs31706341	2	**A**564S	no[(MPD)]	LVKHKPKATAEQLKTVMDDF
Ces1c	Carboxylesterase 1C	NP_031980.2	rs51686553	4	F461**S**	yes[(MPD)]	SVFGAPLLKEGASEEETNL
Egf	Pro-epidermal growth factor	NP_034243.2	rs30619049	2	**V**103L	no[(MPD)]	LYWVDVERQV YWVDVERQVL WVDVERQVL
Kap	Kidney androgen-regulated protein	NP_034724.1	rs31887216	2	**Q**44R	no[(MPD)]	LLNSVFDQLGSY LNSVFDQL LNSVFDQLG LNSVFDQLGSY LNSVFDQLGSYR LNSVFDQLGSYRGTKAPLED NSVFDQLGSY SVFDQLGSY

(Table continued on next page)

Table 3-4 (continued)

Gene	Protein name	NCBI Reference Sequence	Accession number of SNV in dbSNP	Number of ss entries for SNV	SAV resulting from SNV[(a)]	Difference between B6 and BALB/c mice	SAV peptides detected in urine
Kap	Kidney androgen-regulated protein	NP_034724.1	rs31887214	2	**A**53S	no[(MPD)]	LNSVFDQLGSYRGTK**A**PLED
			rs13462485	1	**P**54S	no[(MS: BALB/c)]	LNSVFDQLGSYRGTKA**P**LED
Lrg1	Leucine-rich alpha-2-glycoprotein	NP_084072.1	rs13467361	2	**S**100F	no[(MPD)]	SSNRLQAL**S**PELLAPVPRLRA
							LQAL**S**PELLAPVPR
							QAL**S**PELLAPVPRL
							SPELLAPVPRL
			rs13467359	3	**A**112V	no[(MPD)]	SSNRLQALSPELLAPVPRLR**A**
			rs13467360	1	**Q**153H	no[(MS: BALB/c)]	L**Q**GLDALGHLDLAE
Prol1	Mucin apoprotein	NP_032670.2	rs51790902	2	**G**52S	no[(MPD)]	**G**FIPSSPKFP
Serpina1b	Alpha-1-antitrypsin 1-2	NP_033270.3	rs33061356	2	**D**313N	no[(MS: BALB/c)]	VQIHIPRLSISG**D**YN
Serpina3k	Serine protease inhibitor A3K	NP_035588.2	rs33123758	5	**V**386**I**	yes[(MPD)]	DVAETGTEAAAATGVIGGIRKA**V**L
							DVAETGTEAAAATGVIGGIRKA**I**L
							VAETGTEAAAATGVIGGIRKA**I**
							VAETGTEAAAATGVIGGIRKA**I**L
							VAETGTEAAAATGVIGGIRKA**I**LPAVHFNRP
							AAATGVIGGIRKA**I**L
							AAATGVIGGIRKA**I**LPAVHFNRP
							AATGVIGGIRKA**I**L
							AATGVIGGIRKA**I**LPAVHFNRP
							ATGVIGGIRKA**I**L
							GVIGGIRKA**I**L
							GIRKA**I**LPAVHFNRP
							IRKA**I**LPAVHFNRP
							IRKA**I**LPAVHFNRPF
							IRKA**I**LPAVHFNRPFL
							KA**I**LPAVHFNRPF
							KA**V**LPAVcFNRPFL
							KA**I**LPAVHFNRPFL

(Table continued on next page)

Table 3-4 (continued)

Gene	Protein name	NCBI Reference Sequence	Accession number of SNV in dbSNP	Number of ss entries for SNV	SAV resulting from SNV[a]	Difference between B6 and BALB/c mice	SAV peptides detected in urine
Serpina3k	Serine protease inhibitor A3K	NP_035588.2	the combination of both rs261641732 and rs225117050[b]	1 per SNV	**C**391**H**[b]	yes[(MS: B6 / BALB/c)]	VAETGTEAAAATGVIGGIRKAILPAV<u>H</u>FNRP AAATGVIGGIRKAILPAV<u>H</u>FNRP AATGVIGGIRKAILPAV<u>H</u>FNRP GIRKAILPAV<u>H</u>FNRP IRKAILPAV<u>H</u>FNRP IRKAILPAV<u>H</u>FNRPF IRKAILPAV<u>H</u>FNRPFL KAILPAV<u>H</u>FNRPF KAVLPAV<u>c</u>FNRPFL KAILPAV<u>H</u>FNRPFL PAV<u>c</u>FNRP PAV<u>H</u>FNRPF PAV<u>c</u>FNRPFL PAV<u>H</u>FNRPFL
			rs8273122	5	**I**398**F**	yes[(MPD)]	<u>I</u>VIYHTSAQSIL <u>F</u>VIYHTSAQSIL

Each of the 55 peptides listed in this table can be uniquely assigned to only one gene regardless of the SAV form. Therefore, a variation of the indicated SAV forms will lead to the loss of one SAV peptide form in urine and the appearance of a new one (compare Table 3-5). Note that some peptides are listed twice in this table because they can contain two distinct SAVs at different positions. Each SNV submitted to the dbSNP database obtains a submitted SNP identification number (ss#). Several ss numbers referring to the same SNV are assembled into a reference SNP cluster labelled with an rs number. Consequently, the number of ss entries for each SNV (rs entry) correlates positively with its genomic reliability and is used as a parameter of confidence. The reported deviation from the standard sequence should be considered with particular caution if there is only one ss entry for the SNV. MS reliability information for peptides in this table is given in Supplementary Tables 9-4 and 9-5 as well as in Supplementary Figures 9-21 to 9-40. If the Mouse Phenome Database (MPD) contains information on the distribution of SNV alleles in mouse strains, the respective rs number is printed in blue, and the MPD data is reported in Table 3-6. SAV positions in urinary peptides are underlined and the SAV forms identified by our MS approach are highlighted in green. The small letter "c" in peptide sequences indicates cysteine residues that occurred in a cysteinylated form, i.e. these cysteines had bound free cysteine via a disulfide bridge.

Annotations

(a) The SAV form occurring in C57BL/6J mice corresponds to the NCBI reference sequence and is always indicated in front of the SAV position number. In contrast, the non-reference ("non-B6") SAV form is denoted behind the SAV position number.

(b) The *Serpina3k* SNV rs261641732 alone is predicted to cause the SAV C391R, the *Serpina3k* SNV rs225117050 alone will result in SAV C391Y. However, because both SNVs affect the same codon, their co-occurrence leads to the SAV C391H, thereby changing the protein sequence from the C57BL/6J-form (as given in NCBI Reference Sequence NP_035588.2) to the BALB/c-form (the standard form of UniProtKB/Swiss-Prot entry P07759 and of IPI).

(MPD) According to the data from the Mouse Phenome Database (MPD)[256,257], which are reported in Table 3-6.

(MS: BALB/c) According to our MS data: we detected the NCBI reference form of the SAV (by definition the form occurring in C57BL/6J mice) in BALB/c mice.

(MS: B6 + BALB/c) According to our MS data: we detected the reference form of the SAV in both B6 and BALB/c mice (B6/$\beta_2m^{-/-}$ and B6 mice are congenic and therefore considered as a single group in SAV analyses).

Table 3-5 | Identified urinary SAV candidate peptides that are encoded by more than one gene and which therefore may or may not be affected by the presence of a missense SNV (*Nat Comm*, Theo)**.**

Genes encoding peptide	Genes to which the SNV was assigned by dbSNP	Protein name	NCBI Reference Sequence	Accession number of SNV in dbSNP	Number of ss entries for SNV	SAV resulting from SNV[a]	SNV differing between B6 and BALB/c mice	SAV candidate peptides detected in urine	Peptide encoded more than one time in the genome of *Mus musculus* with SAV in non-B6 form[b]
Acta1	*Acta1*	Actins: alpha skeletal muscle	NP_033736.1	rs13462314	1	**D**25G	no[ENA]	AGFAG**D**DAPR	no
Acta2		aortic smooth muscle	NP_031418.1						
Actb		cytoplasmic 1	NP_031419.1						
Actc1		alpha cardiac muscle 1	NP_033738.1						
Actg1		cytoplasmic 2	NP_033739.1						
Actg2		gamma-enteric smooth muscle	NP_033740.2						
Amy2a2		Pancreatic alpha-amylase: 2A2	NP_001153624.1	rs13460710	1	**L**298**M**[c]	?	LKNWGEGWG**L**VPSDRAL	?[c]
Amy2a3	*Amy2a3*	2A3	NP_001153623.1					LKNWGEGWG**M**VPSDRAL	
Amy2a4		2A4	NP_001153622.1						
Amy2a5	*Amy2a5*	2A5	NP_001036176.1						
Amy2b [c]		2B isoform 1	NP_001177332.1						
		2B isoform 2	NP_001177333.1						
Ces1a		Carboxylesterase: 1A	NP_001013786.2	rs50445108	1	**W**514C	no[MPD]	FARNGNPNGEGLPH**W**PEY	no
Ces1b		1B	NP_001074841.1						
Ces1c		1C	NP_031980.2						
Ces1g	*Ces1g*	1G	NP_067431.2						
Eef1a1	*Eef1a1*	Elongation factor 1-alpha 1	NP_034236.2	rs13470327	1	**T**21S	?	S**T**TTGHLIYK	no
Eef1a2		Elongation factor 1-alpha 2	NP_031932.1						
	LOC101056619*	(Elongation factor 1-alpha 1-like)	XP_003946341.1-XP_003946350.1 XP_003946352.1						
Hba-a1	*Hba-a1*	Hemoglobin subunit alpha	NP_032244.2	rs13459861	1	**K**40R	?	ASFPTT**K**TYFPHF	no
Hba-a2	*Hba-a2*	Hemoglobin alpha, adult chain 2	NP_001077424.1						
	*Hba-4ps**								
Hba-a1	*Hba-a1*	Hemoglobin subunit alpha	NP_032244.2	rs13459859	1	**Y**42C	?	ASFPTTKT**Y**FPHF	no
Hba-a2	*Hba-a2*	Hemoglobin alpha, adult chain 2	NP_001077424.1						
	*Hba-4ps**								

(Table continued on next page)

Table 3-5 (continued)

Genes encoding peptide	Genes to which the SNV was assigned by dbSNP	Protein name	NCBI Reference Sequence	Accession number of SNV in dbSNP	Number of ss entries for SNV	SAV resulting from SNV[a]	SNV differing between B6 and BALB/c mice	SAV candidate peptides detected in urine	Peptide encoded more than one time in the genome of *Mus musculus* with SAV in non-B6 form[b]
Hbb-b1 (=*Hbb-t1*)	*Hbb-b1* (=*Hbb-t1*)	Hemoglobin subunit beta-1	NP_032246.2	rs13463743	1	**Y**35C[d]	no[300]	VV**Y**PWTQRYFDS	no
Hbb-b2 (=*Hbb-t2*)	*Hbb-b2* (=*Hbb-t2*)	Hemoglobin subunit beta-2	NP_058652.1						
(*Beta-s* = *Hbb-t1* and *Hbb-t2*)	LOC100503686* LOC100503605 (1)*		NP_001188320.1						
Hbb-b1 (=*Hbb-t1*)	*Hbb-b1* (=*Hbb-t1*)	Hemoglobin subunit beta-1	NP_032246.2	rs13463746	1	**S**44R[e]	no[300]	VVYPWTQRYFD**S**	no
Hbb-b2 (=*Hbb-t2*)	*Hbb-b2* (=*Hbb-t2*)	Hemoglobin subunit beta-2	NP_058652.1						
(*Beta-s* = *Hbb-t1* and *Hbb-t2*)	LOC100503686* LOC100503605 (1)*		NP_001188320.1						
Klk1	*Klk1*	Kallikrein-1	NP_034769.4	rs16784507	1	**D**139A	no[(ss)]	MLLRLKKPADITDVVKPI**D**L	no
Klk1b5	*Klk1b5*	Kallikrein 1-related peptidase b5	NP_032482.1					LLRLKKPADITDVVKPI**D**L	no
								LRLKKPADITDVVKPI**D**	no
								LRLKKPADITDVVKPI**D**L	no
								RLKKPADITDVVKPI**D**L	no
								LKKPADITDVVKPI**D**L	no
Klk1	*Klk1*	Kallikrein-1	NP_034769.4	rs16784507	1	**D**139A	no[(ss)]	ITDVVKPI**D**L	yes
Klk1b5	*Klk1b5*	Kallikrein 1-related peptidase b5	NP_032482.1						
Klk1b16		Kallikrein 1-related peptidase b16	NP_032480.1						
Klk1b22		Kallikrein 1-related peptidase b22	NP_034244.1						
form with A at pos. 139:									
Egfbp2		Epidermal growth factor-binding protein type B	NP_034245.2						
Klk1b1		Kallikrein 1-related peptidase b1	NP_034775.1						
Klk1b9		Kallikrein 1-related peptidase b9	NP_034246.1						
Klk1b26		Kallikrein 1-related peptidase b26	NP_034774.1						

(Table continued on next page)

Table 3-5 (continued)

Genes encoding peptide	Genes to which the SNV was assigned by dbSNP	Protein name	NCBI Reference Sequence	Accession number of SNV in dbSNP	Number of ss entries for SNV	SAV resulting from SNV[(a)]	SNV differing between B6 and BALB/c mice	SAV candidate peptides detected in urine	Peptide encoded more than one time in the genome of *Mus musculus* with SAV in non-B6 form[(b)]
Serpina1a		Alpha-1-antitrypsin 1-1 isoform 1	NP_033269.1	rs31018094	4	**I**52L	yes[(MPD)]	SPASHEIATNLGDFAI	yes
		Alpha-1-antitrypsin 1-1 isoform 2	NP_001239498.1						
Serpina1b		Alpha-1-antitrypsin 1-2	NP_033270.3						
Serpina1c	*Serpina1c*	Alpha-1-antitrypsin 1-3	NP_033271.1						
Serpina1e	*Serpina1e*	Alpha-1-antitrypsin 1-5	NP_033273.1						
form with L at pos. 52:									
Serpina1d	*Serpina1d;* non-coding locus*	Alpha-1-antitrypsin 1-4	NP_033272.1						
Serpina1a	*Serpina1a*	Alpha-1-antitrypsin 1-1 isoform 1	NP_033269.1	rs13462221	1	L403**V**	?	IIFEEHTQSPIFVG	yes
		Alpha-1-antitrypsin 1-1 isoform 2	NP_001239498.1						
form with V at pos. 403:									
Serpina1b		Alpha-1-antitrypsin 1-2	NP_033270.3						
Serpina1d	*Serpina1d*	Alpha-1-antitrypsin 1-4	NP_033272.1						
Serpina1a		Alpha-1-antitrypsin 1-1 isoform 1	NP_033269.1	rs33060453	2	**K**317E	yes[(MPD)]	KTLMSPLGI	yes
		Alpha-1-antitrypsin 1-1 isoform 2	NP_001239498.1					KTLMSPLGITRI	yes
Serpina1b	*Serpina1b*	Alpha-1-antitrypsin 1-2	NP_033270.3						
Serpina1c		Alpha-1-antitrypsin 1-3	NP_033271.1						
Serpina1d		Alpha-1-antitrypsin 1-4	NP_033272.1						
form with E at pos. 317:									
Serpina1e		Alpha-1-antitrypsin 1-5	NP_033273.1						
Serpina1c	*Serpina1c*	Alpha-1-antitrypsin 1-3	NP_033271.1	rs51746614	4	**V**382L	yes[(MPD)]	YSMPPIVRF	no
Serpina1d		Alpha-1-antitrypsin 1-4	NP_033272.1					YSMPPIVRFD	no
Serpina1c	*Serpina1c*	Alpha-1-antitrypsin 1-3	NP_033271.1	rs51746614	4	**V**382L	yes[(MPD)]	MPPIVRFD	yes
Serpina1d		Alpha-1-antitrypsin 1-4	NP_033272.1						
form with L at pos. 382:									
Serpina1a		Alpha-1-antitrypsin 1-1 isoform 1	NP_033269.1						
		Alpha-1-antitrypsin 1-1 isoform 2	NP_001239498.1						
Serpina1b		Alpha-1-antitrypsin 1-2	NP_033270.3						

(comments to the table on next page)

Table 3-5 (see previous pages): comments and explanations

Although gene conversion can result in identical SNVs occurring in two or more paralogous genes[300], the assignment of a SNV to multiple paralogous genes by the dbSNP database[250,251] could also be a technical artefact: due to the shortness of the established SNV flanking nucleotide sequences, it is often difficult to assign the SNV to a single gene. However, in most cases, except those labelled with an asterisk, all genes assigned to contain the SNV also encode the peptide that we identified. Because we do not know the proportionate contribution of these genes to the urinary peptide concentration, the effect of a SNV in a single gene can range from zero (if the SNV affects a gene not contributing to the urinary peptide concentration) over quantitative (if the SNV gene contributes to the urinary peptide concentration together with other genes) to "all or nothing" (if the SNV gene is the only one relevant for the urinary peptide occurrence). Due to these uncertainties, peptides listed in this table are not considered in the reported minimal percentage of SAV peptides among all urinary peptides, which we determined to be > 7% (> 47 out of 639 peptides) among laboratory or wild-derived inbred mouse strains. The genes encoding the peptide were determined by performing an NCBI Standard Protein BLAST[301] (blastp) against NCBI reference proteins[252] of *Mus musculus*. The matching NCBI reference sequence protein accession numbers (NP_…, XP_...) are indicated in addition to the gene names. Each SNV submitted to the dbSNP obtains a submitted SNP identification number (ss#). Several ss numbers referring to the same SNV are assembled into a reference SNP cluster labelled with an rs number. Consequently, the number of ss entries for each SNV (rs entry) correlates positively with its genomic reliability and is used as a parameter of confidence. If the Mouse Phenome Database (MPD) contains information on the distribution of SNV alleles in mouse strains, the respective rs number is printed in blue and the MPD data is reported in Table 3-6.

Although we rejected SNVs with obvious uncertainties in dbSNP gene assignment from our analyses (see Supplementary Methods), all predicted deviations from the standard amino acid sequence should be considered with particular caution if there is only one ss entry for the SNV. The genes containing the SNV according to the dbSNP web interface[255] ("Integrated Maps" field of the rs entry) are listed in the second column; they are printed in grey if they were not attributed to the SNV in both the early spring 2012 and the October 2012 dbSNP data queries. Because the rate of false positive reports at the nucleotide level (SNVs in dbSNP) appears to be much higher than the rate of false positive peptide identifications (we applied an FDR of < 1% and only report peptides with a Mascot score > 25), we did not check the MS spectra of peptides in this table manually. However, due to the restrictive MS data filtering, the number of incorrectly identified peptides (irrespective of predicted SAVs) in this table is anticipated to be zero or very low. SAV positions in urinary peptides are underlined and the SAV form identified by our MS approach is highlighted in green.

Annotations

* This gene or locus is not known to be transcribed and translated and therefore the contained SNVs are not expected to affect the urinary peptidome.

(a) The SAV form occurring in C57BL/6J mice corresponds to the NCBI reference sequence and is always indicated in front of the SAV position number. In contrast, the non-reference ("non-B6") SAV form is denoted behind the SAV position number.

(b) For all peptides of this table, the "B6 form" of the SAV is encoded by more than one gene. In several cases, which are indicated by "yes" in this column, the "non-B6 form" is also encoded by more than one gene. However, some SAVs result in unique peptides in their "non-B6" form, thereby altering urinary peptide composition qualitatively if the affected gene indeed contributes to the urinary peptide concentration.

(c) According to NCBI Reference Sequence, only the form with L is encoded in the standard (C57BL/6J) proteome of *Mus musculus*. However, EMBL-ENA (European Nucleotide Archive) entry BAC37484.1 reports the form with M for C57BL/6J mice. The form with M is also contained in UniProtKB/TrEMBLentry Q8C5B4 and IPI entry IPI00756078.1.

(d) The major subspecies of *Mus musculus* do not differ at this position[300]. However, it should be noted that plenty of SAVs occur within *Mus musculus* haemoglobin subunits β1 and β2[300]. Although we could identify only one further *Hbb-t* derived peptide (VDPENFRLLGNM, Mascot score = 21) apart from VVYPWTQRYFDS (Mascot score = 43) at an FDR of 1%, other *Hbb-t* derived peptides encompassing the SAV positions reported in ref. [300] can be expected in urine.

(e) In *Mus pahari*, a sympatric species of *Mus musculus* in Southeast Asia, the SAV S44N occurs in both Hbb-t1 and Hbb-t2. However, the major subspecies of *Mus musculus* do not differ at this position[300].

(ENA) According to data accessible at the European Nucleotide Archive (ENA; http://www.ebi.ac.uk/ena/). These data are linked to the UniProtKB entry of a protein.

(MPD) According to the data from the Mouse Phenome Database (MPD)[256,257], which are reported in Table 3-6.

(ss) According to a submitted SNV (ss) entry of the respective reference SNV (rs) entry in the dbSNP.

Table 3-6 | Distribution of SNV alleles affecting identified urinary peptides in laboratory and wild-derived inbred strains of *Mus musculus* (*Nat Comm*, Theo).

				Nucleotides at the SNV position in *Mus musculus* laboratory and wild-derived inbred strains (for the latter the *subspecies* is also indicated)																														
				Laboratory strains																										*domesticus*	*musculus*		*mollossinus*	*castaneus*	*Mus spretus*
														C57																				
														C57BL/10			C57BL/6																	
Gene	Accession number of SNV in dbSNP	SAV resulting from SNV[(a)]	MPD data set	129P2/OlaHsd	129S1/SvImJ	129S5SvEv<Brd>	AKR/J	A/J	BALB/cByJ	**BALB/cJ**	BTBR_T+_tf/J	C3H/HeJ	C57BL/10J	C57BL/10ScNJ	C57BL/10ScSnJ	**C57BL/6J**	C57BL/6NCrl	C57BL/6NJ	C57BL/6NTac	C57BLKS/J	C57BR/cdJ	C57L/J	C58/J	CBA/J	DBA/2J	FVB/NJ	LP/J	NOD/ShiLtJ	NZO/HlLtJ	WSB/EiJ	PWD/PhJ	PWK/PhJ	MOLF/EiJ	CAST/EiJ	SPRET/EiJ
Abca13	rs50794155	S747F	[1]	C	C	C	C	C		C		C				C		C						C	C		C	C	C	C		T		C	C
Alb	rs31706341	A564S	[1,5]	G	G	G	G	G	G	G	G	G				G		G						G	G	G	G	G	G	G	T	T	G	T	G
Ces1c	rs51686553	F461S	[1,3]	G	G	G	G	G	G	G		G	A	A	A	A	A	A	A	A	A	A		G	G	G	G	G	G	G		A		A	G
Egf	rs30619049	V103L	[1,5]	C	C	C	C	C	C	C	C	C				C		C						C	C	C	C	C	C	G	G	G	C	G	G
Kap	rs31887216	Q44R	[1,5]	T	T	T	T	T	T	T	T	T				T		T						T	T	T	T	T	T	T	C	C	T	T	T
	rs31887214	A53S	[1,3,5]	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C				C	C	C	C	C	C	A	C	C	C	C	A	C
Lrg1	rs13467361	S100F	[1]	G	G	G	G	G		G		G				G		G						G	G		G	G	G	G		A		G	G
	rs13467359	A112V	[1,4,18]	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	A	A	A	A	A
Prol1	rs51790902	G52S	[1,3,5]	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	A	G	G	G	G	A	G
Serpina1b	rs33061356	D313N	[5]				C	T			C	T				C									C	C		C			T				
Serpina3k	rs33123758	V386I	[1,3,5]	A	A	A	A	G	A	A	A	G	G	G	G	G	G	G	G	A	A	A	G	A	A	A	A	A	G	A	G	G	G	G	A
	rs8273122	I398F	[1,3,5]	T	T	T	T	A	T	T	T	A	A	A	A	A	A	A	A	T	T	T	A	T	T	T	T	T	A	T	A	A	A	A	T
Ces1g	rs50445108	W514C	[5]		C		C	C	C		A	C				C									C	C		C		C	C		C	C	
Serpina1c, *d*, *e* or noncoding locus	rs31018094	I52L	[5,6]		T		T	G	T		T	G				G									T	T		T		T	G		T	T	
Serpina1b	rs33060453	K317E	[5]		T		T	C	T			C				C									T	T		T		T	T		C	C	
Serpina1c	rs51746614	V382L	[1,3]	G	G	G	G	C	G	G	G	C	C	C	C	C	C	C	C	G	G	G	C	G	G	G	G	G	C	G				C	C

Note that the subspecies of *Mus musculus* can be assigned unequivocally using the diverse predicted SAV peptide pairs of identified urinary peptides. The identified urinary peptides affected by the SAVs are depicted in Tables 3-4 and 3-5. SNV forms of further mouse strains are available at MPD[257]. Nucleotides are colour-coded to support visual comparison.

[(a)] The SAV form occurring in C57BL/6J mice is always indicated in front of the SAV position number. In contrast, the non-reference ("non-B6") SAV form is denoted behind the SAV position number.

Mouse Phenome Database (MPD)[256,257] data sets: (queried on 11 October 2012)

[1] Wellcome Trust Sanger Institute. Sanger SNP data, 65+ million locations, 18 strains of mice. MPD:Sanger1.[80]
[3] Center for Genome Dynamics (CGD). Imputed SNP data for 88 strains (plus Sanger data, 12 strains), for 12+ million locations. MPD:CGD-IMP2. [302]
[4] Center for Genome Dynamics (CGD). SNP data from Mouse Diversity Genotyping Array, 550,000 locations for 123 strains of mice. MPD:CGD-MDA1.[303]
[5] Cox D, Frazer KA. SNP data, 8.2+ million locations for 16 inbred strains of mice. MPD:Perlegen2.[304]
[6] Celera / Applied Biosystems. SNP data, 2.1+ million locations for 5 inbred strains of mice. MPD:Celera2.[305]
[18] The Jackson Laboratory. Ad hoc strains, Mouse Diversity Genotyping Array. MPD:JAX-MDA1.

Table 3-7 | Identified urinary peptides that are encoded in the genome of BALB/c but not B6 mice or vice versa (*Nat Comm*, Theo)**.**

Gene	Peptide sequence	Peptide encoded by allele of strain		Peptide detected in urine of indicated mouse strain in MS experiment no. (1-3)				
		B6J	BALB/cJ	B6	B6/$\beta_2m^{-/-}$	B10	B10.D2	BALB/c
Ces1c	SVFGAPLLKEGASEEETNL	-	+	-	-	n.a.	n.a.	1
Serpina3k	DVAETGTEAAAATGVIGGIRKAVL	+	-	3	n.a.	n.a.	n.a.	n.a.
	DVAETGTEAAAATGVIGGIRKAIL	-	+	-	-	-	-	1, 2
	VAETGTEAAAATGVIGGIRKAI	-	+	-	-	n.a.	n.a.	1
	VAETGTEAAAATGVIGGIRKAIL	-	+	-	-	-	-	1, 2
	VAETGTEAAAATGVIGGIRKAILPAVHFNRP	-	+	-	-	n.a.	n.a.	1
	AAATGVIGGIRKAIL	-	+	-	-	n.a.	n.a.	1
	AAATGVIGGIRKAILPAVHFNRP	-	+	-	-	n.a.	n.a.	1
	AATGVIGGIRKAIL	-	+	-	-	n.a.	n.a.	1
	AATGVIGGIRKAILPAVHFNRP	-	+	-	-	n.a.	n.a.	1
	ATGVIGGIRKAIL	-	+	-	-	n.a.	n.a.	1
	GVIGGIRKAIL	-	+	-	-	n.a.	n.a.	1
	GIRKAILPAVHFNRP	-	+	-	-	n.a.	n.a.	1
	IRKAILPAVHFNRP	-	+	-	-	n.a.	n.a.	1
	IRKAILPAVHFNRPF	-	+	-	-	n.a.	n.a.	1
	IRKAILPAVHFNRPFL	-	+	-	-	n.a.	n.a.	1
	KAILPAVHFNRPF	-	+	-	-	-	-	1, 2
	KAVLPAVCFNRPFL	+	-	-	1	n.a.	n.a.	-
	KAILPAVHFNRPFL	-	+	-	-	n.a.	n.a.	1
	PAVCFNRP	+	-	3	n.a.	n.a.	n.a.	n.a.
	PAVHFNRPF	-	+	-	-	n.a.	n.a.	1
	PAVCFNRPFL	+	-	1	1, 2	2	2	-
	PAVHFNRPFL	-	+	-	-	-	-	1, 2
Serpina3k	IVIYHTSAQSIL	+	-	1, 2	1, 2	2	2	-
	FVIYHTSAQSIL	-	+	-	-	n.a.	n.a.	1

All cysteine residues occurred in a cysteinylated form, i.e. they had bound free cysteine via a disulfide bridge. Because the cysteine residues are not encoded in BALB/c mice, this posttranslational modification differs between mouse strains. See Table 2-1 for details about the setups of MS experiments (no. 1 − 3). Blue print refers to specific features of B6 mice whereas red print refers to peculiarities of BALB/c mice. "n.a." indicates that the respective mouse strain was not analysed in the experiments that led to the identification of the peptide.

The dbSNP database (XML files, revision 14 April 2010) contains 74.696 missense SNVs for *Mus musculus*. By definition, these SNVs are predicted to result in SAVs in the encoded proteins. Consequently, for each gene of *Mus musculus*, on average about three different SAV-coding SNVs are described. The frequency of observed SAV peptides in urine is in agreement with the value obtained when calculating the expected frequency of SAV peptides (based on the dbSNP missense SNV data) in a random mouse peptide set having the same average peptide length.

The biochemical differences of SAV peptide pairs are sometimes very minor, with leucine versus isoleucine SAVs being the most conservative exchange. The SAV peptides IVIYHTSAQSIL and

FVIYHTSAQSIL highlighted above both contain a hydrophobic amino acid at the N-terminal SAV position. Consequently, we aimed to investigate whether such small differences can be discriminated by VSNs, a prerequisite for qualifying SAV peptides as olfactory signals of genetic identity. IVIYHTSAQSIL and FVIYHTSAQSIL are not expected to bind to MHC class I molecules in their full 12-mer length, and their SAV position is unrelated to MHC anchor positions. After I had enriched synthetic IVIYHTSAQSIL and FVIYHTSAQSIL to > 96 % purity, Trese Leinders-Zufall could indeed demonstrate VSNs that are specific for either one of these peptides or react with both when applied at concentrations of 10^{-11} M (Fig. 4 of Sturm *et al.*[207]). Hence, naturally occurring 12-mer SAV peptides are recognised by VSNs, and an MHC motif is neither required for activation of these cells nor for their capability to discriminate between these ligands. Given these observations, it seems likely that VSNs can also discriminate other types of genomic variation which, although less common than SNVs, are expected to alter biochemical properties of individual peptides more profoundly, e.g. strain specific expression of complete peptides, multiple amino acid variations originating from several SNVs in adjacent codons as well as insertions or deletions of amino acids resulting from genomic insertions and deletions respectively. Such genomic variations apart from SAVs are dealt with in the next section and in section "4.3. Genotype specific peptides offer new perspectives for olfactory research" (*Nat Comm*, joint).

3.3. Further genomic variations influence urinary peptides

Amino acid variations in urinary peptides are expected to increase with growing taxonomic divergence. To demonstrate this, we aligned the 115 major urinary protein (MUP)-derived peptides which we identified in urine to the MUP sequence of *Mus macedonicus*, a species able to produce fertile F_1 females with the sympatric *Mus musculus*[306]. In *Mus macedonicus*, urine contains only a single MUP, whereas urine of *Mus musculus* contains multiple different MUPs encoded by several paralogous genes[307]. Accordingly, out of the 115 MUP-derived peptides identified in urine, 58 are not encoded in the *Mus macedonicus* MUP including several peptides predicted to differ in more than three amino acid residues between these two species (Figure 3-2) (*Nat Comm*, joint).

Even within *Mus musculus*, MUPs are a highly polymorphic group of proteins[183] and quantitative differences in urinary concentrations can occur in addition to qualitative ones. For example, the proportion of MUP20 (cf. section section "1.4. Olfactory cues in *Mus musculus*") among all MUPs is only < 0.5% in male BALB/c but about 15% in male B6 mice[8]. Consequently, we found unique peptides from MUP20 more frequently and more abundantly in male urine of B6 than BALB/c origin (Table 3-8 and Supplementary Table 9-6). Females do not express MUP20[8,308], and thereby, unique MUP20-derived urinary peptides are both sex specific and vary in concentration in male urine (*Nat Comm*, joint).

Figure 3-2 | Alignment of identified urinary peptides derived from major urinary proteins (MUPs) of B6, B10 and / or BALB/c mice (*Mus musculus*) to the single MUP sequence of *Mus macedonicus*. Please refer to the following page for further explanations (*Nat Comm*, Theo).

Figure 3-2 (previous page) | Alignment of identified urinary peptides derived from major urinary proteins (MUPs) of B6, B10 and / or BALB/c mice (*Mus musculus*) to the single MUP sequence of *Mus macedonicus*. Note that identified MUP peptides are derived from multiple different MUPs encoded by several paralogous genes in *Mus musculus*. In contrast, urine of *Mus macedonicus* appears to contain only a single MUP whose sequence is depicted at the top with grey shading[307]. The 55 identified MUP peptides encoded by both species are indicated in black whereas those 58 not contained in the *Mus macedonicus* MUP sequence are printed in green with the differing amino acids highlighted in red. The two amino acids of the *Mus macedonicus* MUP not encoded in any *Mus musculus* MUP[108,179] are marked in violet. For the two peptides printed with grey letters the exact sequence remains unclear because both HGIIRENIIDL and HGILRENIIDL are encoded in *Mus musculus*, and isoleucine and leucine cannot be discriminated by the low energy CID MS applied. The MS^2 spectra of most of the indicated MUP peptides have not been controlled manually. However, because of the many overlapping sequences, the applied false discovery rate (FDR) of 1% in target-decoy database searches and the removal of all peptides with Mascot scores below 25, the number of incorrectly identified MUP peptides is anticipated to be very low leaving the percentage of species discriminating MUP peptides among all MUP peptides essentially unchanged. Different urinary MUP expression patterns also occur within wild populations of *Mus musculus*[183,184] and between B6 and BALB/c mice[182]. Due to incomplete strain-specific protein sequence information in protein databases and literature and a high sequence homology between different MUPs, we were not able to reliably assign urinary MUP peptides encoded in B6 but not BALB/c mice or vice versa. Nevertheless, such peptides are likely to exist. Quantitative differences of urinary MUP20 peptides discriminating between B6 and BALB/c mice are covered in Table 3-8 (*Nat Comm*, Theo).

The impact of the metalloendopeptidase meprin A α on the urinary peptidome

Genes with allele-specific occurrence or functionality can affect MUPs[308] as well as other proteins. Correspondingly, we identified 21 peptides derived from meprin A α (that are 3.3% of all identified peptides; cf. Supplementary Table 9-7), a urinary metalloendopeptidase lacking mRNA expression in C3H/He mice (Fig. 5 in ref.[309]; see also section "1.1.1. The formation of urine"). Urine of some other inbred strains also lacks meprin A α and the nonfunctional *mep1a* allele is genetically linked to the $H2^k$ haplotype[36,310]. Mice deficient in this protein will not only lack the peptides encoded by *mep1a* but could also have a changed proteolytic cleavage pattern of urinary peptides (*Nat Comm*, Theo).

To estimate the possible impact of the lack of proteolytic cleavage by meprin A α, I checked whether I can observe urinary peptide patterns that point to the activity of meprin A α. This was done applying the meprin A α positive mouse strains investigated during my thesis (see Table 2-1). Meprin has been demonstrated to present "the major endopeptidase in mouse urine"[37]. The soluble form of meprin in urine is composed of homo-oligomeric meprin A α[38] which is prone to cleave N-terminal to the sequence XP, where X represents any amino acid and P symbolises proline[39,40]. Therefore, a dominant role of homo-oligomeric meprin A α in the generation of urinary peptides should be reflected by an overrepresentation of urinary peptides with proline in position 2. However, our data show the opposite, with proline frequencies significantly diminished in position 2 as compared to positions 3 to 10 (Table 3-9). This discrepancy could be explained in two ways.

Table 3-8 | Unique MUP20 derived peptides are more abundant in urine of B6 and B6/$\beta_2m^{-/-}$ mice as compared to BALB/c mice (*Nat Comm*, Theo).

		Normalised MS intensities, in 1000									Mean of normalised MS intensities, in 1000		
Peptide	Exp. no.	B6			B6/$\beta_2m^{-/-}$			BALB/c			B6 and B6/$\beta_2m^{-/-}$	BALB/c	Ratio (B6 and B6/$\beta_2m^{-/-}$) / BALB/c
EEASSMERNF	1	548	496	n.a.	117	474	460	0	12	0	419	4	102
NVEKINGEWYTIM	1	90	144	n.a.	114	222	249	0	0	0	164	0	∞
VEYIHVLENSL	1	250	395	n.a.	184	0	266	0	49	0	219	16	13
"	2	52	0	0	90	50	0	0	0	0	32	0	∞
"	3	45932	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
IHVLENSLALKF	1	76	68	n.a.	72	98	0	0	0	0	63	0	∞
IMIHLINKKDGETFQL	1	1240	2768	n.a.	895	2404	900	544	278	403	1641	409	4
LINKKDGETFQL	1	0	0	n.a.	26	56	0	0	0	0	16	0	∞

Fractions of urine mixtures of 4-5 mice per mouse strain were analysed by MS (cf. Table 2-1). In exp. no. 1 and 2 MS intensities were normalised so that the mean of the sum of intensities of all identified peptides was equal for all mouse strains. MS intensities correlate only very loosely with peptide concentrations and should therefore be interpreted more qualitatively than quantitatively. Only peptides that are unique for MUP20, i.e. those that cannot be explained by other proteins, are depicted. The $\beta_2m^{-/-}$ mutation is not expected to influence the expression of MUP20. Accordingly, there was no significant difference in the intensities of unique MUP20 peptides between B6 and B6/$\beta_2m^{-/-}$ mouse urine ($P = 0.60$, two-tailed, heteroskedastic Student`s *t*-test). Therefore, B6 and B6/$\beta_2m^{-/-}$ mice were considered as a single entity when calculating the mean MUP20 peptide intensities and performing the statistical comparison with BALB/c mice. Unique MUP20 peptides are significantly more abundant in urine of B6 and B6/$\beta_2m^{-/-}$ mice as compared to BALB/c mice ($P = 0.019$, two-tailed, heteroskedastic Student`s *t*-test with Bonferroni correction for two comparisons). The statistical testing was performed solely with the data of experiment no. 1 because, first, MS intensities cannot be compared between different experiments, second, we could detect only a single peptide unique for MUP20 in experiments no. 2 and 3 and, third, it is thereby avoided that one peptide (VEYIHVLENSL) entails a stronger influence on the comparison than the other peptides. Overall, in contrast to MUP20 derived peptides, more peptides were detected in mouse urine of BALB/c (407 and 181) than B6 (348 and 134) or B6/$\beta_2m^{-/-}$ (373 and 185) origin (numbers for experiment no. 1 and 2 respectively). MS reliability parameters for peptides in this table are specified in Supplementary Table 9-6. MS intensities are colour-coded according to the rank that each intensity value achieved among all measurements of the respective peptide in a given experiment for the indicated mouse strains. n.a. = not analysed

(1) If one assumes that the described cleavage pattern of homo-oligomeric meprin A α[39,40] reflects the natural situation in urine, this peptidase seems to have no dominant impact on the urinary peptide pattern. This would imply that pre-renal and / or renal proteases of the mouse and / or *ex vivo* microbial proteases are more important in generating urinary peptides than mouse proteases in the urine (urinary proteases apart from meprin A α contribute only about 50% to protein degradation within urine[37]).

(2) An alternative, although less likely explanation could be that the cleavage pattern of homo-oligomeric meprin A α determined in artificial buffers *in vitro*[39,40] does not apply to the natural conditions present in urine thereby questioning the statistical data analyses of Table 3-9.

Table 3-9 | Evidence against either a dominant contribution of homo-oligomeric meprin A α to the observed urinary peptide pattern or against its described cleavage pattern.

	Percentage of proline residues within the peptides in position											*P* value for
Exp.	1	2	3	4	5	6	7	8	9	10	3 - 10 (mean)*	position 2*
no. 1	1.9	1.3	5.1	7.0	8.1	8.3	7.1	7.3	6.7	6.8	7.0	$1.8 \cdot 10^{-9}$
no. 2	5.2	2.2	4.5	10.5	12.4	6.5	6.9	5.8	7.0	8.7	7.8	0.0165

Note that mouse homo-oligomeric meprin A α has a preference to cleave N-terminal of the sequence XP, where X represents any amino acid and P symbolises proline[39,40]. Therefore, a dominant role of homo-oligomeric meprin A α in the generation of urinary peptides should lead to increased proline frequencies in position 2. However, our data show the opposite, with proline frequencies significantly reduced in position 2. *P* values were calculated for the observed proline frequencies in position 2 taking positions 3 to 10 as the reference for calculating the Gaussian distribution of proline frequencies expected by chance. The asterisk indicates that N-terminal (i.e. position 1) and C-terminal amino acid residues of peptides were not included in the calculations because these positions are known to prominently influence the cleavage pattern of many different peptidases[60]. Consequently, only peptides longer than *n* amino acids were considered for position *n*, e.g. for position 8 only peptides longer than eight amino acids were taken into account. In exp. no. 3, the proline frequency was also lower in position 2 as compared to each of positions 3 to 10. However, due to the low total number of only 75 peptides in exp. no. 3 (cf. Table 2-1), this experiment appears not appropriate for a pertinent statistical evaluation.

Although the impact of homo-oligomeric meprin A α on urinary peptide patterns might not be dominant, it is unlikely that this abundant, soluble metalloendopeptidase has no impact at all on the cleavage of urinary peptides. For example, the relative lack of proline in position 2 might even be more pronounced in the absence of meprin A α. Furthermore, homo-oligomeric meprin A α has a relatively broad cleavage specificity[40]. Additionally, meprin A α has important proteolytic functions upstream of the urinary tract being involved in the regulatory proteolytic processing of several cytokines[311,312] and the remodelling of the extracellular matrix[38]. These upstream cleavage events could also contribute to urinary peptide pattern. Therefore, the effect of meprin A α deficiency should be considered when comparing mice of the $H2^k$ haplotype with MHC congenic individuals in potentially MHC related behaviours (see Figure 4-1 and section "4.2.2. Non-MHC genes may confound behavioural experiments targeting the MHC" for a discussion).

More genotype-dependent differences in urinary peptides are yet to be discovered

„*Wichtig ist, dass man nicht aufhört zu fragen.*“
(“It is important to never stop asking.”)
Albert Einstein

In the preceding sections, I have highlighted several examples of urinary peptides that differ between BALB/c mice and mice of the C57 lineage qualitatively (SAV peptides) or quantitatively (unique MUP20 peptides). Although I could not pinpoint further conclusive examples of differences in urinary peptides between these mouse strains, they likely do exist due to the separated generation of the BALB/c and the C57 lineage of inbred mice at the beginning of the 20th century[148,195]. To test this hypothesis, I calculated the normalised Euclidean distance of normalised MS peptide intensities between all examined mouse strains in a pair wise manner. The Euclidean distance is a measure of the sum of all observed differences between two given samples. By using the *normalised* Euclidean distance, the observed differences for every single peptide were normalised according the overall standard deviation of MS intensities of this peptide across all examined mouse strains. This means, for example, if an MS peptide intensity differs by 100,000 between two given mouse strains, this is considered a strong difference if all mouse strains are within this range, but it is judged as a small difference if MS intensities of this peptide differ by 10,000,000 between some mouse strains. Thereby, it is respected that very intense peptides tend to have higher absolute differences between mouse strains than low intensive ones. I calculated the normalised Euclidean distance according to the following formula[313]:

$$Normalised\ Euclidean\ distance = \sqrt{\sum_{n=1}^{m}\left(\frac{(Int1_n - Int2_n)}{s_n}\right)^2}$$

m	number of considered peptides in the respective experiment
n	indicator for a single peptide
*Int*1	normalised MS intensity of the peptide *n* in mouse strain 1
*Int*2	normalised MS intensity of the peptide *n* in mouse strain 2
s	standard deviation of the peptide *n* over all mouse strains

The reason for calculating the Euclidean distances was to test if genomic variations *apart* from SAVs and unique MUP20 peptides affect urinary peptide composition in BALB/c mice versus mice of the C57 lineage. Therefore, all SAV peptides with known differences between BALB/c mice and the four mouse strains of the C57 lineage (cf. Tables 3-6 and 3-7) as well as all unique MUP20 peptides (see Table 3-8) were excluded from these analyses. Despite of this exclusion, differences (as measured by the normalised Euclidean distance) of urinary peptides were significantly lower within different mouse strains of the C57 lineage than between BALB/c mice and the four individual C57 mouse strains

($P = 0.002$, two-tailed, heteroskedastic Student`s *t*-test using data of exp. no. 2, see Table 3-10). This adds further evidence to the assumption that differences in urinary peptides increase with decreasing evolutionary relatedness. These differences might be of quantitative and / or qualitative nature as the Euclidean distance measures both types in a single parameter. Further detailed investigation and correlation of gene expression data with urinary peptides would be necessary to unravel additional conclusive examples.

Table 3-10 | Additional, unknown genetic differences between mice of the C57 lineage and BALB/c mice influence urinary peptide composition.

			Normalised Euclidean distance of MS peptide intensities*	
Compared mouse strains			Exp. no. 1	Exp. no. 2
BALB/c	versus	B10	n.a.	20.2
BALB/c	versus	B10.D2	n.a.	20.4
BALB/c	versus	B6	28.8	23.7
BALB/c	versus	B6/β_2m$^{-/-}$	27.9	20.9
Mean of comparisons involving BALB/c			28.3	21.3
B10	versus	B10.D2	n.a.	11.5
B10	versus	B6	n.a.	16.9
B10	versus	B6/β_2m$^{-/-}$	n.a.	13.7
B10.D2	versus	B6	n.a.	19.6
B10.D2	versus	B6/β_2m$^{-/-}$	n.a.	14.5
B6	versus	B6/β_2m$^{-/-}$	19.4	15.2
Mean of comparisons not involving BALB/c, i.e. comparisons within the C57 lineage			19.4	15.2

Note that the normalised Euclidean distance is only comparable within one experiment but not between different experiments due to different overall numbers of peptides and different overall MS intensities in each experiment. Within each experiment, MS intensities were normalised so that the mean of the sum of intensities of all identified peptides was equal for all mouse strains The asterisk denotes that all SAV peptides with known differences between BALB/c mice and the four mouse strains of the C57 lineage (cf. Tables 3-6 and 3-7) as well as all unique MUP20 peptides (see Table 3-8) were excluded from the calculation of Euclidean distances. n.a. indicates that the Euclidean distance cannot be calculated because at least one of the respective mouse strains was not analysed in exp. no. 1.

As expected, the lowest Euclidean distance was observed between B10 and B10.D2 mice. However, Euclidean distances were generally quite similar within the C57 lineage and no specific C57 mouse strain separated from the other ones significantly. Furthermore, most Euclidean distances did not follow the order expected from evolutionary relatedness within C57 mice. This might be due to the differing number of identified peptides in each strain, detailed in the following for exp. no. 2: B10 = *211*, B10.D2 = *225*, B6 = *127*, B6/β_2m$^{-/-}$ = *174*, BALB/c = *172*, not counting SAV peptides differing between B6 and BALB/c and unique MUP20 peptides. The lowest number of peptides was identified in B6 mice and this might account for the slightly elevated (although not statistically significant) Euclidean distance in comparisons involving B6 (see Table 3-10). Biological replicates

covering a larger percentage of the urinary peptidome could possibly lead to Euclidean distances correlating with the expectations also within the C57 lineage. Such replicates would definitely increase the statistical power of the comparisons and would enable statistical evaluation of comparisons between two pairs of strains (e.g. is the distance between B10 and B10.D2 mice lower than the one between B10 and B6 mice?).

3.4. Detection of an MHC-dependent peptide in mouse urine

We were not able to detect MHC-dependent peptides, i.e. those that are only present in urine of mice expressing the corresponding MHC class I protein complex, by our MS approach (see section "3.1. Abundant peptides with MHC class I binding motif occur in mouse urine in an MHC-independent manner"). Assuming that *bona fide* MHC-dependent peptides were present at very low concentrations, we reasoned that they would be missed in MS analyses due to the presence of a multitude of other, more abundant peptides. In fact, control experiments showed that a synthetic MHC peptide spiked into urine is not reproducibly detected at concentrations below 10^{-11} M.

Therefore, we developed a highly specific model system based on the well-studied H2-K^b-presented peptide SIINFEKL corresponding to amino acid residues 257-264 of chicken ovalbumin (OVA)[314,315]. The transgenic mouse line B6/OVA^+ expresses a cell surface protein composed of the entire chicken ovalbumin sequence and a transmembrane sequence under the control of the chicken β-actin promoter in all organs. Ovalbumin is naturally processed in this strain resulting in SIINFEKL-H2-K^b complexes[202]. B6/OVA^+ mice were crossed with B6/$\beta_2m^{-/-}$ mice[201], resulting in the strain B6/$OVA^+/\beta_2m^{-/-}$, which produces ovalbumin but lacks the expression of functional MHC class I molecules (for a phenotypic characterisation of lymphocytes isolated from these mouse strains see Figure 2-1). As a sensitive reagent for detecting SIINFEKL, we used the monoclonal antibody 25-D1.16, which is specific for the C-terminal residues of this peptide in association with H2-K^b [218]. We developed an enzyme-linked immunosorbent assay (ELISA) based on recombinant H2-K^b-molecules refolded *in vitro* in the presence of the human β_2m-subunit and the peptide TEYGFLNL. This assay can be used with[293] or without removing the bound peptide at alkaline pH before loading of SIINFEKL, the latter mode being about three times more sensitive[203, (*Nat Comm*, joint)].

3.4.1. Challenging the results of my diploma thesis

The question of an MHC-dependent occurrence of SIINFEKL in B6/OVA^+ mouse urine had already been examined in my diploma thesis. We ended up with the conclusion that SIINFEKL is present in this body fluid of B6/OVA^+ mice at a mean concentration of about $2 \cdot 10^{-11}$ M in an MHC-dependent

manner. However, during my PhD thesis, it turned out that this initial observation just resulted from a mouse with severely impaired, non-physiological composition of urine and did therefore not reflect the SIINFEKL concentrations at physiological levels. We were not able to recognise the problem at that time due to a suboptimal experimental design. In the following, it is detailed how I realised and finally solved the problem leading to the conclusion that SIINFEKL is indeed present in B6/OVA^{+} mouse urine in an MHC-dependent manner but at a median concentration of only $4 \cdot 10^{-12}$ M.

In the experiments of my diploma thesis, we always pooled urine samples of about 12 individual mice per strain for measurements of urinary SIINFEKL concentrations. Urine mixtures from different mice instead of urine samples of individual mice were applied for two reasons. First and most importantly, previous experience of our laboratory has shown that composition of urine differs substantively even between individuals of the same inbreed strain[62]. Pooling of urine within each strain did balance the individual differences thereby focusing analyses on common features of that strain. This greatly simplified the inter-strain comparisons, helped to reduce the number of samples (urine of several mice is measured at once and not separately) and therewith decreased the work input as well as measurement times. The second advantage of urine pooling is also time-related. Due to the very low urinary concentration of SIINFEKL, 5 to 10 ml of urine are required for detection even with our highly sensitive SIINFEKL-H2-K^{b}-ELISA. Using about 12 mice per sample, this volume can be gathered within a few days. However, if this amount needs to be derived from a single mouse, collection times for each sample often exceed one month.

Pooling of urine within a strain works perfect if the substance of interest does not differ significantly between the mouse strains compared. If one finds no difference between the mixtures, a significant strain-related difference cannot be expected when analysing the individual mice instead. However, if a difference between the mixtures is detected, one cannot be sure if this variation really reflects all mice of the considered groups or just results from a few or even a single outlier mouse. A general prerequisite for meaningful statistical testing is that the individual data points are independent of one another. In the case of inter-strain comparisons of urine this means that each individual mouse should contribute to only one data point of the comparison. During my diploma thesis we only ensured a (partial) independence in collection time meaning that urine obtained at one day usually contributed to only one measured urine mixture. However, this degree of independence of data points turned out to be insufficient for the analysis of SIINFEKL (see below). If mixtures of urine from several mice are examined, the rule of independence of data points would require that a completely non-overlapping set of urine donor mice is used for each measurement − a prerequisite that is unjustifiable in terms of animal protection, breeding efforts and costs when using more than 10 mice for each mixture. Therefore, during my diploma thesis, we repeatedly measured urine mixtures derived from the same mice neglecting the rule of independence.

At the beginning of my PhD thesis (without foreseeing the tremendous consequences, but having in mind the rule of independence), I decided to examine the variation of urinary SIINFEKL concentrations between six individual B6/OVA$^+$ mice. I collected 10 ml of urine of each mouse with the help of metabolic cages and subjected 20 s HPLC fractions thereof to the SIINFEKL-H2-K^b-ELISA in the 50 µl standard format. The amounts and procedures equalled those used in my diploma thesis but SIINFEKL was not detectable in three of the six B6/OVA$^+$ mice. In the other three B6/OVA$^+$ mice, SIINFEKL was only borderline detectable at a concentration of about 10^{-12} M being as little as factor 2 to 4 above the corresponding signal of B6 mice (Table 3-11, experiment (exp.) no. 15). This finding was in obvious conflict with the results of the previous exp. no. 14 (see Table 3-11) and the SIINFEKL experiments of my diploma thesis (cf. Table 3-5 in [203]) where concentrations of SIINFEKL in B6/OVA$^+$ mouse urine were between 0.8 and $3.7 \cdot 10^{-11}$ M. However, for the positive control consisting of synthetic SIINFEKL spiked into B6 mouse urine, the measured SIINFEKL concentrations in exp. no. 15 resembled those obtained during my diploma thesis. Therefore, technical problems appeared unlikely to be responsible for the discrepancy regarding B6/OVA$^+$ mice. Nevertheless, it was necessary to reproduce the results of exp. no. 15 to have a solid basis for questioning the multitude of previous experiments. Consequently, I designed exp. no. 16 taking into account the following points:

1.) The conflict between exp. no. 14 and exp. no. 15 could result from one or some of the six B6/OVA$^+$ mice whose urine was measured in exp. no. 14 but not in exp. no. 15. I therefore included all B6/OVA$^+$ mice of exp. no. 14 into a dedicated urine mixture of exp. no. 16 for which we could obtain sufficient amounts of urine within the available time (cf. Table 3-11).

2.) When urine was collected from the six B6/OVA$^+$ mice for exp. no. 15 (Table 3-11), these mice were about three months older than when their urine had been collected for exp. no. 14. In order to rule out that aging was the cause of the conflict, I additionally included urine of B6/OVA$^+$ mice being at least 9 month younger (not depicted in Table 3-11) than the other B6/OVA$^+$ mice of exp. no. 16 (indicated in Table 3-11). For the younger mice, I measured urine mixtures from three non-overlapping groups each consisting of 4 − 5 mice. Like in all experiments preceding exp. no. 16, urine was collected in metabolic cages.

3.) The urine collection method was not changed between exp. no. 14 and exp. no. 15 and can therefore be ruled out as the cause of the discrepancy. However, because the urinary concentration of SIINFEKL was around the detection limit in B6/OVA$^+$ mice of exp. no. 15, improvements in sensitivity were required for future experiments to accurately examine the MHC-dependence of SIINFEKL in the picomolar concentration range. It appeared possible that some SIINFEKL is degraded over night during collection in metabolic cages despite of the presence of protease inhibitors in the collection tubes of the metabolic cages. Therefore, I decided to compare urine

collected over night in metabolic cages versus manually collected urine immediately frozen on dry ice after voiding from the bladder using the same set of 10 B6/OVA$^+$ mice (cf. Table 3-11).

Table 3-11 | The search for "outlier mice" and tentative testing of the effect of urine collection on measured urinary SIINFEKL concentration.

Exp. no.	Urine collection	B6/OVA$^+$ mice examined (ear no.)												Urinary concentration of SIINFEKL
14	Metabolic cages	52	53	54	55	58	59	60	67	87	88	89	90	$2.5 \cdot 10^{-11}$ M (mixture of all 12 mice)
15	Metabolic cages							60	67	87	88	89	90	Ear no. 60: $1.0 \cdot 10^{-12}$ M Ear no. 87: $1.5 \cdot 10^{-12}$ M Ear no. 89: $1.2 \cdot 10^{-12}$ M Ear no. 67, 88 and 90: below LOD LOD = $7.6 \cdot 10^{-13}$ M
16	Metabolic cages		53		55	58	59	60	67	87	88	89	90	Slightly below LOD of $6.0 \cdot 10^{-13}$ M (mixture of all 10 mice)
	Manually		53		55	58	59	60	67	87	88	89	90	$2.0 \cdot 10^{-12}$ M (mixture of all 10 mice)

The limit of detection (LOD) was determined combining two criteria. First, the calculated SIINFEKL concentration in the ELISA plate had to exceed the 95% confidence interval of all measured HPLC fractions of the negative control (cf. Table 3-13). Additionally, the calculated urinary SIINFEKL concentration was required to be at least factor 2 above the negative control. Generally, B6 mouse urine was used as a negative control for determination of LOD. However, in exp. no. 16, the LOD was determined utilising an HPLC blank run because, in this case, the background signal of the blank run was higher than that of B6 urine.

In summary, exp. no. 16 did confirm the results of exp. no. 15, again challenging the results of my diploma thesis and exp. no. 14. However, due to the customised experimental design, it also provided valuable hints for troubleshooting and future improvements of analysis regarding the three points mentioned above:

Regarding point 1:
B6/OVA$^+$ mice no. 53, 55, 58 and 59 (cf. Table 3-11) cannot be responsible for the discrepancy between exp. no. 14 and 15. In contrast, B6/OVA$^+$ mice no. 52 and 54 remained candidate "outlier mice" (which later indeed turned out to be the case for B6/OVA$^+$ no. 52, see page 76 and Figure 3-3).

Regarding point 2:
Of the three measured urine samples of the younger B6/OVA$^+$ mice (not included in Table 3-11), two had SIINFEKL concentrations below the limit of detection (LOD) and one slightly exceeded the LOD. This is in good congruence with the results obtained for the older B6/OVA$^+$ mice in exp. no. 15 and 16 (see Table 3-11) ruling out aging of mice as the reason for the conflict with previous experiments.

Regarding point 3:
I could measure a urinary SIINFEKL concentration being more than three times above the LOD when urine was collected manually. In contrast, using urine of exactly the same B6/OVA$^+$ mice collected in metabolic cages yielded a value slightly below the LOD (see Table 3-11). Although statistical testing is not possible due to the very low number of samples, this finding provided a tentative hint that peptides are indeed degraded during urine collection in metabolic cages. Later experiments also support this suspicion*.

A possible explanation for the discrepancy between exp. no. 15 and 16 versus former results might have been the loss of ovalbumin expression in B6/OVA$^+$ mice for some unknown reason. Therefore, my supervisors encouraged me to retype the utilised B6/OVA$^+$ mice for their ovalbumin expression (Table 3-12). However, the mean fluorescence intensity (MFI) of the ovalbumin staining was not decreased when compared to former ovalbumin stainings (see Supplementary Table 9-1). Surprisingly, the ovalbumin signal was significantly higher in older B6/OVA$^+$ mice than in younger B6/OVA$^+$ mice (Table 3-12). This might reflect an increase of ovalbumin expression with age but definite conclusions would require additional control samples consisting of old and young B6 mice. For example, the proportion of lymphocytes with certain Fc receptors could have changed with age altering the ovalbumin signal independent of ovalbumin expression. In any case, ovalbumin expression can be ruled out as the factor responsible for the discrepancies in measured urinary SIINFEKL concentrations.

Having performed exp. no. 15 and 16 as well as the rechecking of ovalbumin expression, the conclusion of my diploma thesis that SIINFEKL occurs in urine of B6/OVA$^+$ mice at a mean concentration of about $2 \cdot 10^{-11}$ M no longer seemed trustworthy. Unfortunately, using the methods established in my diploma thesis, the newly obtained data appeared to be in the range of the LOD and thereby not suitable for a reliable determination of urinary SIINFEKL concentrations. Most importantly, our ultimate goal was the examination of an MHC-dependent occurrence of SIINFEKL in urine. The B6/OVA$^+$ mice might have expressed more ovalbumin than the B6/OVA$^+$/β_2m$^{-/-}$ mice utilised as a control (see Figure 2-1, Supplementary Table 9-1 and methods section “2.1. Mice”). It is perfectly possible that B6/OVA/β_2m$^{-/-}$ mice express only halve as much ovalbumin as B6/OVA$^+$ mice because the former were heterozygous whereas the latter were homozygous for the ovalbumin transgene. Therefore, in order to demonstrate that the reduced urinary SIINFEKL concentration in

* I quantified nine peptides (LNSVFDQL, LNSVFDQLGSY, TRVLNLGPI, IDQTRVLNLGPI, VYRPDQVSIL, PAVRGFSL, PAVRGFSLEL, PVESKIYF and ALDYDPVESKIYF, cf. Tables 3-1 and 3-3 as well as Supplementary Table 9-3) using $^{13}C_5{}^{15}N_1$ valine in MS exp. no. 4 (cf. Table 2-1), where urine was collected manually. I had performed an additional MS experiment essentially identical to MS exp. no. 4 but with urine collected in metabolic cages over night in the presence of protease inhibitors. Comparing these two MS experiments, it turned out that the measured concentration of the nine peptides was on average about 10 times higher when urine was collected manually. However, different urine collection methods were not used within the same experiment, and due to the lack of independent biological replicates, these results represent only a tentative hint.

B6/OVA/$\beta_2m^{-/-}$ mice is due to an MHC-dependent effect, a difference clearly exceeding factor 2 has to be shown for these two mouse strains. Even the highest SIINFEKL concentration, obtained with 10 ml of manually collected B6/OVA$^+$ mouse urine, was only about three times above the LOD (Table 3-11). This is clearly not sufficient to demonstrate the required difference of factor >2. Hence, we strived for a more sensitive assay.

Table 3-12 | B6/OVA$^+$ mice of exp. no. 15 and 16 express normal levels of ovalbumin and the expression of ovalbumin does not decrease with the age of the mice.

Mouse strain	Ear no.	Age in month	MFI of anti-ovalbumin staining
B6/OVA$^+$	60	20.9	798.6
	87	17.6	857.9
	88	17.6	805.6
	89	17.1	836.3
	90	16.0	774.2
	mean of „old“ B6/OVA$^+$ mice (60, 87-90)	17.9	814.5
	98	5.8	734.2
	103	5.5	704.5
	104	5.4	716.7
	108	4.0	699.6
	111	2.9	656.7
	mean of „young“ B6/OVA$^+$ mice (98, 103, 104, 108, 111)	4.7	702.3
B6	15	7.2	45.8

All B6/OVA$^+$ mice were typed for expression of ovalbumin and CD8$^+$ T cells before their use for urine collection (cf. section “2.1. Mice” and Supplementary Table 9-1). However, data shown in this table were obtained four days after the completion of exp. no. 16 to rule out that B6/OVA$^+$ mice lost expression of ovalbumin for some unknown reason which would have explained the very low urinary concentrations of SIINFEKL in exp. no. 15 and 16. Surprisingly, the MFI of the anti-ovalbumin staining was significantly higher in older B6/OVA$^+$ mice as compared to younger B6/OVA$^+$ mice (P = 0.00046; two tailed, heteroskedastic Student`s t-test). The age of the mice is indicated for the timepoint of ovalbumin retyping (i.e. four days after completion of exp. no. 16). MFI = mean fluorescence intensity. Data depicted in this table are derived from a flow cytometry experiment jointly performed by Theo Sturm and Beate Pömmerl.

Peter Overath searched the literature for ELISA formats providing increased sensitivity and encountered the intelligent multifunctional analytical plates[294,316] (IMAPlates, see section “2.6.3. SIINFEKL-H2-Kb-ELISA in a 5 µl miniaturised format (IMAPlates)”). I established this 5 µl miniaturised ELISA format in our laboratory and compared its LOD with the LOD of the 50 µl standard ELISA format (described in sections “2.6.3. SIINFEKL-H2-Kb-ELISA in a 5 µl miniaturised format (IMAPlates)” and “2.6.2. SIINFEKL-H2-Kb-ELISA in 50 µl standard format”). I calculated the LODs as outlined in the following.

1.) The LOD is determined as the upper limit of the 95% confidence interval of the putative SIINFEKL concentrations in all measured HPLC fractions of the negative control.

2.) Urine of B6 mice is used as the negative control instead of an HPLC blank run or ELISA blank values obtained with the buffer solution. Thereby, I take into account that the applied 25-D1.16 monoclonal antibody can also bind to peptides other than SIINFEKL although this cross-reactivity is generally very low in the 10 or 20 s HPLC fractions (cf. my diploma thesis[203]).

3.) Putative SIINFEKL concentrations and the corresponding standard deviations (across the diverse HPLC fractions) of the negative control are based on the mean of the ELISA triplicate measurements, i.e. they were calculated in the same way that I applied for the other urine samples. Therefore, the standard deviations indicated in Table 3-13 represent standard deviations of the mean of triplicates and not standard deviations of single values, thereby decreasing the impact of outlier values[†].

The results of the LOD comparison between the 50 µl standard and the 5 µl miniaturised ELISA format are depicted in Table 3-13. In terms of SIINFEKL concentrations within the wells of the ELISA plates, I found no significant difference between the two formats. However, the 5 µl IMAPlate format requires a sample volume of only about 17 µl for a triplicate measurement, whereas 120 µl are necessary for a duplicate measurement in the 50 µl standard format. Consequently, when considering absolute amounts of required SIINFEKL, the LOD was about six times lower in the 5 µl versus the 50 µl format (0.25 fmol versus 1.51 fmol). Although the difference was only borderline significant ($P = 0.073$), I therefore decided to use the 5 µl IMAPlate format for all future measurements of SIINFEKL in mouse urine.

Identification of a B6/OVA$^+$ mouse with extraordinarily high urinary SIINFEKL concentrations

During manual collection of urine, I realised that the urine of the B6/OVA$^+$ mouse no. 52 had an orange colour and was unusually turbid. This mouse is a potential explanation for the discrepancies between the results of my diploma thesis and exp. no. 14 versus exps. no. 15 and 16 (cf. page 73 and Table 3-11). Having established the highly sensitive 5 µl IMAPlate ELISA, I was now able to measure urinary SIINFEKL concentrations in individual mice (see section "3.4.2. SIINFEKL occurs in urine of ovalbumin transgenic mice in an MHC-dependent manner"). I also included the B6/OVA$^+$ mouse no. 52 into these measurements. Already suspecting that this mouse might have exceedingly high urinary SIINFEKL concentrations, I measured only 1 ml instead of the usual 5 ml of urine. A separate 1 ml sample of B6 urine supplemented with 10^{-11} M synthetic SIINFEKL was applied for accurate determination of yield losses in the 1 ml versus the 5 ml preparation. Despite the low urine volume applied for B6/OVA$^+$ mouse no. 52, the measured absorbance in the ELISA exceeded the maximal value of the calibration curve in two neighbouring HPLC fractions. Therefore, I could not accurately measure the urinary SIINFEKL concentration of the B6/OVA$^+$ mouse no. 52, but it is possible to state

[†] The calculation of the median instead of the mean would have been a suitable alternative to cope with outlier values in all IMAPlate triplicate measurements, but due to laboratory tradition I did not apply the median for calculating putative SIINFEKL concentrations in individual HPLC fractions.

Table 3-13 | Comparison of the detection limit of the SIINFEKL-H2-K^b-ELISA in the 50 µl standard versus the 5 µl miniaturised format.

Format of ELISA	Exp. no.	Length of HPLC fractions in s	Total sample volume in µl	Calculated concentration of SIINFEKL in ELISA plate in 10^{-10} M (mean of duplicates or triplicates respectively); the HPLC fraction of urine of B6 mice is indicated											Calculated concentration of SIINFEKL in ELISA plate in 10^{-10} M			Detection limit of SIINFEKL in fmol
				1	2	3	4	5	6	7	8	9	10	11	Mean	Standard deviation	Detection limit	
50 µl standard format	11	20	120	0.007	0.040	0.138	0.250	0.086	0.222	0.054	0.005	0.010	0.026	n.a.	0.084	0.090	0.261	3.13
	12	20	120	0.075	0.004	-0.047	0.056	0.077	0.079	0.033	0.002	-0.002	-0.003	n.a.	0.027	0.043	0.112	1.34
	13	20	120	0.102	0.022	0.025	0.118	0.062	0.156	0.144	0.053	0.043	0.006	n.a.	0.073	0.053	0.177	2.13
	14	10	120	0.060	0.015	0.015	0.019	0.056	-0.005	0.053	0.019	0.012	0.015	n.a.	0.026	0.022	0.069	0.83
	16	20	120	-0.042	-0.031	-0.058	0.004	-0.019	-0.027	-0.046	-0.042	0.000	-0.038	n.a.	-0.030	0.020	0.010	0.11
	Mean of 11 - 14, 16														0.036	0.046	0.126	1.51
5 µl miniaturised format (IMAPlates)	17	20	17.0	-0.141	-0.089	0.153	-0.007	0.007	0.162	-0.015	0.077	0.048	0.054	-0.138	0.010	0.103	0.213	0.36
	18	10	17.0	-0.170	-0.072	0.012	-0.145	0.100	0.055	0.028	0.006	0.010	0.129	-0.161	-0.019	0.104	0.184	0.31
	19	10	17.0	-0.141	-0.121	-0.052	-0.074	0.044	0.052	-0.018	0.038	0.074	0.107	0.061	-0.003	0.083	0.161	0.27
	20	10	17.5	-0.194	-0.220	-0.075	-0.194	0.016	-0.067	-0.078	-0.059	0.014	0.069	-0.012	-0.073	0.095	0.114	0.20
	21	10	17.5	-0.038	-0.157	-0.015	-0.190	0.082	-0.002	-0.066	-0.020	-0.013	0.046	-0.035	-0.037	0.079	0.117	0.21
	22	10	17.5	-0.150	-0.159	-0.018	-0.161	-0.009	-0.066	-0.083	-0.092	0.015	0.051	-0.030	-0.064	0.072	0.078	0.14
	Mean of 17 - 22														-0.031	0.089	0.144	0.25
P-value (two-tailed, heteroskedastic t-test; miniaturised vs. standard format)															0.028	0.023	0.710	0.073
Ratio [standard; mean of exp. no. 11 − 15] / [miniaturised; mean of exp. no. 16 − 21]																0.51	0.87	6.08

Note that the difference in sensitivity resulting from the analysis of 10 s versus 20 s HPLC fractions has not been examined systematically. The length of HPLC fractions as well as the presence of selected, abundant (about 10^{-5} M) H2-K^b binding peptides apart from SIINFEKL can influence the signal measured in SIINFEKL-H2-K^b-ELISA (see my diploma thesis[203]). However, this possible effect is supposed to be of minor importance in the analytical setup of experiments no. 11 − 22 and was therefore not included in the calculations depicted in this table. In the 50 µl standard ELISA, samples were measured in duplicate whereas I employed triplicates for the 5 µl miniaturised format to compensate for the higher standard deviation associated with this format. The putative concentrations of SIINFEKL in the ELISA plate are derived from the measured absorbance values employing the individual SIINFEKL calibration curve of the respective experiment. The calculated concentrations of SIINFEKL in the B6 urine fractions do not necessarily represent (contaminating synthetic) SIINFEKL but can also indicate cross-reactivity of the 25-D1.16 antibody with urinary peptides bound to the ELISA plate or just reflect assay variability. Negative numbers for calculated concentrations of SIINFEKL originate if measured ELISA absorbance values are below those of ELISA loading buffer. They are mathematical artefacts resulting from normal assay variability. The "detection limit of SIINFEKL in fmol" refers to the total amount of SIINFEKL required for duplicate or triplicate measurements respectively but does not include the losses derived from urine sample preparation (usually 40 − 70%). A few SIINFEKL-H2-K^b-ELISA experiments with B6 mouse urine are not considered in this table because they gave rise to a signal above the detection limit in exactly the same HPLC fraction (or the neighbouring fraction) in which synthetic SIINFEKL eluted in the positive control of the respective experiment. Therefore, in these cases, it is likely that the significant signal resulted from a contamination of the B6 mouse urine sample with synthetic SIINFEKL leading to an increased standard deviation unrelated to the performance of the ELISA itself. Although the numbering of experiments is different from that used in my diploma thesis[203], I performed experiments no. 11 − 13 during the diploma studies. n.a. = not analysed

that it is (probably much) higher than $6.7 \cdot 10^{-11}$ M. If urine of this mouse is mixed with urine of about eleven other B6/OVA$^+$ mice containing only minor amounts of SIINFEKL, one will measure relatively high SIINFEKL concentrations in the mixture. This was apparently the problem during my diploma thesis and in exp. no. 14 providing a conclusive explanation for the discrepancies in relation to exps. no. 15 and 16 (cf. Table 3-11) as well as with regard to the data depicted in Figure 3-4.

To objectively demonstrate that the urine of the B6/OVA$^+$ mouse no. 52 represents an unusual condition that should not be included into the calculation of the mean urinary SIINFEKL concentration of B6/OVA$^+$ mice, I acquired the absorbance spectra of urine of B6/OVA$^+$/β_2m$^{-/-}$ mice and B6/OVA$^+$ mice including B6/OVA$^+$ no. 52. The results are depicted in Figure 3-3. At wavelengths from 385 nm to about 600 nm the absorbance of B6/OVA$^+$ no. 52 urine significantly exceeds that of the other B6/OVA$^+$(/β_2m$^{-/-}$) mice (based on the calculation of 95% confidence intervals). Neither absorbance at 280 nm nor a Lowry protein assay performed by Peter Overath did reveal evidence for increased urinary protein concentrations in B6/OVA$^+$ mouse no. 52. Applying Combur^9 test stripes (cobas, Roche Diagnostics, Mannheim) designed for clinical evaluation of human urine samples, I obtained positive reactions for granulocytes, haemoglobin, bilirubin and urobilinogen in urine of B6/OVA$^+$ mouse no. 52 but not in urine of the other B6/OVA$^+$(/β_2m$^{-/-}$) mice. Hence, B6/OVA$^+$ mouse no. 52 appeared to be in a pathologic condition. Because (neutrophil) granulocytes express much more MHC class I molecules than non-immune cells[89], it is tempting to speculate that the urinary granulocytes released critical amounts of SIINFEKL from their H2-K^b molecules into urine leading to the elevated SIINFEKL concentrations measured. It would be interesting to examine whether elevated concentrations of MHC-dependent peptides in urine always coincide with abnormal levels of the urinary biomarkers of Combur^9 test stripes, or whether such peptides can serve as indicators for specific diseases in their own right when the standard biomarkers of urine are at normal levels. However, given the technical possibilities available today, detection of MHC-dependent peptides in human urine appears to be a very challenging task at least in the healthy condition.

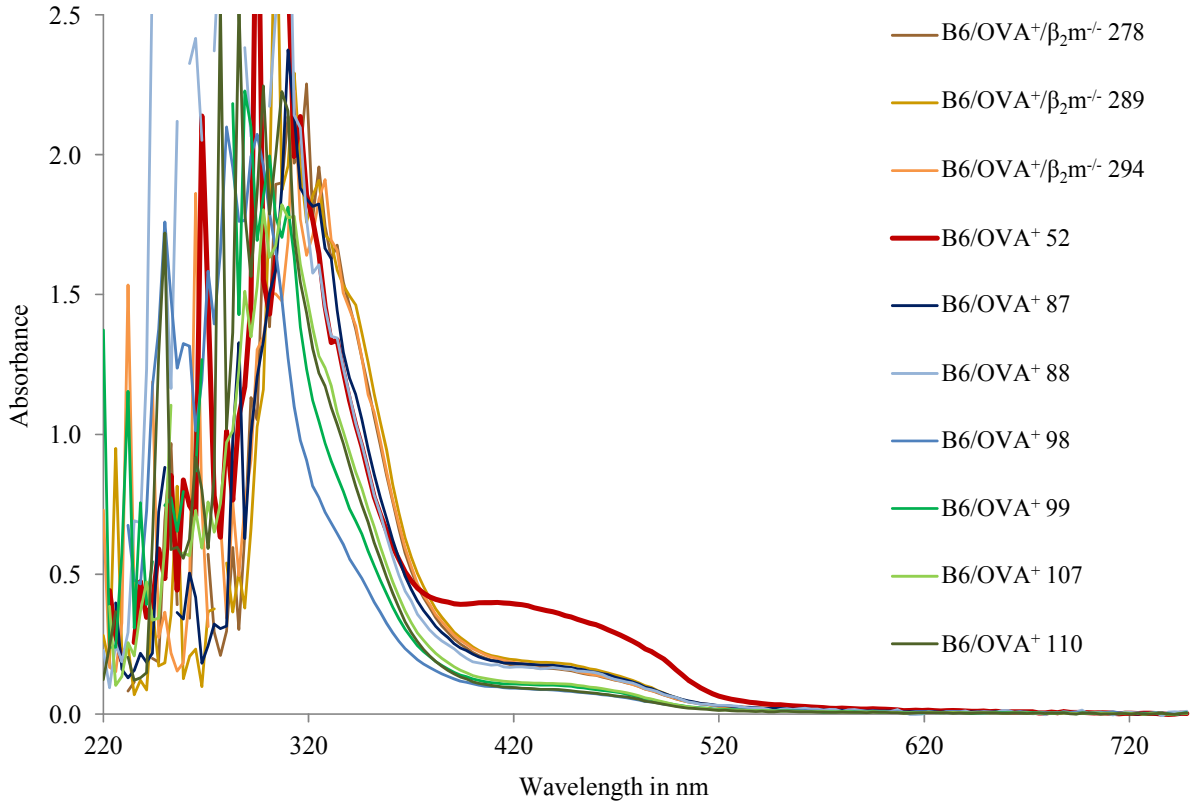


Figure 3-3 | The absorbance spectrum of urine of B6/OVA$^+$ mouse no. 52 differs significantly from that of other B6/OVA$^+$ mice and B6/OVA$^+$/β_2m$^{-/-}$ mice. Absorbance was measured applying a NanoDrop 1000 Spectrophotometer from Thermo Scientific.

3.4.2. SIINFEKL occurs in urine of ovalbumin transgenic mice in an MHC-dependent manner

Having optimised the sensitivity of the SIINFEKL-H2-K^b-ELISA and the urine collection procedure (see section "3.4.1. Challenging the results of my diploma thesis"), I was able to successfully measure the urinary SIINFEKL concentrations in manually collected urine of individual B6/OVA$^+$, B6/OVA$^+$/β_2m$^{-/-}$ and B6 mice. Representative results of this assay performed on HPLC fractions of free urinary peptides (not bound to carrier molecules) from three mouse strains and a sample of B6 urine spiked with $5 \cdot 10^{-12}$ M SIINFEKL are shown in Figure 3-4a. B6/OVA$^+$ mice gave rise to a peak at the same retention time as the spiked peptide. Importantly, SIINFEKL was not detectable in B6/OVA$^+$/β_2m$^{-/-}$ and B6 mice. Therefore, the appearance of this peptide in urine requires the expression of both the protein of origin and the cognate MHC class I molecule as is the case for cellular MHC ligands[90]. Urine from B6/OVA$^+$ mice contained a median concentration of $4 \cdot 10^{-12}$ M SIINFEKL (Figure 3-4b). This amount was present in free form because it was not retained by the 3 kDa ultrafilter used as a first step for peptide purification. The corresponding concentration detected in B6/OVA$^+$/β_2m$^{-/-}$ and B6 urine was $\leq 0.1 \cdot 10^{-12}$ M and $\leq 0.3 \cdot 10^{-12}$ M, respectively (Figure 3-4b). By

eluting peptides from potential carrier molecules using 20% acetonitrile[317] at pH 2, the fraction of SIINFEKL bound to undefined high molecular weight components was determined as 10-35% of total urinary SIINFEKL (*Nat Comm*, joint).

Milinski and co-workers reported that the putative MHC-dependent signal of sticklebacks (*Gasterosteus aculeatus*), which this group suggested to consist of MHC-dependent peptides[100], was only produced by reproductively active males[318]. These authors therefore assumed that the MHC signal is not just a by-product of normal immunological functions but rather associated with costs for the organism. Such costs were hypothesised to originate from cellular shedding of MHC-peptide complexes leading to local suboptimal T cell immunosurveillance of the MHC-depleted cells[144,318]. The mice used for the generation of the data in Figure 3-4 were singly kept males, although female mice were present in different cages in the same room. Male mice older than about 16 months are usually infertile and their androgen levels will be decreased in comparison to younger males. I therefore analysed whether there is a correlation between the urinary SIINFEKL concentrations and the age of the mice. However, I obtained no evidence for this idea (Figure 3-5). Obviously, MHC-dependent peptides are also present in urine of reproductively inactive male mice.

Originally, the MHC anchor amino acid residues were reported to be of crucial importance for the recognition of MHC peptide ligands by VSNs, and all peptides fitting to the same MHC binding motif were apparently assumed to be recognised as a single unit[96]. This model raised a problem for the chemosensory detection of MHC-dependent peptides when considering our new biochemical analyses of urinary peptides. We found that the MHC-dependent model peptide SIIN**F**EK**L** occurs at urinary concentrations of only about $4 \cdot 10^{-12}$ M, while other peptides fitting to the H2-K^b binding motif were much more abundant (Tables 3-1 and 3-2). LNSV**F**DQ**L**, which has exactly the same anchor amino acid residues as SIIN**F**EK**L**, occurred at concentrations of up to 10^{-6} M (Table 3-3). We therefore had to test whether VSNs are capable of selective detection of natural concentrations of SIINFEKL within the complex background of urine. Applying urine of the same B6/OVA^+ mice that I used for the SIINFEKL-H2-K^b-ELISA (Figure 3-4), Trese Leinders-Zufall demonstrated this to be the case. Remarkably, taking advantage of VSN dose response curves obtained with synthetic SIINFEKL, she measured the mean concentration of urinary SIINFEKL to be about $3 \cdot 10^{-12}$ M in B6/OVA^+ mice (Figure 2 of Sturm *et al.*[207]). This is in close accordance with the results that I obtained using the SIINFEKL-H2-K^b-ELISA (Figure 3-4).

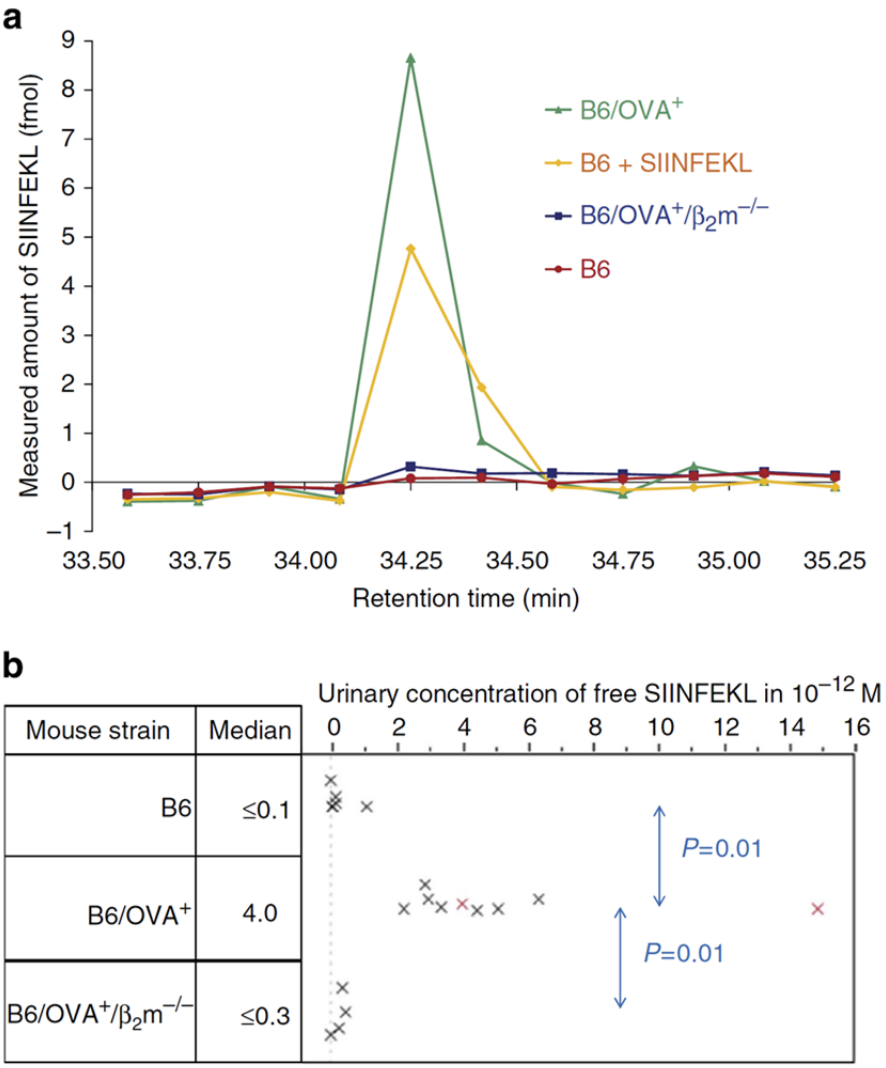


Figure 3-4 | HPLC fractions derived from urine are tested for the MHC model peptide SIINFEKL by ELISA. (**a**) A fraction of urine from a B6/OVA$^+$ mouse gives rise to a strong signal while the corresponding preparations from a B6/OVA$^+$/β_2m$^{-/-}$ mouse and B6 mice are negative. Spiked synthetic SIINFEKL elutes at the same retention time as the substance detected in B6/OVA$^+$ urine. (**b**) SIINFEKL concentrations measured in urine of individual B6/OVA$^+$ mice (n = 9). For comparison, results for individual B6/OVA$^+$/β_2m$^{-/-}$ (n = 4) and B6 wild type (n = 5) mice are shown. The two data points printed in red are derived from two separate urine samples of the same mouse. A two-tailed, heteroskedastic Student`s t-test was applied with Bonferroni correction for two comparisons (*Nat Comm*, Theo).

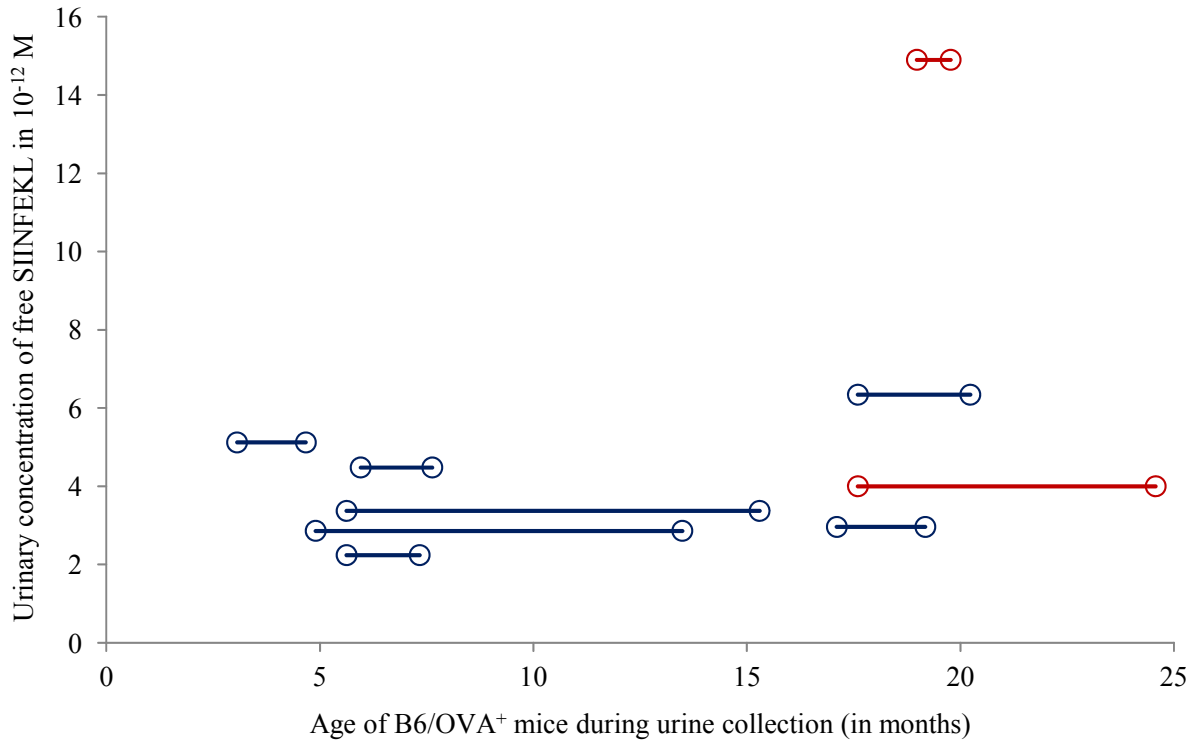


Figure 3-5 | The urinary concentration of the MHC model peptide SIINFEKL shows no obvious correlation with the age of the B6/OVA$^+$ mice. Urine samples obtained from collections on different days were pooled for each individual mouse. The age of the B6/OVA$^+$ mice at the first and the last day of urine collection is marked by circles. Each blue data bar represents a different individual B6/OVA$^+$ mouse. The two data bars printed in red are derived from two separate urine samples of the same mouse. Every data bar is based on a single ELISA measurement and therefore contains no information about the time dependency of the urinary concentration of SIINFEKL in an individual mouse. Because SIINFEKL was present in B6/OVA$^+$ mice of all ages between 3 and 25 months and the age of the control mice (B6 and B6/OVA$^+$/β_2m$^{-/-}$) was within this range, an influence of the age on the MHC-dependent occurrence of SIINFEKL in urine can be ruled out (*Nat Comm*, Theo).

4. Discussion

This study provides a molecular basis for the evaluation of genotype by nasal chemosensory neurons. By analysing the identity and genomic origin of naturally occurring peptides in mouse urine, we have discovered multiple, partly overlapping peptide pools that contain distinct types of information related to genomic individuality. We have identified a novel family of MHC-independent peptides that differ in strains by single amino acid exchanges, and we show for the first time that a *bona fide* MHC peptide ligand exists in urine in an MHC-dependent manner at biologically relevant concentrations. Hence, the emerging picture is a complex scenario in which multiple peptide families together with previously characterised major urinary proteins (cf. section "1.6.2. Recognition of genomic identity based on non-MHC genes"), formylated and maternally-inherited peptides (not yet shown to exist in urine)[131] and possibly other yet undiscovered ligand families link the genotype of an individual with nasal chemoreception mechanisms[207,319]. Future research should examine the contribution of these peptides to genotype discrimination in diverse behavioural contexts applying physiologically relevant concentrations as provided in this thesis (*Nat Comm*, joint).

4.1. MHC-dependent peptides in mouse urine

We demonstrate for the first time that a prototypic MHC class I peptide ligand is present in urine in an MHC-dependent manner at a concentration of $4 \cdot 10^{-12}$ M in a free form. This finding provides evidence both for and against the hypothesis that such immune-related peptides mediate recognition of individuality in mice. The demonstration of the *occurrence* of an MHC-dependent peptide in urine fills the most critical gap that has remained for this hypothesis thereby essentially supporting it. However, the observed urinary *concentration* of the MHC peptide ligand is about eight orders of magnitude lower than that previously applied in behavioural experiments[96,99]. Selected aspects of behavioural experiments are discussed in section "4.2. Questioning the role of the MHC as an olfactory signal in social behaviours". I will first focus on the MHC-dependent urinary peptides themselves (*Nat Comm*, joint).

The following arguments support the conclusion that we really detected SIINFEKL in urine: First, the peptide must descend from ovalbumin because it is not present in urine of B6 wild type mice ($OVA^{-/-}/\beta_2m^{+/+}$). Second, the peptide elutes from a reversed phase HPLC column at the same retention time as synthetic SIINFEKL (in almost all replicates exactly the same 10 s fraction within a 1 h gradient). Third, the peptide is detected by a SIINFEKL-H2-K^b-specific monoclonal antibody. Forth, urine from ovalbumin-positive mice stimulates SIINFEKL-reactive VSNs, whereas urine from $B6/OVA^{+}/\beta_2m^{-/-}$ mice does not. Fifth, the concentration of free SIINFEKL in urine of $B6/OVA^{+}$ mice

determined by ELISA ($4 \cdot 10^{-12}$ M) corresponds to the estimate obtained by comparison of B6/OVA$^+$ urine-evoked neuronal activity with the dose-response curve for synthetic SIINFEKL ($\approx 3 \cdot 10^{-12}$ M) (*Nat Comm*, joint).

The free form of urinary peptides is considered to be responsible for neuronal stimulation[98]. I found that this free form is the major form of urinary SIINFEKL. Only 10 to 35% of total urinary SIINFEKL were bound to undefined high molecular weight components. Such carrier molecules could include MHC molecules with the appropriate binding motif. During my diploma thesis, I determined the concentration of H2-K^b molecules in urine of B6 mice to be about 1 nM[203]. However, this assay was performed with the B8-24-3 monoclonal antibody[320] directed against the α1 domain of H2-K^b. One can assume that the majority of urinary MHC molecules has lost its original peptide ligand because most urinary MHC molecules consist only of the α1 and α2 domains[321]. It would be revealing to immunoprecipitate urinary MHC molecules and to analyse the eluted peptide ligands by LC-MS. In doing so, monoclonal antibodies that are directed against properly folded, intact MHC molecules should be included (e.g. W6/32 [322] for MHC class I molecules of human urine). It is also conceivable that some of the urinary SIINFEKL is non-specifically adsorbed to abundant urinary proteins like MUPs and uromodulin. For blood plasma, it is well known that dominant plasma proteins including serum albumin and apolipoprotein act as molecular sponges binding a multitude of smaller proteins and peptides[323]. Regardless of the nature of the carrier molecules of bound urinary SIINFEKL, it remains unknown whether the observed free urinary SIINFEKL is released from H2-K^b molecules in the body tissues, the blood, the urothelium and / or directly within urine.

From our data for SIINFEKL, we infer that other endogenously processed, but wild type-encoded MHC-dependent peptides are likely present in urine in similar or even lower concentrations although they cannot be detected by today`s mass spectrometry. The urinary SIINFEKL concentration determined in B6/OVA$^+$ mice is probably higher than the concentration of most other MHC class I peptides in urine of outbred or wild mice. There are four reasons for this assumption. First, SIINFEKL is an optimal H2-K^b ligand having a sequence perfectly fitting to the H2-K^b binding motif. The phenylalanine at position 5 is the best anchor residue for H2-K^b and the leucine at the C-terminus shares this first rank only with methionine[324]. SIINFEKL`s affinity for H2-K^b is higher than that of many other peptides bearing an H2-K^b binding motif even when compared to those having exactly the same anchor amino acids (Deres *et al.*[324] and Figure 3-1). Second, SIINFEKL is special because it can bind not only to one but two different MHC class I molecules, namely H2-K^b and H2-D^b [324]. The presentation of SIINFEKL by H2-D^b molecules is not known from *in vivo* studies. However, such a presentation is likely because of SIINFEKL`s H2-D^b binding capabilities and its efficient generation in the MHC class I processing pathway (which is demonstrated by the *in vivo* presentation on H2-K^b molecules)[325]. The third special parameter of our model system is expected to have pushed urinary SIINFEKL concentrations by factor two. Inbred mice like B6/OVA$^+$ are homozygous for all

autosomal genes, whereas outbred or wild mice are usually heterozygous at the MHC. The homozygous expression of the *H2*^{b} haplotype is supposed to lead to a two-fold increase of the presentation of H2-K^{b} and H2-D^{b} peptide ligands like SIINFEKL as compared to a heterozygous *H2*^{b} setting. This difference is probably reflected in the corresponding MHC-dependent urinary peptides. The fourth variable is very important but the one with the highest uncertainties - it is the expression of SIINFEKL`s source protein, ovalbumin, in the B6/OVA^{+} mice. It is known that ovalbumin is expressed in all organs of B6/OVA^{+} mice. In principal, this will elevate urinary concentrations of ovalbumin derived peptides relative to other peptides descending from proteins with tissue-specific expression. Notably, we cannot exclude the possibility that only peptides presented on MHC molecules of the urinary tract occur in urine at concentrations relevant for detection by VSNs. Besides the tissue distribution of source protein expression, the cellular abundance and turn-over[119,326] of the source protein will be crucial in determining the overall number of SIINFEKL molecules presented on MHC molecules of a B6/OVA^{+} mouse. However, both the average copy number of ovalbumin molecules per cell and the turnover rates of ovalbumin are unknown.

In L929 cells transfected with H2-K^{b} and infected with a recombinant vaccinia virus expressing soluble ovalbumin, SIINFEKL occupied 9% of all H2-K^{b} molecules. When the vaccinia virus contained a SIINFEKL-coding minigene instead of the ovalbumin gene, this percentage even exceeded 85% [218]. But the biological setup in the L929 cell line differs largely from B6/OVA^{+} mice which express ovalbumin as an artificial membrane protein under the control of the β-actin promoter of *Gallus gallus*[202]. Despite the widespread occurrence of ovalbumin in all organs, expression of SIINFEKL-H2-K^{b} complexes was not high enough for immunohistochemical detection on all cells with H2-K^{b} expression in B6/OVA^{+} mice. In tail sections of B6/OVA^{+} mice only cells with dendritic morphology could be stained using the 25-D1.16 αSIINFEKL-H2-K^{b} antibody[202]. These findings argue against the assumption that MHC-associated SIINFEKL was extremely abundant throughout the body of our mice.

The urinary concentrations of all MHC-dependent peptides may add up to approximately 10^{-10} to 10^{-8} M if one assumes that SIINFEKL represents between 0.1 and 10% of the H2-K^{b} ligands[325,327]. However, for the detection by VSNs, concentrations of individual peptides appear to be more relevant than the summed concentration of all MHC-dependent peptides corresponding to a certain MHC allele. VSNs also discriminate amino acids apart from the MHC anchor positions (Leinders-Zufall *et al.*[98] and Figure 4 of Sturm *et al.*[207]). Accordingly, VSNs were able to detect SIIN**F**EK**L** within B6/OVA^{+} urine (Figure 2 of Sturm *et al.*[207]) although much more abundant MHC-independent peptides with the same H2-K^{b} binding motif are present in mouse urine, and some of them (e.g. LNSV**F**DQ**L**, cf. Tables 3-1, 3-2 and 3-3) have exactly the same anchor amino acid residues as SIIN**F**EK**L**.

Our finding of a median concentration of free urinary SIINFEKL of only about $4 \cdot 10^{-12}$ M is in obvious conflict with a report of Hovis *et al.*[176] which was published when our own study[207] was under revision for *Nature Communications*. Hovis *et al.*suggested that the MHC model peptide S**Y**FPEITH**I** occurs in mouse urine at a concentration of $6 \cdot 10^{-9}$ M. However, since this report neither listed the mouse strain or its MHC type nor included a control for the possible contamination of urine samples with the extensively used synthetic SYFPEITHI, it provides no information on the MHC-dependence of this peptide in urine. Of note, the amount of SYFPEITHI was at the very low end of the calibration curve applied for LC-MS detection (see Figure 3C of Hovis *et al.*[176]). In this context, it should be mentioned that I myself struggled with contamination of our B6 urine samples with synthetic SIINFEKL for several months. I got rid of this problem only by extensive cleaning of all laboratory equipment and thorough, repeated removal of all dust from shelves and other surfaces in the working rooms. I performed blank runs between every preparative HPLC run and did not handle synthetic SIINFEKL powder in our own laboratory but went to the Hubert Kalbacher group instead to dissolve and purify synthetic SIINFEKL. Such extensive efforts are not standard in many laboratories. In any case, negative controls appear essential in this context (*Nat Comm*, Theo).

Given the new observation that an MHC motif is required neither for the detection of a peptide nor for the discrimination of its individual amino acid residues by VSNs, another type of MHC-dependent urinary peptides might play a role in recognition of genomic identity. These are MHC peptides, i.e. peptides directly derived from the MHC class I α chain or from MHC class II α and / or β chains. During my diploma thesis[203], I measured the H2-K^b concentration in urine of B6 mice to be about 1 nM (cf. page 84). This is in good congruence with the findings of Singh *et al.* for rat urine[171]. Analysis of rat urine has revealed that fragments of MHC molecules having a size of 27 or 39 kDa enormously dominate over intact MHC molecules in this fluid[321] providing evidence for the proteolysis of urinary MHC molecules. Therefore, it is conceivable that MHC peptides occur at higher urinary concentrations than MHC peptide ligands although both of these peptide groups were not contained in our set of 639 peptides identified by MS. The MHC protein sequences are highly polymorphic and some MHC genes are expressed in an allele-specific manner. For example, H2-L is present in mice of the *H2*^{d} haplotype but not in those of the *H2*^{b} haplotype. Consequently, many H2-L derived peptides will discriminate e.g. B6 and BALB/c mice in an MHC-dependent manner if they should be present in urine.

4.2. Questioning the role of the MHC as an olfactory signal in social behaviours

Until now, there is no conclusive evidence for the importance of MHC-dependent olfactory cues in social communication. Studies claiming an influence of the MHC on behavioural reactions have mostly been based on correlations of the behaviour with the MHC type of the individuals but often miss controls for non-MHC genes. A conclusive chain of evidence for MHC-related behaviour should comprise (1) at least one MHC-dependent substance occurring in a body fluid accessible for olfaction, (2) the demonstration of chemosensory neurons recognising this (these) substance(s) within the body fluid, and (3) behavioural reactions evoked by physiological concentrations of the substance(s) in a natural context. Such a comprehensive chain of evidence is still missing.

4.2.1. The role of MHC-dependent peptides in social behaviours

There is only a single MHC-dependent molecule that has been demonstrated to occur in urine and for which the observed MHC-dependent occurrence can be plausibly linked to a single MHC gene. This molecule is SIINFEKL as demonstrated in this thesis. The synthetic MHC peptide ligands applied in behavioural experiments so far[96,97,99,100] have only been hypothesised but not demonstrated to occur in urine or other body fluids accessible for olfactory evaluation. The putative model MHC ligand AAPD**N**RET**F** utilised in several neurobiological and behavioural studies of key importance for the olfactory MHC hypothesis[6,96,97,99,100,176] represents an especially problematic case. This peptide is neither encoded in the mouse genome nor contained in any protein of the UniProtKB database[108] (BLAST search from 17 June 2013). The H2-D^b ligand A**A**PD**N**RET**F** had been suggested after Edman peptide sequencing[328], but subsequent MS analyses of B6 mice demonstrated it to be **K**APD**N**RET**L** instead[329]. Whereas **K**APD**N**RET**L** perfectly fits to the H2-D^b binding motif, the C-terminal phenylalanine of A**A**PD**N**RET**F** does not[91,93,94].

The most important experiments that suggested a role for MHC class I peptide ligands in social behaviours are those related to the pregnancy block[96,99]. These experiments were performed several years before the urinary concentration of an MHC-dependent peptide was measured directly. Indeed, the peptide concentration used for the pregnancy blocks was about eight orders of magnitude (!) higher than that of SIINFEKL in B6/OVA$^+$ urine. Therefore, applied peptide concentrations were far away from the physiological situation. This circumstance makes it possible that the findings just represent laboratory artefacts although negative control peptides were applied at the same concentrations. Apart from concentrations, the major problem is how the mice shall discriminate the MHC-related information contained in MHC-dependent peptides from the information on genomic relatedness

contained in all urinary peptides. In the following, I will discuss three different scenarios to explain MHC signalling via MHC peptide ligands in urine.

Scenario 1: MHC-dependent peptides are recognised solely via their anchor amino acid residues

Previous pregnancy block experiments were based on the assumption that only the MHC binding motif (consisting of the anchor residues and an appropriate peptide length) is important in determining responses of sensory neurons in the VNO[96] and MOE[97], i.e. it was assumed that self-encoded and nonself-encoded peptides are recognised as a unit as long as they fit to the same MHC binding motif. Accordingly, Leinders-Zufall *et al.*[96] included MHC peptide ligands encoded by *Plasmodium berghei*, influenza virus or simian virus 40 into the pregnancy block experiments. Surprisingly, the pregnant mice considered these peptides as familiar (associated with the mating male) or unfamiliar depending on the *H2* haplotype of the mating male and the MHC motif of the peptides. For these experiments, MHC-dependent urinary peptides were hypothesised to constitute a dominant fraction of the urinary peptidome, and MHC-independent MHC motif peptides were not considered[95,96]. In this theoretical model, the signal received by peptide-sensitive VSNs and OSNs is clearly attributable to the MHC, and confounding interactions with other genes are not a problem. However, if all peptides with a binding motif corresponding to a given MHC allele were considered as a unit, and only their summed concentrations were relevant for olfactory recognition and associated behavioural consequences, a role for MHC-dependent urinary peptides could be clearly ruled out with the insights obtained in the present study. As now demonstrated by us, the sum of the urinary concentrations of MHC-dependent peptides is just negligible in comparison to the concentration of MHC-independent peptides with MHC binding motif. Therefore, scenario 1 is not an option anymore.

However, in 2009, Leinders-Zufall *et al.*[98] demonstrated that amino acids apart from the anchor positions are also discriminated by VSNs. During the cooperation with our group, Trese Leinders-Zufall showed that SIINFEKL is selectively recognised in the complex background of other peptides of B6/OVA$^+$ urine (Fig. 2 of Sturm *et al.*[207]) despite of the presence of urinary MHC motif peptides having exactly the same anchor residues as SIINFEKL and being up to six orders of magnitude more abundant (Tables 3-1, 3-2, 3-3 and 4-1). As demonstrated with a 12-mer SAV peptide pair, an MHC motif is required neither for the detection nor the sequence specific discrimination of peptides by VSNs (Fig. 4 in Sturm *et al.*[207]). Consequently, MHC-dependent peptides are detected on an individual basis establishing scenarios 2 and 3 discussed in the following. In these scenarios, the MHC anchor positions cannot be used as an indicator for a specific MHC allele because this would lead to confusions with MHC-independent MHC motif peptides in natural settings.

Many but not all VSNs responding to synthetic SIINFEKL also responded to B6 urine, which is devoid of this peptide[207]. For this reason, and because of the relatively low total number of VSNs (about 300.000 per mouse) in comparison to the vast multitude of potential peptides, individual peptides can only be discriminated by the combinatorial activation pattern of many different VSNs.

This conclusion is in accordance with a systematic study applying several synthetic peptides[98]. At least some MHC-independent MHC motif peptides are indistinguishable from MHC-dependent ones in terms of principal structural features and MHC binding properties (see section "3.1. Abundant peptides with MHC class I binding motif occur in mouse urine in an MHC-independent manner"). MHC-independent urinary MHC motif peptides can indeed be recognised by VSNs (Figure 3 of Sturm *et al.*[207]). Hence, specific VSNs specialised on the detection of MHC-dependent peptide ligands do probably not exist. This creates the problem of discriminating MHC-dependent from MHC-independent differences, a challenge not present in scenario 1. In other words, how shall the olfactory system recognise that the detected MHC-dependent difference is associated with the immune system of a conspecific and not just one out of hundreds or thousands of different signals of genomic relatedness transferred by other peptides? In the following, I will discuss two options how to achieve this discrimination, but I will finally conclude that both scenarios are very unlikely. For both scenarios, the mice would first have to learn all MHC-independent MHC motif peptides of their own urine to be "self" and irrelevant for MHC evaluation.

Scenario 2: The presence of unfamiliar, individually recognised MHC-dependent peptides constitutes the MHC signal

In theory, a mouse could consider only unfamiliar peptides with MHC motif to gain information about the MHC of a different mouse. Having learned the peptides with MHC motif (MHC-dependent and MHC-independent ones) from its own urine, unfamiliar peptides with MHC motif would point to a different MHC type in scenario 2. However, there are also differences in MHC-independent MHC motif peptides between mouse strains. If, for example, a *Mus musculus domesticus* mouse of the $H2^d$ haplotype would get used to the dominant H2-K^b motif peptide LNSV**F**D**Q**__**L**__ from its own urine, it should consider the SAV version LNSV**F**D**R**__**L**__ occurring in *Mus musculus musculus* as an unfamiliar peptide (cf. Tables 3-4 and 3-6). This might indeed be the case. But if the *M. m. domesticus* mouse would then assume that the *M. m. musculus* mouse is of a different MHC haplotype ($H2^b$) than itself, this might be a wrong conclusion. Given the multitude of different MHC alleles in *Mus musculus*, each characterised by a distinct MHC binding motif, a major problem for scenario 2 is also how a mouse should know whether an unfamiliar peptide fits to any of the many MHC binding motifs present in the population, and whether it is therefore relevant for the evaluation of the MHC type. This problem is obviated by scenario 3.

Scenario 3: The absence of familiar, individually recognised MHC-dependent peptides constitutes the MHC signal

Another theoretical option would be that a mouse uses its own MHC peptide ligands as a reference. However, this reference cannot be obtained from urine because this would lead to the same confusions with MHC motif peptides that I have already described above. Instead, a mouse could take advantage of the individual MHC peptide ligands presented on its own cell surfaces. Although the underlying

biochemical and cellular mechanisms are hard to imagine, one could speculate that the olfactory system is specifically educated for the recognition of self MHC peptide during early developmental processes (cf. the "education" of T cells early in live). While the learned individual MHC-dependent peptides are present in the mouse`s own urine, they will be absent in the urine of mice of different MHC types, and this lack might be used to recognise MHC dissimilarity. But apart from the biochemical and cellular mechanisms of referencing, scenario 3 implies a second major problem. MHC class I molecules are expressed by almost all cell types. Consequently, I would assume that, in analogy to urine, self MHC peptide ligands are continuously released into the luminal fluid of the VNO and into the olfactory mucus of the MOE. Such peptides might also be pumped to the chemosensory neurons from the cells lining the nasal ducts and nasal cavities during sniffing of urine. Because MHC-dependent peptides are present in urine only at picomolar concentrations, and a dilution of the urine of about factor 100 occurs during sniffing[7], the effective concentration of self MHC peptide ligands surrounding VSNs and OSNs might be essentially the same when sniffing urine either containing or not containing MHC peptide ligands corresponding to self MHC alleles.

Summarising scenarios 1, 2 and 3, I see no consistent hypothesis of how a mouse should use MHC-dependent urinary peptides to gain information about the MHC type of a conspecific.

Table 4-1 | Comparison of peptide concentrations in mouse urine with the sensitivity of chemosensory neurons and the peptide concentrations applied in behavioural experiments.

Peptide	Concentration (M)
MHC-dependent peptide in urine of B6/OVA^{+} mice: SIIN**F**EK**L**	$\approx 4 \cdot 10^{-12}$
MHC motif peptides in urine, MHC-independent: e.g. LNSV**F**DQ**L**	up to $\geq 10^{-6}$
SAV peptides in urine: e.g. LNSVFDQL	up to $\geq 10^{-6}$
EC_{50} at VSNs of mice[98]	10^{-14} to 10^{-12}
EC_{50} at OSNs of mice[97]	$\approx 10^{-10}$
Required for discrimination of urine by mice in the "cotton tip test" (synthetic peptides)[97]	$\approx 10^{-6}$
Used for pregnancy block experiments (synthetic peptides)[96,99]	$\approx 10^{-4}$

MHC-dependent urinary peptides could still contribute to the general pattern of relatedness sensed by a mouse without providing interpretable information about the MHC type. But the idea that, in the pregnancy block paradigm, pregnant mice just consider whether they know a given peptide from their own urine or not is also in conflict with the published experiments[96,99]. If this assumption was true, blocking of pregnancy should have occurred in all cases with peptide supplementation in Leinders-Zufall *et al.*[96] (because all peptide mixtures contained peptides not encoded in *Mus musculus*) and at least in all cases with addition of **A**APD**N**RET**F** (not known to occur anywhere in nature, see above)

in Thompson *et al.*[99]. However, the mentioned pregnancy block experiments were not designed to test the hypothesis of recognition of "familiar", i.e. occurring in the urine of the mating male, versus "unfamiliar", i.e. not occurring in the urine of the mating male, peptides. Such experiments should be performed in the future using peptides really present in urine as the "familiar" set and applying both "familiar" and "unfamiliar" peptide mixtures at physiological concentrations.

The interpretation of the pregnancy block experiments of Thompson *et al.*[99] remains a mystery to me for another reason. The synthetic peptides S**Y**FPEITH**I** and AAPD**N**RET**F** elicited the pregnancy block in an MHC haplotype specific manner even in a pure form, i.e. in the absence of urine as a background solution. However, it would be evolutionary ruinous if a pregnant female aborted the pregnancy solely by encountering an "unfamiliar" peptide having some similarity (AAPD**N**RET**F** does not even properly fit to the H2-D^b motif, see above) with an MHC peptide ligand not corresponding to the MHC haplotype of the mating male. Because, for example, the food of mice will just by chance contain some MHC motif peptides corresponding to MHC haplotypes other than those of the mating male, pregnancy should be aborted also when encountering a new type of food after insemination. This is probably not the case. To enable a meaningful interpretation of a peptide signal, a mouse should also consider the biochemical environment.

To test the behavioural relevance of MHC class I-dependent urinary peptides, selective depletion of these peptides from urine samples would be highly desirable, but I see no possibility to accomplish this. At first glance, $\beta_2m^{-/-}$ mice might appear to be an option. However, the urine of these mice will not only lack MHC class I-dependent peptides but also peptides derived from the β_2m protein. The β_2m protein is an established component of mammalian urine (see Hard[42] and Supplementary Information of Nagaraj and Mann[64]). The concentration of β_2m in human urine is approximately $5 \cdot 10^{-9}$ M, but mouse urine is generally more concentrated than human urine[42]. Therefore, β_2m-derived peptides might occur at relevant concentrations in wild type mouse urine. Accordingly, we obtained some evidence for two β_2m-derived peptides (SFSKDWSF and SKDWSFY, overlapping sequence underlined) in urine of B10.D2 mice in MS experiment no. 2 (cf. Table 2-1). However, despite of the overlapping sequence, the confidence of the identification is low for these two peptides. They were discovered at an FDR of 1% but failed the Mascot score threshold of 25 which we required for the 639 peptides reported in this study. The statistical probability for being wrong identifications (PEP value[213,242] calculated by MaxQuant) is 8.4% for SFSKDWSF (Mascot score 14) and 11.8% for SKDWSFY (Mascot score 23). Nevertheless, the probability that both of these peptides are false positive identifications is only ($0.084 \cdot 0.118$ =) 1% thereby providing a strong hint that β_2m-derived peptides are indeed present in mouse urine at relevant concentrations (*Nat Comm*, Theo).

Because selective depletion of MHC-dependent peptides from urine appears impossible, future behavioural experiments will have to test whether peptides added in picomolar concentrations to whole urine samples do have an effect. Notably, synthetic MHC peptide ligands could only be

discriminated down to a concentration of about 10^{-6} M in the cotton tip test[97]. Apparently, this level of sensitivity would make a role of some abundant urinary SAV peptides like LNSVFDQL conceivable, whereas MHC-dependent urinary peptides are clearly below the behavioural response threshold in this test (Table 4-1). Although one would intuitively assume that peptides of higher concentrations will overlay the impact of low abundant ones, this idea must be verified experimentally in different behavioural contexts. Leinders-Zufall *et al.*[98] demonstrated that VSN activation patterns did not change when the concentration of a synthetic peptide was increased from 10^{-11} M to 10^{-5} M. Thereby, the relative contribution of low abundant peptides to a behavioural outcome might be higher than estimated by the huge difference in concentration when compared to the most abundant urinary peptides. One can surmise that the extreme sensitivity of VSNs of about 10^{-14} to 10^{-12} M [98] evolved not by chance, but has a function yet to be discovered.

4.2.2. Non-MHC genes may confound behavioural experiments targeting the MHC

As detailed above, I see no conclusive possibility of how MHC-dependent urinary peptides should transmit information about the MHC type of a conspecific. However, as a matter of fact, mice are able to discriminate MHC congenic mice. This observation has led to the hypothesis that the MHC might be important for recognition of genomic identity, but the crucial question is whether the observed behavioural effects are really related to the MHC genes. MHC congenic mice differ not only in the MHC but also in the adjacent genes, and the discriminating chromosomal region comprises at least 14 Mbp in the strains B6 versus B6-H2^k (Klein *et al.*[330], Figure 4-1), which represent the most frequently used[20,63,149,167,173,331] pair of MHC congenic mice in experiments investigating MHC-related odour types and behaviour. Probably, the whole DNA segment between the MHC and the telomeric end of this arm of chromosome 17 differs between B6 and B6-H2^k mice[332]. In 1990, Jan Kleins laboratory showed that the differential segment encompasses more than one third of chromosome 17 in several MHC congenic mouse strains[332], i.e. several hundred genes.

Interestingly, as depicted in Figure 4-1, the chromosomal segment differing between B6 and B6-H2^k mice contains both the ESP gene cluster (introduced in section ”1.6.2. Recognition of genomic identity based on non-MHC genes”) and *mep1a* coding for meprin A α (see section “1.1.1. The formation of urine”). ESPs appear to be produced specifically for purposes of olfactory signalling; some are released into the tears or expressed in the submaxillary glands producing saliva[143]. ESPs have not yet been described to occur in urine, but experiments on MHC-related mating preferences allowed the mice to investigate both facial and anogenital regions of the partners[20,331]. Some ESPs occur in different alleles within laboratory inbred strains discriminating e.g. B6 and BALB/c mice[143]. Consequently, testing of MHC-related mating preferences between mice of the *H2^b* and *H2^d*

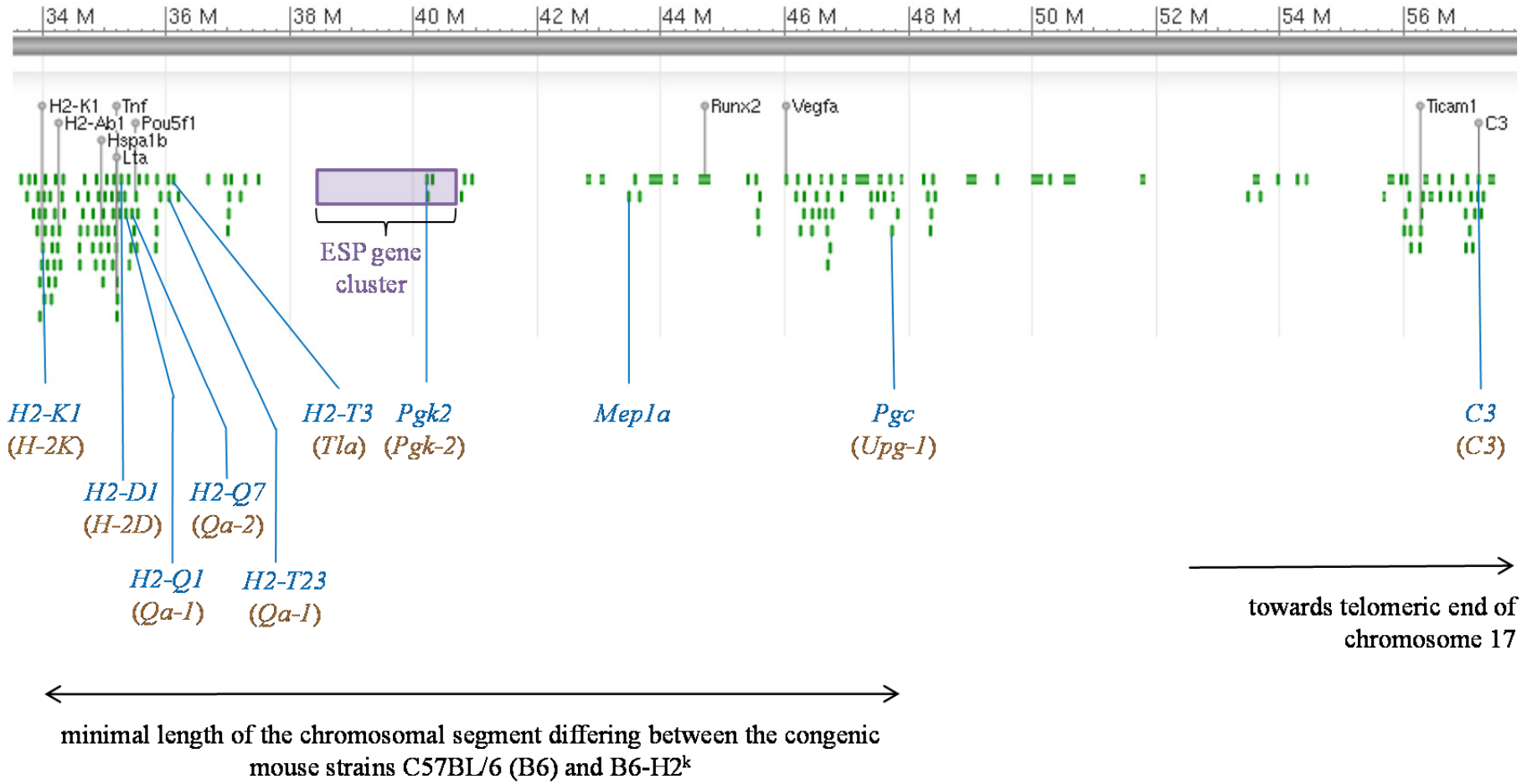


Figure 4-1 | Meprin A α and exocrine gland-secreting peptides (ESPs) are encoded in close proximity to the MHC and could, therefore, confound behavioural experiments with MHC congenic mice. A section of chromosome 17 of *Mus musculus* is depicted (based on NCBI Nucleotide Database, accession number NC_000083.6; GRCm38.p1 C57BL/6J; [252]). The top line indicates the chromosomal nucleotide positions in mega base pairs (Mbp, M). The minimal length of the chromosomal segment differing between B6 and B6-H2^k mice is given according to Figure 2 of Klein *et al.*[330]. This region of chromosome 17 also differs between many other congenic mouse strains[330]. As emphasised by Klein *et al.*[330], the real length of the differential segment is probably much longer and it was shown to comprise more than one third of chromosome 17 in several congenic lines[332]. Green bars represent genes. Some genes, including most ESP genes, were omitted by NCBI Nucleotide Database to enable a simple overview. Selected genes of interest are annotated with blue script representing contemporary reference gene names. The analyses of allelic variants in *H2* linked genes in 104 MHC congenic mouse strains[330] occasionally applied different gene names. For an easy comparison, the gene names used in Klein *et al.*[330] are indicated in brackets in brown font.

haplotypes might be influenced by ESPs. I am not aware of a study investigating differences in ESPs between B6 and AKR mice, the latter strain being the donor of the *H2*^{k} chromosomal segment in B6-H2^k mice. However, when observing mating preferences between B6 and B6-H2^k mice, such a possible difference should be controlled before assigning the effect to the MHC. In the case of meprin A α, it is clearly known that the nonfunctional *mep1a* allele coincides with the *H2*^{k} haplotype[36,310]. As detailed on page 65, this will have prominent direct and probably additional indirect effects on the urinary peptidome. The meprin A α related differences can clearly be expected to be far more dominant than those of MHC-dependent urinary peptides in the B6 versus B6-H2^k comparison, questioning the assignment of the mating preference to the MHC.

For each gene of *Mus musculus*, on average about three different SAV-coding SNVs are described (see section "3.2. The urinary peptidome comprises a respectable set of peptides with single amino acid variations (SAVs)"). Given that the chromosomal segment differing between MHC congenic mice comprises several hundred genes, it is likely that some of the SNVs of this DNA region will affect urinary peptides, again questioning the assignment of observed mating preferences to the MHC genes. Furthermore, the MHC itself contains a lot of genes apart from classical MHC genes and their variability might also contribute to observed behaviours[333].

Many experiments on MHC-related odour differences have suggested the MHC-dependent olfactory cues to be volatile ruling out peptides and proteins as the relevant molecules[63,167,172-174,334,335]. However, the basic problem of MHC-linked genes is the same as with regard to the peptides. Urinary volatiles are the result of complex metabolic pathways, and these will be affected by thousands of different genes. Given the fact that hundreds of genes are derived from different donor mice in MHC congenic mouse strains[330,332], it is very likely that some of these genes will be involved in one or several pathways affecting urinary volatiles, just by chance. The genomic diversity influencing urinary peptides (cf. section "4.3. Genotype specific peptides offer new perspectives for olfactory research") will also be reflected in the complex metabolic processes of a mouse. Apart from the many MHC-linked genes that are highly likely to differ between MHC congenic mice, all genes of a mouse are subject to continuous mutation and a substantial genetic drift has been reported even for closely related substrains of laboratory mice[77]. The latter problem also applies to experiments using MHC mutant mice instead of congenic ones[170,172,335]. Accordingly, Röck *et al.*[62] demonstrated differences in urinary volatiles for substrains of B6 mice, and these differences were more pronounced than those observed between B6 and B6/β_2m$^{-/-}$ mice derived from the same vendor, namely the Jackson Laboratories. Therefore, the mere statistical association of specific volatile profiles with the MHC type of MHC congenic mouse strains cannot validate a specific effect of the MHC genes[62,174]. There are several proposals for how MHC genes might influence urinary volatiles (see Pearse-Pratt *et al.*[321]), but a specific biochemical mechanism has not been demonstrated experimentally. In summary, no urinary volatile has been demonstrated to be causatively linked to a specific MHC allele.

Scientists early got aware of the possibility that background genes might be a confounding factor in potentially MHC-related behaviours[331], but nevertheless most of them did not stop to assign the observed effects to the MHC genes for a very long time[63,149,167,170-173,335]. Already in 1978, Yamazaki *et al.* applied F_2 segregants to reduce the impact of non-MHC genes[331]. While non-MHC linked genes are randomly distributed in F_2 animals and can therefore be ruled out as systematic confounding factors, all genes encoded on chromosome 17 (which contains the MHC in mice) are, on average, inevitably more similar between F_2 mice of the same MHC type than between MHC-dissimilar F_2 mice. Consequently, all genes encoded on chromosome 17 could potentially contribute to the differential odour profiles of MHC-dissimilar F_2 segregants and to the related behavioural effects.

It would be interesting to use congenic mice differing at other loci than the MHC for performing the same behavioural experiments that have been performed with MHC congenic mice. I guess that mice can easily be trained to discriminate the odours of such control-congenic mice. My assumption is based on the finding that even odours of different B6 substrains are distinguishable by gas chromatography (GC-) MS[62], and that the sensory capabilities of mice seem to outperform the sensitivity of MS. Note that the volatile odour of B6 mice can be discriminated from that of B6/$\beta_2m^{-/-}$ mice by conspecifics[334]. However, it proved to be difficult to achieve this discrimination in GC-MS analyses[62]. It is tempting to speculate that, apart from mere odour discrimination, assortative and disassortative mating preferences for one or the other control-congenic mouse strain might occur paralleling the results obtained with MHC congenic mice (cf. section “1.6.1. Recognition of genetic identity based on MHC type”).

In summary, there is no conclusive line of evidence linking an observed behaviour to a specific MHC-dependent molecule occurring in a natural context. It is undeniable that MHC genes and / or MHC-linked genes influence urinary odour profiles. However, the crucial question is whether mice specifically use the originating signature to make decisions related to the immunological competence, i.e. MHC configuration, of a conspecific, or whether they consider the MHC-associated effect just as one of hundreds or thousands of possibly redundant signals of genomic relatedness. This core issue is not yet solved.

4.2.3. Is there an evolutionary need for a direct MHC signal in social communication?

Some scientists suggest that mediation of kin recognition was the primordial function of the MHC in evolution, and that MHC`s function in adaptive immunity only evolved later[336,337]. From a theoretical point of view, the MHC indeed appears highly suitable for olfactory recognition of genomic relatedness because it represents the most polymorphic genomic region of vertebrates[84]. However, this diversity can be explained with immunological reasoning alone - the polymorphism is a mechanism to avoid immune evasion by pathogen adaptation to certain MHC alleles. In contrast to the MHC, MUPs really appear to have evolved for signalling reasons as this is their only known function[178]; but instead of signalling genomic relatedness themselves, I assume the primordial function of MUPs to be the enhancement of the duration of volatile release from urine scent marks facilitating signalling of sex, reproductive and social status as well as genomic identity encoded by the volatiles[180]; this assumption is based on the fact that MUPs are present in urine of most mammals but are polymorphic and polygenic only in a few species, and this polymorphic polygeny is evolutionary young[144,179]. *Mus musculus* really makes use of its MUP polymorphism to discriminate individuals[178], but, in general, it appears much more likely to me that nature takes advantage of polymorphisms that are already present for other reasons than to design molecules solely for recognition of genomic relatedness. There are more than 20,000 protein coding genes in *Mus musculus*[338], and many of them will either influence metabolic pathways that affect urinary volatiles, or they will be directly reflected by urinary peptides. Although the polymorphism of the MHC is unrivalled, the sum of the genetic variation within the several thousand other genes will easily outperform the diversity contained in the MHC eliminating any theoretical need for the MHC as a signal of genomic identity.

Most evolutionary advantageous behaviours ascribed to the MHC, e.g. cooperative behaviour among kin and between parents and their progeny, territorial scent marking, inbreeding avoidance and mate recognition in the context of the pregnancy block, do *a priori* just require a signal of genomic relatedness or individuality but do not need to match the MHC. Note that some of these behaviours also occur in invertebrates, which lack the MHC. Mate choice of jawed vertebrates is a different case because in this setting, selection of partners with certain MHC alleles really makes sense from a theoretical point of view (see section “1.6.1. Recognition of genetic identity based on MHC type”). However, the crucial question is whether this mate selection requires a specific MHC signal, or whether the required information can simply be gathered indirectly. MHC alleles which are not suitable at a given time point in a given environment are detrimental to an individual`s health. Bad health usually reduces optical features of fitness, e.g. well-kept appearance of the fur, body weight and muscle mass, and also makes urine scent less attractive to potential mating partners[10-12]. Thereby, suboptimal MHC alleles will be counterselected indirectly in mate choice or even lead to the dead of the animal before mating. Furthermore, because the MHC is highly polymorphic in most species, it is

unlikely that an individual mates with a partner sharing all MHC alleles as long as the individual avoids mating with relatives; this general inbreeding avoidance could easily be mediated by non-MHC genes. Because the frequency of MHC alleles is usually different in different subpopulations of a species, the probability that two mating partners have identical MHC alleles decreases in inter-subpopulation matings, or, more generally spoken, it decreases with decreasing genomic relatedness. A bias to mate with genetically more unrelated partners will automatically reduce the likelihood for low MHC diversity in the offspring. Therefore, although tempting at the first glance, I see no evolutionary need for a direct MHC signal in social communication of vertebrates.

4.3. Genotype specific peptides offer new perspectives for olfactory research

Peptides with MHC-independent allele-specific features occur with or without an MHC motif at urinary concentrations readily detectable by MS. Notably, MHC-independent MHC motif peptides additionally occur in N- and C-terminally extended versions also able to stimulate VSNs (Figure 3 of Sturm *et al.*[207]). Such variations of length are generally not observed for peptides loaded onto MHC class I molecules via the antigen presentation pathway because overlong peptides do not fit into the binding groove of MHC class I molecules, which is closed at both ends[90,339], and the extended forms are therefore subject to rapid intracellular degradation[119]. Apparently, many urinary peptides with MHC motif and their extended forms are likely produced pre-renally, post-renally or both by proteases such as the copious proteases of blood plasma[323] (see also section "3.3. Further genomic variations influence urinary peptides" and Table 3-9). In contrast, the detectable SIINFEKL is not produced by this mechanism because it was not found in urine from B6/OVA^{+}/β_2m$^{-/-}$ mice (*Nat Comm*, joint).

As demonstrated with a 12-mer SAV peptide pair (Figure 4 of Sturm *et al.*[207]), an MHC motif is required neither for stimulation of specific VSNs nor for their capability to differentiate between SAVs. Therefore, most if not all urinary peptides of a suitable length (comprising at least the range between 3 (ref. [98]) and 18 amino acids (Figure 3 of Sturm *et al.*[207])) may contribute to the individual peptide pattern detected by the ensemble of VSNs. However, the information content of very short unmodified peptides is supposed to be much smaller than that of longer peptides because peptides shorter than about seven amino acids are usually encoded by more than one gene making qualitative differences in unmodified urinary tri- and tetrapeptides conceivably rare. Nevertheless, I expect very short urinary peptides to contain information about individuality due to quantitative effects and variability in modifications. From a theoretical point of view, peptides of a length typical for MHC class I and short MHC class II peptide ligands (about 8 - 15 amino acids) would be optimal to signal genetic relatedness because peptides longer than seven amino acids can usually be assigned to exactly

one gene or at least a small group of homologous genes which makes some qualitative differences between mouse strains likely. On the other hand, peptides longer than about 15 amino acids are prone to form secondary structures thereby probably restricting the access of VSN receptors to some of their amino acid residues and adding a further level of complexity also depending on non-genome related factors. However, it should be highlighted that even defined groups of small proteins (exocrine gland-secreting peptides[143] and intact major urinary proteins) have been described to stimulate VSNs (*Nat Comm*, joint).

It remains to be shown whether VSNs can differentiate all kinds of SAVs, especially whether they are able to discriminate amino acids with very similar biochemical properties, e.g. leucine and isoleucine. However, SAVs are the most common type of intra- and interspecies variation of homologous proteins. The respectable frequency of diverse SAVs in urinary peptides will compensate for a possible occasional failure of VSNs to discriminate a particular SAV. In contrast, genes with functional and non-functional alleles, e.g. certain MUP genes, *mep1a* and *H2-L*, are less widespread than genes containing missense SNVs. Nevertheless, their behavioural impact might easily rival that of SAV peptides because a single case of such an allele-specific lack affects all peptides encoded by this gene and not only peptides containing the SAV position. Furthermore, the biochemical difference resulting from the loss of a complete peptide is much larger than the one resulting from a single SAV and this might translate into VSN activation patterns. Despite our failure to demonstrate nonsense (stop codon gain) mutations affecting urinary peptides applying the dbSNP, such cases probably occur and will also lead to the complete loss of certain peptides in urine. Although not addressed in the current study, other types of genomic variation are likely to additionally alter urinary peptide patterns. These variations include insertions, deletions and frameshift mutations in protein-coding regions as well as splice variations. Moreover, SNVs, insertions, deletions and frameshift mutations in non-coding regions, e.g. in the promoter, in enhancer and silencer sequences or at the polyadenylation signal sequence, may either delete protein expression completely, as discussed above, or lead to quantitative changes finally affecting urinary peptides.

Apart from peptides with differences in single amino acids, variations in urinary peptide composition can also arise from non-functional or multiplicated protein coding genes as exemplified for peptides derived from the metalloendopeptidase meprin A α or from MUPs. Therefore, SAV peptides and other MHC-independent urinary peptides contain sufficient information on genotype to enable discrimination of subspecies and sympatric species of *Mus musculus*. Genotype discrimination via these peptides could be evolutionary advantageous in a variety of processes ranging from kin and parent-progeny recognition to inbreeding avoidance to sympatric speciation. To circumvent inbreeding, a mouse needs to search for partners with a genotype differing clearly from its own and that of its relatives. On the other hand, a mouse should also avoid mating with individuals differing too strongly from its own genotype because the progeny of matings between individuals of closely related

species are usually infertile and progeny of intersubspecies matings often have a reduced fitness[306]. Importantly, SAV peptides will provide a signature of genomic individuality also in species lacking MUP polymorphism (e.g. *Mus macedonicus*) and, because peptides can additionally be detected by the main olfactory epithelium[97], they may even play a role in species lacking a vomeronasal organ. Despite all the differences in urinary peptides, one should be aware that, when comparing two given strains of *Mus musculus*, most urinary peptides are shared (e.g. our MS analyses showed that at least 70% of peptides identified in B6 mice were also present in BALB/c mice) (*Nat Comm*, joint).

The picture on the variability of urinary peptides emerging from this study is very similar to that described for urinary volatiles[62]. In natural wild populations, urinary peptide patterns will differ qualitatively and quantitatively between individuals. However, within an inbred strain, I do not expect qualitative differences but possibly some variation in peptide abundance. Qualitative as well as quantitative differences will increase with decreasing evolutionary relatedness, i.e. in the order: a given inbred strain (taken as reference) < sub-strains of this inbred strain < congenic strains (derived decades ago) < independently derived inbred strains ≤ subspecies of *Mus musculus* < different species. Given that subsets of nasal chemosensory neurons exhibit a remarkable ability to differentiate a wide variety of peptides[98,207], the urinary peptide pattern represents a real time sampling of the expressed genome available for chemosensory assessment by other individuals. MHC-dependent olfactory cues are the proverbial needle in the haystack of the urinary signalling mixture which provides plenty MHC-independent information on genomic identity via volatiles[62], peptides[207,319] and proteins[178]. After decades of tenacious searching for MHC-dependent urinary molecules, the solution might be that the haystack is more important than the searched needle (*Nat Comm*, joint).

5. Summary / Zusammenfassung

Genotype-specific urinary peptides as a molecular basis for olfactory recognition of individuality and genomic relatedness in mice

The degree of genomic relatedness influences many social behaviours in humans and animals. Mice, as nocturnal rodents, take advantage of their excellent sense of smell to identify kin, conspecifics and members of other species via body odours or urine scent marks. Selected groups of peptides, including those that are presented by major histocompatibility complex (MHC) proteins, are recognised by mouse chemosensory neurons with very high sensitivity (up to 10^{-14} M). MHC peptide ligands have been proposed to influence social interactions in many species, including mice and humans. However, the lack of knowledge about such peptides in natural sources accessible for nasal recognition has been a major barrier for this hypothesis (*Nat Comm*, joint).

Here we analyse urinary peptides from selected inbred strains of *Mus musculus* with respect to genotype-related differences. We discover many abundant peptides with single amino acid variations (SAVs) corresponding to genomic nucleotide variations. The polymorphism of major urinary proteins (MUPs) is reflected by variations in prominent urinary peptides, and further types of genomic variability are predicted to affect the urinary peptidome. Several peptides with MHC binding motifs are readily detectable in urine at concentrations of up to 10^{-6} M, but their occurrence is independent of the MHC. On the other hand, by comparing mouse strains expressing chicken ovalbumin together with, or without, functional MHC class I molecules, we provide conclusive evidence that the prototypic H2-K^b ligand SIINFEKL (derived from ovalbumin) exists in urine at a concentration of $4 \cdot 10^{-12}$ M in an MHC-dependent manner. In cooperation with neurobiologists, we show that chemoreceptive neurons in the vomeronasal organ detect and discriminate SAV peptides as well as SIINFEKL (*Nat Comm*, joint).

The present thesis fills a critical gap in the hypothesis that MHC peptide ligands may serve as signals of individuality in mice by providing the first evidence for such a ligand occurring in urine in an MHC-dependent manner. However, at the same time, this study profoundly questions the same hypothesis by showing, first, that the respective peptide concentrations in urine are about eight orders of magnitude lower than those previously applied in behavioural experiments and second, by demonstrating abundant MHC-independent MHC motif peptides. I therefore suppose that the copious peptides which reflect genomic individuality independent of MHC genes are far more important in determining social behaviours than the MHC-dependent ones. Collectively, urinary peptides represent a real-time sampling of the expressed genome available for chemosensory assessment by other individuals (*Nat Comm*, joint).

Genotyp-spezifische Peptide im Harn als molekulare Grundlage für die olfaktorische Erkennung der Individualität und des genetischen Verwandtschaftsgrades bei Mäusen

Der genetische Verwandtschaftsgrad wirkt sich auf viele soziale Verhaltensweisen von Mensch und Tier aus. Als nachtaktive Nagetiere nutzen Mäuse ihren hervorragenden Geruchsinn, um Familienangehörige, Artgenossen und Vertreter fremder Arten zu identifizieren. Die benötigten Informationen können sie dabei auch aus Urinduftmarken gewinnen. Chemosensorische Nervenzellen der Maus erkennen bestimmte Peptidsorten, einschließlich denjenigen, die von den Proteinen des Haupthistokompatibilitätskomplexes (*major histocompatibility complex*, MHC) präsentiert werden, mit sehr hohen Sensitivitäten von bis zu 10^{-14} M. Es wurde postuliert, dass MHC-Peptidliganden soziale Beziehungen bei vielen verschiedenen Arten, einschließlich Maus und Mensch, beeinflussen. Ein grundlegendes Problem bei dieser Hypothese ist jedoch das fehlende Wissen über solche Peptide in Duftquellen, die natürlicherweise für die Geruchserkennung zur Verfügung stehen.

Hier untersuchen wir Urinpeptide von ausgewählten Inzuchtstämmen von *Mus musculus* in Bezug auf Unterschiede, die mit dem Genotyp zusammenhängen. Dabei entdecken wir viele häufige Peptide mit Einzelaminosäurevariationen (EAVs), welche genomischen Nukleotidvariationen entsprechen. Der Polymorphismus der *Major Urinary* Proteine (MUPs) spiegelt sich in Unterschieden in prominenten Urinpeptiden wider, und es kann vorhergesagt werden, dass weitere Arten genomischer Vielfalt das Urinpeptidom beeinflussen. Mehrere Peptide mit MHC-Bindemotiven sind im Harn in Konzentrationen von $\leq 10^{-6}$ M leicht nachzuweisen, treten aber unabhängig vom MHC auf. Andererseits können wir einen schlüssigen Nachweis liefern, dass der prototypische H2-K^b-Ligand SIINFEKL, welcher aus Hühner-Ovalbumin stammt, MHC-abhängig mit einer Konzentration von $4 \cdot 10^{-12}$ M im Harn Ovalbumin-transgener Mäuse vorkommt. Der Nachweis der MHC-Abhängigkeit erfolgt dabei unter Nutzung eines Mausstammes, der Ovalbumin, aber keine funktionellen MHC-Klasse-I-Moleküle exprimiert. In Zusammenarbeit mit Neurobiologen zeigen wir, dass chemosensorische Nervenzellen im Vomeronasalorgan sowohl EAV-Peptide als auch SIINFEKL erkennen und unterscheiden können.

Die vorliegende Doktorarbeit schließt eine entscheidende Lücke in der Hypothese, dass MHC-Peptidliganden als Signale der Individualität von Mäusen dienen können, indem sie den ersten Nachweis für einen solchen Liganden erbringt, der MHC-abhängig im Harn auftritt. Gleichzeitig stellt unsere Studie dieselbe Hypothese jedoch grundlegend in Frage, weil sie erstens zeigt, dass die entsprechenden Peptidkonzentrationen im Urin etwa acht Größenordnungen unter denjenigen liegen, welche in vorangegangenen Verhaltensexperimenten verwendet wurden und weil sie zweitens häufige MHC-unabhängige MHC-Motivpeptide nachweist. Ich vermute daher, dass die zahlreichen Peptide, welche die genomische Individualität unabhängig von MHC-Genen widerspiegeln, für das Sozialverhalten weitaus wichtiger sind als die MHC-abhängigen. In ihrer Gesamtheit bilden Urinpeptide Teile des exprimierten Genoms in Echtzeit ab und machen es somit für die chemosensorische Analyse durch andere Individuen zugänglich.

6. Abbreviations

Abbreviation	Full name and / or explanation
AC	alternating current
AOB	accessory olfactory bulb
APC	antigen presenting cell
B6	C57BL/6 (mouse strain)
B6/OVA^{+}	C57BL/6-Tg(ACTB-OVA)916Jen (mouse strain)
B6/β_2m$^{-/-}$	B6.129P2-*B2m*^{tm1Unc}/J (mouse strain)
B6/OVA^{+}/β_2m$^{-/-}$	cross of B6/OVA^{+} and B6/β_2m$^{-/-}$ mice; mice express artificial ovalbumin protein construct and are homozygous carriers of the *B2m*^{tm1Unc} mutation
B10	C57BL/10Sn (mouse strain)
B10.D2	B10.D2-*Hc*^{1} *H2*^{d} *H2-T18*^{c}/nSnJ (mouse strain)
β_2m	β_2-microglobulin
BSA	bovine serum albumin
CID	collision-induced dissociation
CTL	cytotoxic T lymphocyte
DAG	diacylglycerol
DC	direct current
(Dipl)	see note in section “7. References”
e	extracellular
EC	Enzyme Commission
EGF	epidermal growth factor
ELISA	enzyme-linked immunosorbent assay
ER	endoplasmic reticulum
ESI	electrospray ionisation
ESP	exocrine gland-secreting peptide
exp.	experiment
FITC	fluorescein isothiocyanate
FT	Fourier transformation
GC	gas chromatography
HCD	higher energy collisional dissociation
hetero	heterozygous
homo	homozygous
HPLC	high performance liquid chromatography
i	intracellular
k	constant specific for the orbitrap
LC	liquid chromatography
LOD	limit of detection
LTQ	linear trap quadrupole
m	mass
Δm	mass deviation = theoretical mass − measured mass
mem	plasma membrane
MALDI	matrix-assisted laser desorption ionisation
MFI	mean fluorescence intensity
MHC	major histocompatibility complex
MOE	main olfactory epithelium
MS	mass spectrometry

MS spectrum	mass spectrum obtained without targeted fragmentation of ions
MS^2 spectrum	mass spectrum obtained after first targeted fragmentation of ions
MUP	mature urinary protein
MWCO	molecular weight cut-off
n.a.	not analysed
(*Nat Comm*, joint)	see note in section "7. References"
(*Nat Comm*, Theo)	see note in section "7. References"
ω	frequency of oscillations in axial direction in the orbitrap
OR	olfactory receptor
OSN	olfactory sensory neuron
PBS	phosphate-buffered saline
PerCP	peridinin chlorophyll
Q-TOF	quadrupole-time of flight
ref.	reference
RF	radiofrequency
SAV	single amino acid variation
s.d.	standard deviation
SNV	single nucleotide variation
SV	structural variant
TAP	transporter associated with antigen processing
TFA	trifluoroacetic acid
Th	Thomson = Dalton / elementary charge
VNO	vomeronasal organ
VSN	vomeronasal sensory neuron
z	number of charges of an ion

7. References

Two references used in this PhD thesis are cited very frequently and make up the fundament of this work. The first one is my diploma thesis which focused on peptides with MHC-class I binding motif in mouse urine, a topic that I elaborated during my PhD time. Secondly, we published the key results of the present work in *Nature Communications*[207]. Because I am not only the first author of that article but also wrote major parts of the main text as well as virtually the whole Supplementary Information and additionally created all tables and figures (except Figures 2 − 4 of that article[207]), it was obvious to use large parts of the publication for the present PhD thesis. To designate these circumstances, I introduced special tags in my PhD thesis, explained in the following.

(Dipl) The preceding paragraph is inspired by my German language diploma thesis[203].

(*Nat Comm*, Theo) At least one of the sentences in the preceding paragraph was directly taken from Sturm *et al.*[207]. All sentences were originally written by me. Tables and figures labelled with this tag are also based on Sturm *et al.*[207] but were fully created by me (exception: minor improvements in formatting of Figure 3-4 were performed by *Nature Communications*).

(*Nat Comm*, joint) At least one of the sentences in the preceding paragraph was directly taken from Sturm *et al.*[207]. The corresponding sentences were originally written by Frank Zufall, Peter Overath, Hans-Georg Rammensee and me, with a variable degree of individual contributions.

Some sentences of this thesis were directly taken from our response to the reviewers of *Nature Communication*. However, these sentences were originally written by me, and they have not been published yet. Therefore, they are not specifically labelled.

List of references

1. Wyatt, T. D. Pheromones and signature mixtures: defining species-wide signals and variable cues for identity in both invertebrates and vertebrates. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* **196**, 685-700 (2010).
2. Brennan, P. A. & Kendrick, K. M. Mammalian social odours: attraction and individual recognition. *Philos Trans R Soc Lond B Biol Sci* **361**, 2061-2078 (2006).
3. Logothetis, N. K. & Sheinberg, D. L. Visual object recognition. *Annu Rev Neurosci* **19**, 577-621 (1996).
4. Tirindelli, R., Dibattista, M., Pifferi, S. & Menini, A. From pheromones to behavior. *Physiol Rev* **89**, 921-956 (2009).
5. Wyatt, T. *Pheromones and animal behaviour*. eds (Cambridge University Press, Cambridge, UK, 2003).
6. He, J., Ma, L., Kim, S., Nakai, J. & Yu, C. R. Encoding gender and individual information in the mouse vomeronasal organ. *Science* **320**, 535-538 (2008).
7. He, J. *et al.* Distinct signals conveyed by pheromone concentrations to the mouse vomeronasal organ. *J Neurosci* **30**, 7473-7483 (2010).
8. Roberts, S. A. *et al.* Darcin: a male pheromone that stimulates female memory and sexual attraction to an individual male's odour. *BMC Biol* **8**, 75 (2010).

9. Jones, R. B. & Nowell, N. W. A comparison of the aversive and female attractant properties of urine from dominant and subordinate male mice. *Animal learning & behavior* **2**, 141-144 (1974).
10. Ehman, K. D. & Scott, M. E. Urinary odour preferences of MHCcongenic female mice, Mus domesticus: implications for kin recognition and detection of parasitized males. *Animal behaviour* **62**, 781-789 (2001).
11. Penn, D., Schneider, G., White, K., Slev, P. & Potts, W. Influenza infection neutralizes the attractiveness of male odour to female mice (Mus musculus). *Ethology* **104**, 685–694 (1998).
12. Zala, S. M., Potts, W. K. & Penn, D. J. Scent-marking displays provide honest signals of health and infection. *Behav Ecol* **15**, 338–344 (2004).
13. Zalaquett, C. & Thiessen, D. The effects of odors from stressed mice on conspecific behavior. *Physiol Behav* **50**, 221-227 (1991).
14. Nodari, F. *et al.* Sulfated steroids as natural ligands of mouse pheromone-sensing neurons. *J Neurosci* **28**, 6407-6418 (2008).
15. Hurst, J. L. Female recognition and assessment of males through scent. *Behav Brain Res* **200**, 295-303 (2009).
16. Cheetham, S. A. *et al.* The genetic basis of individual-recognition signals in the mouse. *Curr Biol* **17**, 1771-1777 (2007).
17. Rich, T. J. & Hurst, J. L. Scent marks as reliable signals of the competitive ability of mates. *Animal behaviour* **56**, 727-735 (1998).
18. Rich, T. J. & Hurst, J. L. The competing countermarks hypothesis: reliable assessment of competitive ability by potential mates. *Animal behaviour* **58**, 1027-1037 (1999).
19. Chamero, P. *et al.* Identification of protein pheromones that promote aggressive behaviour. *Nature* **450**, 899-902 (2007).
20. Yamazaki, K. *et al.* Control of mating preferences in mice by genes in the major histocompatibility complex. *J Exp Med* **144**, 1324-1335 (1976).
21. Loconto, J. *et al.* Functional expression of murine V2R pheromone receptors involves selective association with the M10 and M1 families of MHC class Ib molecules. *Cell* **112**, 607-618 (2003).
22. Haraldsson, B. & Sorensson, J. Why do we not all have proteinuria? An update of our current understanding of the glomerular barrier. *News Physiol Sci* **19**, 7-10 (2004).
23. Alberts, B. *et al. Molecular Biology of the Cell (5th edition)*. (Garland Science, New York and Abingdon, 2008).
24. Maack, T., Park, C. H. & Camargo, M. J. F. *Renal filtration, transport, and metabolism of proteins. In The Kidney: Physiology and Pathophysiology.* 1773-1803, eds (Raven Press, New York, 1985).
25. Cui, S., Verroust, P. J., Moestrup, S. K. & Christensen, E. I. Megalin/gp330 mediates uptake of albumin in renal proximal tubule. *Am J Physiol* **271**, F900-907 (1996).
26. Birn, H. *et al.* Cubilin is an albumin binding protein important for renal tubular albumin reabsorption. *J Clin Invest* **105**, 1353-1361 (2000).
27. Batuman, V. *et al.* Myeloma light chains are ligands for cubilin (gp280). *Am J Physiol* **275**, F246-254 (1998).
28. Gburek, J. *et al.* Megalin and cubilin are endocytic receptors involved in renal clearance of hemoglobin. *Journal of the American Society of Nephrology : JASN* **13**, 423-430 (2002).
29. Gburek, J. *et al.* Renal uptake of myoglobin is mediated by the endocytic receptors megalin and cubilin. *American journal of physiology. Renal physiology* **285**, F451-458 (2003).
30. Christensen, E. I. & Gburek, J. Protein reabsorption in renal proximal tubule-function and dysfunction in kidney pathophysiology. *Pediatric nephrology* **19**, 714-721 (2004).
31. Rubio-Aliaga, I. & Daniel, H. Peptide transporters and their roles in physiological processes and drug disposition. *Xenobiotica; the fate of foreign compounds in biological systems* **38**, 1022-1042 (2008).
32. Wehner, R. & Gehring, W. *Zoologie*. 23. edn, eds (Georg Thieme Verlag, Stuttgart, New York, 1995).
33. Abrink, M., Larsson, E., Gobl, A. & Hellman, L. Expression of lactoferrin in the kidney: implications for innate immunity and iron metabolism. *Kidney Int* **57**, 2004-2010 (2000).
34. Vorland, L. H. Lactoferrin: a multifunctional glycoprotein. *APMIS* **107**, 971-981 (1999).
35. *BRENDA - The Comprehensive Enzyme Information System*, <http://www.brenda-enzymes.org> (2013).
36. Beynon, R. J., Oliver, S. & Robertson, D. H. Characterization of the soluble, secreted form of urinary meprin. *Biochem J* **315**, 461-465 (1996).
37. Flannery, A. V., Dalzell, G. N. & Beynon, R. J. Proteolytic activity in mouse urine: relationship to the kidney metallo-endopeptidase, meprin. *Biochimica et biophysica acta* **1041**, 64-70 (1990).
38. Bond, J. S., Matters, G. L., Banerjee, S. & Dusheck, R. E. Meprin metalloprotease expression and regulation in kidney, intestine, urinary tract infections and cancer. *FEBS letters* **579**, 3317-3322 (2005).
39. Bertenshaw, G. P. *et al.* Marked differences between metalloproteases meprin A and B in substrate and peptide bond specificity. *J Biol Chem* **276**, 13248-13255 (2001).
40. Bylander, J. E., Bertenshaw, G. P., Matters, G. L., Hubbard, S. J. & Bond, J. S. Human and mouse homo-oligomeric meprin A metalloendopeptidase: substrate and inhibitor specificities. *Biological chemistry* **388**, 1163-1172 (2007).
41. Olson, M. J., Johnson, J. T. & Reidy, C. A. A comparison of male rat and human urinary proteins: implications for human resistance to hyaline droplet nephropathy. *Toxicol Appl Pharmacol* **102**, 524-536 (1990).
42. Hard, G. C. Species comparison of the content and composition of urinary proteins. *Food Chem Toxicol* **33**, 731-746 (1995).
43. Lehman-McKeeman, L. D. & Caudill, D. Biochemical basis for mouse resistance to hyaline droplet nephropathy: lack of relevance of the alpha 2u-globulin protein superfamily in this male rat-specific syndrome. *Toxicol Appl Pharmacol* **112**, 214-221 (1992).

44. Humphries, R. E., Robertson, D. H., Beynon, R. J. & Hurst, J. L. Unravelling the chemical basis of competitive scent marking in house mice. *Animal behaviour* **58**, 1177-1190 (1999).
45. Beynon, R. J. & Hurst, J. L. Urinary proteins and the modulation of chemical scents in mice and rats. *Peptides* **25**, 1553-1563 (2004).
46. Bates, J. M. *et al.* Tamm-Horsfall protein knockout mice are more prone to urinary tract infection: rapid communication. *Kidney Int* **65**, 791-797 (2004).
47. Mo, L. *et al.* Ablation of the Tamm-Horsfall protein gene increases susceptibility of mice to bladder colonization by type 1-fimbriated Escherichia coli. *American journal of physiology. Renal physiology* **286**, F795-802 (2004).
48. Pak, J., Pu, Y., Zhang, Z. T., Hasty, D. L. & Wu, X. R. Tamm-Horsfall protein binds to type 1 fimbriated Escherichia coli and prevents E. coli from binding to uroplakin Ia and Ib receptors. *J Biol Chem* **276**, 9924-9930 (2001).
49. Su, S. J., Chang, K. L., Lin, T. M., Huang, Y. H. & Yeh, T. M. Uromodulin and Tamm-Horsfall protein induce human monocytes to secrete TNF and express tissue factor. *J Immunol* **158**, 3449-3456 (1997).
50. Saemann, M. D. *et al.* Tamm-Horsfall glycoprotein links innate immune cell activation with adaptive immunity via a Toll-like receptor-4-dependent mechanism. *J Clin Invest* **115**, 468-475 (2005).
51. Saemann, M. D., Weichhart, T., Horl, W. H. & Zlabinger, G. J. Tamm-Horsfall protein: a multilayered defence molecule against urinary tract infection. *Eur J Clin Invest* **35**, 227-235 (2005).
52. Kuranami, M. *et al.* Effect of urine on clonal growth of human bladder cancer cell lines. *Cancer Res* **51**, 4631-4635 (1991).
53. Olsen, P. S., Nexo, E., Poulsen, S. S., Hansen, H. F. & Kirkegaard, P. Renal origin of rat urinary epidermal growth factor. *Regulatory Peptides* **10**, 37-45 (1984).
54. Perheentupa, J., Lakshmanan, J. & Fisher, D. A. Epidermal growth factor in mouse urine: non-blood origin, and increase by sialoadenectomy and T4 therapy. *Acta Endocrinol (Copenh)* **108**, 428-432 (1985).
55. Perheentupa, J., Lakshmanan, J. & Fisher, D. A. Urine and kidney epidermal growth factor: ontogeny and sex difference in the mouse. *Pediatr Res* **19**, 428-432 (1985).
56. Knepper, M. A. & Pisitkun, T. Exosomes in urine: who would have thought...? *Kidney Int* **72**, 1043-1045 (2007).
57. Frommberger, M. *et al.* Peptidomic analysis of rat urine using capillary electrophoresis coupled to mass spectrometry. *Proteomics Clin Appl* **1**, 650-660 (2007).
58. Jurgens, M. *et al.* Towards characterization of the human urinary peptidome. *Comb Chem High Throughput Screen* **8**, 757-765 (2005).
59. Selevsek, N., Matondo, M., Sanchez Carbayo, M., Aebersold, R. & Domon, B. Systematic quantification of peptides/proteins in urine using selected reaction monitoring. *Proteomics* **11**, 1135-1147 (2011).
60. Voet, D. & Voet, J. G. *Biochemistry*. 3rd edn, 1023, eds (John Wiley & Sons, Inc., Hoboken, 2004).
61. Röck, F., Mueller, S., Weimar, U., Rammensee, H. G. & Overath, P. Comparative analysis of volatile constituents from mice and their urine. *J Chem Ecol* **32**, 1333-1346 (2006).
62. Röck, F., Hadeler, K. P., Rammensee, H. G. & Overath, P. Quantitative analysis of mouse urine volatiles: in search of MHC-dependent differences. *PLoS ONE* **2**, e429 (2007).
63. Willse, A. *et al.* Identification of major histocompatibility complex-regulated body odorants by statistical analysis of a comparative gas chromatography/mass spectrometry experiment. *Anal Chem* **77**, 2348-2361 (2005).
64. Nagaraj, N. & Mann, M. Quantitative analysis of the intra- and inter-individual variability of the normal urinary proteome. *J Proteome Res* **10**, 637-645 (2011).
65. Cho, Y. H., Kim, D., Choi, I. & Bae, K. Identification of transcriptional regulatory elements required for the Mup2 expression in circadian clock mutant mice. *Biochem Biophys Res Commun* **410**, 834-840 (2011).
66. Leinders-Zufall, T. *et al.* Ultrasensitive pheromone detection by mammalian vomeronasal neurons. *Nature* **405**, 792-796 (2000).
67. Harvey, S., Jemiolo, B. & Novotny, M. Pattern of volatile compounds in dominant and subordinate male-mouse urine. *J Chem Ecol* **15**, 2061-2072 (1989).
68. Müller-Velten, H. G. Über den Angstgeruch bei der Hausmaus. *Z. Verge. Physiol.* **52**, 401-429 (1966).
69. Kavaliers, M., Choleris, E. & Pfaff, D. W. Recognition and avoidance of the odors of parasitized conspecifics and predators: differential genomic correlates. *Neurosci Biobehav Rev* **29**, 1347-1359 (2005).
70. Zasloff, M. Antimicrobial peptides, innate immunity, and the normally sterile urinary tract. *Journal of the American Society of Nephrology : JASN* **18**, 2810-2816 (2007).
71. Chu, B. C. *et al.* Siderophore uptake in bacteria and the battle for iron with the host; a bird's eye view. *Biometals* **23**, 601-611 (2010).
72. Goetz, D. H. *et al.* The neutrophil lipocalin NGAL is a bacteriostatic agent that interferes with siderophore-mediated iron acquisition. *Mol Cell* **10**, 1033-1043 (2002).
73. Flo, T. H. *et al.* Lipocalin 2 mediates an innate immune response to bacterial infection by sequestrating iron. *Nature* **432**, 917-921 (2004).
74. Yilmaz, A. *et al.* Early prediction of urinary tract infection with urinary neutrophil gelatinase associated lipocalin. *Pediatr Nephrol* **24**, 2387-2392 (2009).
75. Chromek, M. *et al.* The antimicrobial peptide cathelicidin protects the urinary tract against invasive bacterial infection. *Nat Med* **12**, 636-641 (2006).
76. Vyoral, D. & Petrak, J. Hepcidin: a direct link between iron metabolism and immunity. *Int J Biochem Cell Biol* **37**, 1768-1773 (2005).
77. Simpson, E. M. *et al.* Genetic variation among 129 substrains and its importance for targeted mutagenesis in mice. *Nat Genet* **16**, 19-27 (1997).

78. Hasin-Brumshtein, Y., Lancet, D. & Olender, T. Human olfaction: from genomic variation to phenotypic diversity. *Trends Genet* **25**, 178-184 (2009).
79. Feuk, L., Carson, A. R. & Scherer, S. W. Structural variation in the human genome. *Nat Rev Genet* **7**, 85-97 (2006).
80. Keane, T. M. *et al.* Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature* **477**, 289-294 (2011).
81. Strachan, T. & Read, A. P. *Molekulare Humangenetik*. 3 edn, eds (Elsevier GmbH, Munich, 2005).
82. Chen, J. M., Cooper, D. N., Chuzhanova, N., Ferec, C. & Patrinos, G. P. Gene conversion: mechanisms, evolution and human disease. *Nat Rev Genet* **8**, 762-775 (2007).
83. Church, D. M. *et al.* Lineage-specific biology revealed by a finished genome assembly of the mouse. *PLoS Biol* **7**, e1000112 (2009).
84. Piertney, S. B. & Oliver, M. K. The evolutionary ecology of the major histocompatibility complex. *Heredity* **96**, 7-21 (2006).
85. Murken, J., Grimm, T. & Holinski-Feder, E. *Taschenlehrbuch Humangenetik*. 7 edn, eds (Georg Thieme Verlag, Stuttgart, 2006).
86. Collins, D. W. & Jukes, T. H. Rates of transition and transversion in coding sequences since the human-rodent divergence. *Genomics* **20**, 386-396 (1994).
87. Mühlemann, O. Recognition of nonsense mRNA: towards a unified model. *Biochem Soc Trans* **36**, 497-501 (2008).
88. Klein, J. The major histocompatibility complex of the mouse. *Science* **203**, 516-521 (1979).
89. Murphy, K. *Janeway`s Immunobiology*. 8th edn, eds (Garland Science, New York and Abingdon, 2012).
90. Falk, K., Rötzschke, O. & Rammensee, H. G. Cellular peptide composition governed by major histocompatibility complex class I molecules. *Nature* **348**, 248-251 (1990).
91. Falk, K., Rötzschke, O., Stevanović, S., Jung, G. & Rammensee, H. G. Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. *Nature* **351**, 290-296 (1991).
92. Rammensee, H. G., Bachmann, J. & Stefanović, S. *MHC Ligands and Peptide Motifs*. eds (Landes Bioscience, Georgetown, TX, 1997).
93. *SYFPEITHI*, <http://www.syfpeithi.de> (2012).
94. Rammensee, H., Bachmann, J., Emmerich, N. P., Bachor, O. A. & Stevanovic, S. SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics* **50**, 213-219 (1999).
95. Boehm, T. & Zufall, F. MHC peptides and the sensory evaluation of genotype. *Trends Neurosci* **29**, 100-107 (2006).
96. Leinders-Zufall, T. *et al.* MHC class I peptides as chemosensory signals in the vomeronasal organ. *Science* **306**, 1033-1037 (2004).
97. Spehr, M. *et al.* Essential role of the main olfactory system in social recognition of major histocompatibility complex peptide ligands. *J Neurosci* **26**, 1961-1970 (2006).
98. Leinders-Zufall, T., Ishii, T., Mombaerts, P., Zufall, F. & Boehm, T. Structural requirements for the activation of vomeronasal sensory neurons by MHC peptides. *Nat Neurosci* **12**, 1551-1558 (2009).
99. Thompson, R. N., McMillon, R., Napier, A. & Wekesa, K. S. Pregnancy block by MHC class I peptides is mediated via the production of inositol 1,4,5-trisphosphate in the mouse vomeronasal organ. *J Exp Biol* **210**, 1406-1412 (2007).
100. Milinski, M. *et al.* Mate choice decisions of stickleback females predictably modified by MHC peptide ligands. *Proc Natl Acad Sci U S A* **102**, 4414-4418 (2005).
101. Janeway, C. A., Travers, P., Walport, M. & Shlomchik, M. J. *Immunobiology*. 6 edn, 183, eds (Garland Science, 2005).
102. Bouvier, M. & Wiley, D. C. Importance of peptide amino and carboxyl termini to the stability of MHC class I molecules. *Science* **265**, 398-402 (1994).
103. Rammensee, H. G., Friede, T. & Stevanovic, S. MHC ligands and peptide motifs: first listing. *Immunogenetics* **41**, 178-228 (1995).
104. Ruppert, J. *et al.* Prominent role of secondary anchor residues in peptide binding to HLA-A2.1 molecules. *Cell* **74**, 929-937 (1993).
105. Fremont, D. H., Matsumura, M., Stura, E. A., Peterson, P. A. & Wilson, I. A. Crystal structures of two viral peptides in complex with murine MHC class I H-2Kb. *Science* **257**, 919-927 (1992).
106. Matsumura, M., Saito, Y., Jackson, M. R., Song, E. S. & Peterson, P. A. In vitro peptide binding to soluble empty class I major histocompatibility complex molecules isolated from transfected Drosophila melanogaster cells. *J Biol Chem* **267**, 23589-23595 (1992).
107. Collins, E. J., Garboczi, D. N. & Wiley, D. C. Three-dimensional structure of a peptide extending from one end of a class I MHC binding site. *Nature* **371**, 626-629 (1994).
108. *UniProt*, <http://www.uniprot.org> (2011 - 2013).
109. PyMOL. (2009).
110. *Protein Data Base*, <www.rcsb.org/pdb> (2009).
111. Rudolph, M. G. & Wilson, I. A. (RCSB Protein Data Base, 2003).
112. Rodgers, J. R. & Cook, R. G. MHC class Ib molecules bridge innate and acquired immunity. *Nat Rev Immunol* **5**, 459-471 (2005).
113. Rötzschke, O. *et al.* Qa-2 molecules are peptide receptors of higher stringency than ordinary class I molecules. *Nature* **361**, 642-644 (1993).
114. Shawar, S. M., Vyas, J. M., Rodgers, J. R. & Rich, R. R. Antigen presentation by major histocompatibility complex class I-B molecules. *Annu Rev Immunol* **12**, 839-880 (1994).
115. Olson, R., Dulac, C. & Bjorkman, P. J. MHC homologs in the nervous system--they haven't lost their groove. *Curr Opin Neurobiol* **16**, 351-357 (2006).

116. Olson, R., Huey-Tubman, K. E., Dulac, C. & Bjorkman, P. J. Structure of a pheromone receptor-associated MHC molecule with an open and empty groove. *PLoS Biol* **3**, e257 (2005).
117. Schubert, U. *et al.* Rapid degradation of a large fraction of newly synthesized proteins by proteasomes. *Nature* **404**, 770-774 (2000).
118. Yewdell, J. W., Schubert, U. & Bennink, J. R. At the crossroads of cell biology and immunology: DRiPs and other sources of peptide ligands for MHC class I molecules. *J Cell Sci* **114**, 845-851 (2001).
119. Yewdell, J. W., Reits, E. & Neefjes, J. Making sense of mass destruction: quantitating MHC class I antigen presentation. *Nat Rev Immunol* **3**, 952-961 (2003).
120. Rock, K. L., York, I. A. & Goldberg, A. L. Post-proteasomal antigen processing for major histocompatibility complex class I presentation. *Nat Immunol* **5**, 670-677 (2004).
121. Ackerman, A. L., Kyritsis, C., Tampe, R. & Cresswell, P. Access of soluble antigens to the endoplasmic reticulum can explain cross-presentation by dendritic cells. *Nat Immunol* **6**, 107-113 (2005).
122. Ackerman, A. L., Giodini, A. & Cresswell, P. A role for the endoplasmic reticulum protein retrotranslocation machinery during crosspresentation by dendritic cells. *Immunity* **25**, 607-617 (2006).
123. Gromme, M. *et al.* Recycling MHC class I molecules and endosomal peptide loading. *Proc Natl Acad Sci USA* **96**, 10326-10331 (1999).
124. Kleijmeer, M. J. *et al.* Antigen loading of MHC class I molecules in the endocytic tract. *Traffic* **2**, 124-137 (2001).
125. Spehr, M. *et al.* Parallel processing of social signals by the mammalian main and accessory olfactory systems. *Cell Mol Life Sci* **63**, 1476-1484 (2006).
126. Riviere, S., Challet, L., Fluegge, D., Spehr, M. & Rodriguez, I. Formyl peptide receptor-like proteins are a novel family of vomeronasal chemosensors. *Nature* **459**, 574-577 (2009).
127. Liberles, S. D. *et al.* Formyl peptide receptors are candidate chemosensory receptors in the vomeronasal organ. *Proc Natl Acad Sci U S A* **106**, 9842-9847 (2009).
128. Papes, F., Logan, D. W. & Stowers, L. The vomeronasal organ mediates interspecies defensive behaviors through detection of protein pheromone homologs. *Cell* **141**, 692-703 (2010).
129. Karlson, P. & Luscher, M. Pheromones': a new term for a class of biologically active substances. *Nature* **183**, 55-56 (1959).
130. Roberts, S. A., Davidson, A. J., McLean, L., Beynon, R. J. & Hurst, J. L. Pheromonal induction of spatial learning in mice. *Science* **338**, 1462-1465 (2012).
131. Chamero, P., Leinders-Zufall, T. & Zufall, F. From genes to social communication: molecular sensing by the vomeronasal organ. *Trends Neurosci* **35**, 597-606 (2012).
132. Elias, J. E. & Gygi, S. P. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods* **4**, 207-214 (2007).
133. Apfelbach, R., Blanchard, C. D., Blanchard, R. J., Hayes, R. A. & McGregor, I. S. The effects of predator odors in mammalian prey species: a review of field and laboratory studies. *Neurosci Biobehav Rev* **29**, 1123-1144 (2005).
134. Tian, H. & Ma, M. Molecular organization of the olfactory septal organ. *J Neurosci* **24**, 8383-8390 (2004).
135. Fuss, S. H., Omura, M. & Mombaerts, P. The Grueneberg ganglion of the mouse projects axons to glomeruli in the olfactory bulb. *Eur J Neurosci* **22**, 2649-2654 (2005).
136. Storan, M. J. & Key, B. Septal organ of Gruneberg is part of the olfactory system. *J Comp Neurol* **494**, 834-844 (2006).
137. Breer, H., Fleischer, J. & Strotmann, J. The sense of smell: multiple olfactory subsystems. *Cell Mol Life Sci* **63**, 1465-1475 (2006).
138. Zhang, X. & Firestein, S. The olfactory receptor gene superfamily of the mouse. *Nat Neurosci* **5**, 124-133 (2002).
139. Brennan, P. A. Outstanding issues surrounding vomeronasal mechanisms of pregnancy block and individual recognition in mice. *Behav Brain Res* **200**, 287-294 (2009).
140. Rodriguez, I., Del Punta, K., Rothman, A., Ishii, T. & Mombaerts, P. Multiple new and isolated families within the mouse superfamily of V1r vomeronasal receptors. *Nat Neurosci* **5**, 134-140 (2002).
141. Haga, S. *et al.* The male mouse pheromone ESP1 enhances female sexual receptive behaviour through a specific vomeronasal receptor. *Nature* **466**, 118-122 (2010).
142. Wynn, E. H., Sanchez-Andrade, G., Carss, K. J. & Logan, D. W. Genomic variation in the vomeronasal receptor gene repertoires of inbred mice. *BMC Genomics* **13**, 415 (2012).
143. Kimoto, H. *et al.* Sex- and strain-specific expression and vomeronasal activity of mouse ESP family peptides. *Curr Biol* **17**, 1879-1884 (2007).
144. Ruff, J. S., Nelson, A. C., Kubinak, J. L. & Potts, W. K. MHC signaling during social communication. *Adv Exp Med Biol* **738**, 290-313 (2012).
145. Bruce, H. M. An exteroceptive block to pregnancy in the mouse. *Nature* **184**, 105 (1959).
146. Bruce, H. M. A block to pregnancy in the mouse caused by proximity of strange males. *J Reprod Fertil* **1**, 96-103 (1960).
147. Thomas, L. in *4th International Convocation on Immunology.* (eds E. Neter and F. Milgrom) 2-11 (Karger, 1975).
148. Klein, J. *Natural History of the Major Histocompatibility Complex*. (Wiley, New York, 1986).
149. Yamazaki, K. *et al.* Recognition of H-2 types in relation to the blocking of pregnancy in mice. *Science* **221**, 186-188 (1983).
150. Sommer, S. Major histocompatibility complex and mate choice in a monogamous rodent. *Behav Ecol Sociobiol* **58**, 181-189 (2005).
151. Bos, D. H., Williams, R. N., Gopurenko, D., Bulut, Z. & DeWoody, J. A. Condition-dependent mate choice and a reproductive disadvantage for MHC-divergent male tiger salamanders. *Mol Ecol* **18**, 3307-3315 (2009).

152. Paterson, S. & Pemberton, J. M. No evidence for major histocompatibility complex-dependent mating patterns in a free-living ruminant population. *Proc Biol Sci* **264**, 1813-1819 (1997).
153. Ober, C. *et al.* HLA and mate choice in humans. *Am J Hum Genet* **61**, 497-504 (1997).
154. Chaix, R., Cao, C. & Donnelly, P. Is mate choice in humans MHC-dependent? *PLoS Genet* **4**, e1000184 (2008).
155. Hedrick, P. W. & Black, F. L. HLA and mate selection: no evidence in South Amerindians. *Am J Hum Genet* **61**, 505-511 (1997).
156. Lawlor, D. A., Zemmour, J., Ennis, P. D. & Parham, P. Evolution of class-I MHC genes and proteins: from natural selection to thymic selection. *Annu Rev Immunol* **8**, 23-63 (1990).
157. McClelland, E. E., Penn, D. J. & Potts, W. K. Major histocompatibility complex heterozygote superiority during coinfection. *Infect Immun* **71**, 2079-2086 (2003).
158. Oliver, M. K., Telfer, S. & Piertney, S. B. Major histocompatibility complex (MHC) heterozygote superiority to natural multi-parasite infections in the water vole (Arvicola terrestris). *Proc Biol Sci* **276**, 1119-1128 (2009).
159. Bertoletti, A. *et al.* Natural variants of cytotoxic epitopes are T-cell receptor antagonists for antiviral cytotoxic T cells. *Nature* **369**, 407-410 (1994).
160. Moskophidis, D. & Zinkernagel, R. M. Immunobiology of cytotoxic T-cell escape mutants of lymphocytic choriomeningitis virus. *J Virol* **69**, 2187-2193 (1995).
161. Jeffery, K. J. *et al.* HLA alleles determine human T-lymphotropic virus-I (HTLV-I) proviral load and the risk of HTLV-I-associated myelopathy. *Proc Natl Acad Sci U S A* **96**, 3848-3853 (1999).
162. Penn, D. & Potts, W. The evolution of mating preferences and major histocompatibility genes. *Am Nat* **153**, 145‑164 (1999).
163. Jordan, W. C. & Bruford, M. W. New perspectives on mate choice and the MHC. *Heredity (Edinb)* **81 (Pt 2)**, 127-133 (1998).
164. Manning, C. J., Potts, W. K., Wakeland, E. K. & Dewsbury, D. A. in *Chemical signals in vertebrates* Vol. 6 (eds R.L. Doty and D. Muller-Schwarze) pp. 229-235 (Plenum, New York, 1992).
165. Potts, W. K., Manning, C. J. & Wakeland, E. K. Mating patterns in seminatural populations of mice influenced by MHC genotype. *Nature* **352**, 619-621 (1991).
166. Potts, W. K., Manning, C. J. & Wakeland, E. K. The role of infectious disease, inbreeding and mating preferences in maintaining MHC genetic diversity: an experimental test. *Philos Trans R Soc Lond B Biol Sci* **346**, 369-378 (1994).
167. Yamazaki, K. *et al.* Recognition among mice. Evidence from the use of a Y-maze differentially scented by congenic mice of different major histocompatibility types. *J Exp Med* **150**, 755-760 (1979).
168. Yamazaki, K. & Beauchamp, G. K. Genetic basis for MHC-dependent mate choice. *Adv Genet* **59**, 129-145 (2007).
169. Penn, D. & Potts, W. K. Untrained mice discriminate MHC-determined odors. *Physiol Behav* **64**, 235-243 (1998).
170. Carroll, L. S., Penn, D. J. & Potts, W. K. Discrimination of MHC-derived odors by untrained mice is consistent with divergence in peptide-binding region residues. *Proc Natl Acad Sci USA* **99**, 2187-2192 (2002).
171. Singh, P. B., Brown, R. E. & Roser, B. MHC antigens in urine as olfactory recognition cues. *Nature* **327**, 161-164 (1987).
172. Montag, S. *et al.* "Electronic nose" detects major histocompatibility complex-dependent prerenal and postrenal odor components. *Proc Natl Acad Sci USA* **98**, 9249-9254 (2001).
173. Kwak, J. *et al.* Major histocompatibility complex-regulated odortypes: peptide-free urinary volatile signals. *Physiol Behav* **96**, 184-188 (2009).
174. Kwak, J., Willse, A., Preti, G., Yamazaki, K. & Beauchamp, G. K. In search of the chemical basis for MHC odourtypes. *Proc Biol Sci* **277**, 2417-2425 (2010).
175. Chamero, P. *et al.* G protein G(alpha)o is essential for vomeronasal function and aggressive behavior in mice. *Proc Natl Acad Sci U S A* **108**, 12898-12903 (2011).
176. Hovis, K. R. *et al.* Activity regulates functional connectivity from the vomeronasal organ to the accessory olfactory bulb. *J Neurosci* **32**, 7907-7916 (2012).
177. Novotny, M. *et al.* in *Chemical signals: vertebrates and aquatic invertebrates* (eds D. Muller-Schwarze and R.M. Silverstein) pp. 377-390 (Plenum Press, New York, 1980).
178. Hurst, J. L. *et al.* Individual recognition in mice mediated by major urinary proteins. *Nature* **414**, 631-634 (2001).
179. Logan, D. W., Marton, T. F. & Stowers, L. Species specificity in major urinary proteins by parallel evolution. *PLoS One* **3**, e3280 (2008).
180. Hurst, J. L. & Beynon, R. J. Scent wars: the chemobiology of competitive signalling in mice. *BioEssays : news and reviews in molecular, cellular and developmental biology* **26**, 1288-1298 (2004).
181. Brennan, P. A., Schellinck, H. M. & Keverne, E. B. Patterns of expression of the immediate-early gene egr-1 in the accessory olfactory bulb of female mice exposed to pheromonal constituents of male urine. *Neuroscience* **90**, 1463-1470 (1999).
182. Robertson, D. H., Cox, K. A., Gaskell, S. J., Evershed, R. P. & Beynon, R. J. Molecular heterogeneity in the Major Urinary Proteins of the house mouse Mus musculus. *Biochem J* **316 (Pt 1)**, 265-272 (1996).
183. Robertson, D. H., Hurst, J. L., Bolgar, M. S., Gaskell, S. J. & Beynon, R. J. Molecular heterogeneity of urinary proteins in wild house mouse populations. *Rapid Commun Mass Spectrom* **11**, 786-790 (1997).
184. Beynon, R. J. *et al.* Polymorphism in major urinary proteins: molecular heterogeneity in a wild mouse population. *J Chem Ecol* **28**, 1429-1446 (2002).
185. Bacchini, A., Gaetani, E. & Cavaggioni, A. Pheromone binding proteins of the mouse, Mus musculus. *Experientia* **48**, 419-421 (1992).

186. Novotny, M. V., Ma, W., Wiesler, D. & Zidek, L. Positive identification of the puberty-accelerating pheromone of the house mouse: the volatile ligands associating with the major urinary protein. *Proc Biol Sci* **266**, 2017-2022 (1999).
187. Robertson, D. H. L., Beynon, R. J. & Evershed, R. P. Extraction, characterization and binding analysis of two pheromonally active ligands associated with major urinary protein of house mouse (Mus musculus). *J Chem Ecol* **19**, 1405–1416 (1993).
188. Robertson, D. H. L., Marie, A. D., Veggerby, C., Hurst, J. L. & Beynon, R. J. in *Chemical signals in vertebrates* Vol. 9 (eds A. Marchlewska-Koj, D. Muller-Schwarze and J. Lepri) pp. 169-176 (Plenum Press, New York, 2001).
189. Hurst, J. L., Robertson, D. H. L., Tolladay, U. & Beynon, R. J. Proteins in urine scent marks of male house mice extend the longevity of olfactory signals. *Animal behaviour* **55**, 1289-1297 (1998).
190. Armstrong, S. D., Robertson, D. H., Cheetham, S. A., Hurst, J. L. & Beynon, R. J. Structural and functional differences in isoforms of mouse major urinary proteins: a male-specific protein that preferentially binds a male pheromone. *Biochem J* **391**, 343-350 (2005).
191. Darwish Marie, A. *et al.* Effect of polymorphisms on ligand binding by mouse major urinary proteins. *Protein Sci* **10**, 411-417 (2001).
192. Sharrow, S. D., Vaughn, J. L., Zidek, L., Novotny, M. V. & Stone, M. J. Pheromone binding by polymorphic mouse major urinary proteins. *Protein Sci* **11**, 2247-2256 (2002).
193. Kwak, J. *et al.* Differential binding between volatile ligands and major urinary proteins due to genetic variation in mice. *Physiol Behav* **107**, 112-120 (2012).
194. Cheetham, S. A., Smith, A. L., Armstrong, S. D., Beynon, R. J. & Hurst, J. L. Limited variation in the major urinary proteins of laboratory mice. *Physiol Behav* **96**, 253-261 (2009).
195. Beck, J. A. *et al.* Genealogies of mouse inbred strains. *Nat Genet* **24**, 23-25 (2000).
196. Ferris, S. D., Sage, R. D. & Wilson, A. C. Evidence from mtDNA sequences that common laboratory strains of inbred mice are descended from a single female. *Nature* **295**, 163-165 (1982).
197. Goios, A., Pereira, L., Bogue, M., Macaulay, V. & Amorim, A. mtDNA phylogeny and evolution of laboratory mouse strains. *Genome Res* **17**, 293-298 (2007).
198. *Gene Partner*, <www.genepartner.com> (2013).
199. Milinski, M., Croy, I., Hummel, T. & Boehm, T. Major histocompatibility complex peptide ligands as olfactory cues in human body odour assessment. *Proc Biol Sci* **280**, 20122889 (2013).
200. Sherborne, A. L. *et al.* The genetic basis of inbreeding avoidance in house mice. *Curr Biol* **17**, 2061-2066 (2007).
201. Koller, B. H., Marrack, P., Kappler, J. W. & Smithies, O. Normal development of mice deficient in beta 2M, MHC class I proteins, and CD8+ T cells. *Science* **248**, 1227-1230 (1990).
202. Ehst, B. D., Ingulli, E. & Jenkins, M. K. Development of a novel transgenic mouse for the study of interactions between CD4 and CD8 T cells during graft rejection. *Am J Transplant* **3**, 1355-1362 (2003).
203. Sturm, T. *MHC-Klasse I-Bindemotivpeptide im Urin der Hausmaus (Mus musculus)*, Diploma thesis, University of Tübingen (2009).
204. Jones, B. & Janeway, C. A., Jr. Cooperative interaction of B lymphocytes with antigen-specific helper T lymphocytes is MHC restricted. *Nature* **292**, 547-549 (1981).
205. Hämmerling, G. J., Rüsch, E., Tada, N., Kimura, S. & Hämmerling, U. Localization of allodeterminants on H-2Kb antigens determined with monoclonal antibodies and H-2 mutant mice. *Proc Natl Acad Sci USA* **79**, 4737-4741 (1982).
206. Lemke, H., Hämmerling, G. J., Hohmann, C. & Rajewsky, K. Hybrid cell lines secreting monoclonal antibody specific for major histocompatibility antigens of the mouse. *Nature* **271**, 249-251 (1978).
207. Sturm, T. *et al.* Mouse urinary peptides provide a molecular basis for genotype discrimination by nasal sensory neurons. *Nature communications* **4**, 1616 (2013).
208. Hortin, G. L., Meilinger, B. & Drake, S. K. Size-selective extraction of peptides from urine for mass spectrometric analysis. *Clin Chem* **50**, 1092-1095 (2004).
209. Tinoco, A. D. & Saghatelian, A. Investigating endogenous peptides and peptidases using peptidomics. *Biochemistry* **50**, 7447-7461 (2011).
210. Michalski, A. *et al.* Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. *Mol Cell Proteomics* **10**, M111 011015 (2011).
211. Picotti, P. *et al.* A complete mass-spectrometric map of the yeast proteome applied to quantitative trait analysis. *Nature* **494**, 266-270 (2013).
212. Nesvizhskii, A. I., Vitek, O. & Aebersold, R. Analysis and validation of proteomic data generated by tandem mass spectrometry. *Nat Methods* **4**, 787-797 (2007).
213. Käll, L., Storey, J. D., MacCoss, M. J. & Noble, W. S. Posterior error probabilities and false discovery rates: two sides of the same coin. *J Proteome Res* **7**, 40-44 (2008).
214. Edman, P. A method for the determination of amino acid sequence in peptides. *Arch Biochem* **22**, 475 (1949).
215. Edman, P. & Begg, G. A protein sequenator. *Eur J Biochem* **1**, 80-91 (1967).
216. Steen, H. & Mann, M. The ABC's (and XYZ's) of peptide sequencing. *Nat Rev Mol Cell Biol* **5**, 699-711 (2004).
217. Dick, T. P. *et al.* The making of the dominant MHC class I ligand SYFPEITHI. *Eur J Immunol* **28**, 2478-2486 (1998).
218. Porgador, A., Yewdell, J. W., Deng, Y., Bennink, J. R. & Germain, R. N. Localization, quantitation, and in situ detection of specific peptide-MHC class I complexes using a monoclonal antibody. *Immunity* **6**, 715-726 (1997).
219. *UNIMOD*, <www.unimod.org> (2012).
220. Rappsilber, J., Mann, M. & Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat Protoc* **2**, 1896-1906 (2007).

221. Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. & Whitehouse, C. M. Electrospray ionization for mass spectrometry of large biomolecules. *Science* **246**, 64-71 (1989).
222. Walther, T. C. & Mann, M. Mass spectrometry-based proteomics in cell biology. *The Journal of cell biology* **190**, 491-500 (2010).
223. Karas, M. & Hillenkamp, F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Analytical chemistry* **60**, 2299-2301 (1988).
224. Kebarle, P. A brief overview of the present status of the mechanisms involved in electrospray mass spectrometry. *J Mass Spectrom* **35**, 804-817 (2000).
225. Blades, A. T., Ikonomou, M. G. & Kebarle, P. Mechanism of Electrospray Mass Spectrometry. Electrospray as an Electrolysis Cell. *Analytical chemistry* **63**, 2109-2114 (1991).
226. Taylor, G. Disintegration of Water Drops in an Electric Field. *Proc. R. Soc. Lond. A* **280**, 383-397 (1964).
227. Dole, M., Mack, L. L. & Hines, R. L. Molecular Beams of Macroions. *J. Chem. Phys.* **49**, 2240-2249 (1968).
228. Rayleigh, L. On the equilibrium of liquid conducting masses charged with electricity. *Philosophical Magazine* **14**, 184-186 (1882).
229. Schmelzeisen-Redeker, G., Bütfering, L. & Röllgen, F. W. Desolvation of ions and molecules in thermospray mass spectrometry. *Int. J. Mass Spectrom. Ion Processes* **90**, 139-150 (1989).
230. Nehring, H., Thiebes, S., Bütfering, L. & Röllgen, F. W. Cluster ion formation in thermospray mass spectrometry of ammonium salts. *Int. J. Mass Spectrom. Ion Processes* **128**, 123-132 (1993).
231. Iribarne, J. V. & Thomson, B. A. On the evaporation of small ions from charged droplets. *J Chem Phys* **64(6)**, 2287–2294 (1976).
232. Thomson, B. A. & Iribarne, J. V. Field induced ion evaporation from liquid surfaces at atmospheric pressure. *J Chem Phys* **71**, 4451-4463 (1979).
233. Olsen, J. V. *et al.* A dual pressure linear ion trap Orbitrap instrument with very high sequencing speed. *Mol Cell Proteomics* **8**, 2759-2769 (2009).
234. *Planet Orbitrap*, <http://planetorbitrap.com> (2013).
235. Michalski, A. *et al.* Ultra high resolution linear ion trap Orbitrap mass spectrometer (Orbitrap Elite) facilitates top down LC MS/MS and versatile peptide fragmentation modes. *Mol Cell Proteomics* **11**, O111 013698 (2012).
236. Lange, V., Picotti, P., Domon, B. & Aebersold, R. Selected reaction monitoring for quantitative proteomics: a tutorial. *Molecular systems biology* **4**, 222 (2008).
237. Douglas, D. J., Frank, A. J. & Mao, D. Linear ion traps in mass spectrometry. *Mass Spectrom Rev* **24**, 1-29 (2005).
238. Perry, R. H., Cooks, R. G. & Noll, R. J. Orbitrap mass spectrometry: instrumentation, ion motion and applications. *Mass Spectrom Rev* **27**, 661-699 (2008).
239. Makarov, A. *et al.* Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. *Analytical chemistry* **78**, 2113-2120 (2006).
240. Scigelova, M. & Makarov, A. Orbitrap mass analyzer--overview and applications in proteomics. *Proteomics* **6 Suppl 2**, 16-21 (2006).
241. Olsen, J. V. *et al.* Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol Cell Proteomics* **4**, 2010-2021 (2005).
242. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**, 1367-1372 (2008).
243. Cox, J. *et al.* Andromeda: a peptide search engine integrated into the MaxQuant environment. *J Proteome Res* **10**, 1794-1805 (2011).
244. Varga, J., Due, M., Frisvad, J. C. & Samson, R. A. Taxonomic revision of Aspergillus section Clavati based on molecular, morphological and physiological data. *Stud Mycol* **59**, 89-106 (2007).
245. Olsen, J. V. *et al.* Higher-energy C-trap dissociation for peptide modification analysis. *Nat Methods* **4**, 709-712 (2007).
246. *ProteinProspector*, <http://prospector.ucsf.edu/prospector/> (2013).
247. Schroeder, M. J., Shabanowitz, J., Schwartz, J. C., Hunt, D. F. & Coon, J. J. A neutral loss activation method for improved phosphopeptide sequence analysis by quadrupole ion trap mass spectrometry. *Analytical chemistry* **76**, 3590-3598 (2004).
248. *WideBand Activation Technology - Product Support Bulletin*. Thermo Fisher Scientific (2012).
249. Perkins, D. N., Pappin, D. J., Creasy, D. M. & Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551-3567 (1999).
250. Sherry, S. T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* **29**, 308-311 (2001).
251. Kitts, A. & Sherry, S. *The NCBI handbook, Chapter 5, The Single Nucleotide Polymorphism Database (dbSNP) of Nucleotide Sequence Variation.*, Vol. Chapter 5 (National Library of Medicine (US), National Center for Biotechnology Information, 2002), http://www.ncbi.nlm.nih.gov/books/NBK21088/.
252. Pruitt, K., Brown, G., Tatusova, T. & Maglott, D. *The NCBI handbook, Chapter 18, The Reference Sequence (RefSeq) Project.*, Vol. Chapter 18 (National Library of Medicine (US), National Center for Biotechnology Information, 2002), http://www.ncbi.nlm.nih.gov/books/NBK21091/.
253. Dreszer, T. R. *et al.* The UCSC Genome Browser database: extensions and updates 2011. *Nucleic Acids Res* **40**, D918-923 (2012).
254. *UCSC Genome Browser database*, <http://genome.ucsc.edu/> (2012).
255. *dbSNP Short Genetic Variations*, <http://www.ncbi.nlm.nih.gov/projects/SNP/> (2012).
256. Maddatu, T. P., Grubb, S. C., Bult, C. J. & Bogue, M. A. Mouse Phenome Database (MPD). *Nucleic Acids Res* **40**, D887-894 (2012).

257. *Mouse Phenome Database (MPD)*, <http://phenome.jax.org/SNP > (2012).
258. Cox, J., Michalski, A. & Mann, M. Software lock mass by two-dimensional minimization of peptide mass errors. *J Am Soc Mass Spectrom* **22**, 1373-1380 (2011).
259. Creasy, D. M. & Cottrell, J. S. Unimod: Protein modifications for mass spectrometry. *Proteomics* **4**, 1534-1536 (2004).
260. Ying, J., Clavreul, N., Sethuraman, M., Adachi, T. & Cohen, R. A. Thiol oxidation in signaling and response to stress: detection and quantification of physiological and pathophysiological thiol modifications. *Free Radic Biol Med* **43**, 1099-1108 (2007).
261. Griffiths, S. W., King, J. & Cooney, C. L. The reactivity and oxidation pathway of cysteine 232 in recombinant human alpha 1-antitrypsin. *J Biol Chem* **277**, 25486-25492 (2002).
262. Woo, H. A. *et al.* Reduction of cysteine sulfinic acid by sulfiredoxin is specific to 2-cys peroxiredoxins. *J Biol Chem* **280**, 3125-3128 (2005).
263. Paron, I. *et al.* A proteomic approach to identify early molecular targets of oxidative stress in human epithelial lens cells. *Biochem J* **378**, 929-937 (2004).
264. Dormann, P. *et al.* Amino acid exchange and covalent modification by cysteine and glutathione explain isoforms of fatty acid-binding protein occurring in bovine liver. *J Biol Chem* **268**, 16286-16292 (1993).
265. Kolarich, D., Weber, A., Turecek, P. L., Schwarz, H. P. & Altmann, F. Comprehensive glyco-proteomic analysis of human alpha1-antitrypsin and its charge isoforms. *Proteomics* **6**, 3369-3380 (2006).
266. Bergenhem, N., Carlsson, U. & Strid, L. The existence of glutathione and cysteine disulfide-linked to erythrocyte carbonic anhydrase from tiger shark. *Biochim Biophys Acta* **871**, 55-60 (1986).
267. Perla-Kajan, J., Twardowski, T. & Jakubowski, H. Mechanisms of homocysteine toxicity in humans. *Amino Acids* **32**, 561-572 (2007).
268. Laurell, C. B. Complexes formed in vivo between immunoglobulin light chain kappa, prealbumin and-or alpha-1-antitrypsin in myeloma sera. *Immunochemistry* **7**, 461-465 (1970).
269. Laurell, C. B. & Thulin, E. Complexes in plasma between light chain kappa immunoglobulins and alpha 1-antitrypsin respectively prealbumin. *Immunochemistry* **11**, 703-709 (1974).
270. Musiani, P., Lauriola, L. & Piantelli, M. Inhibitory activity of alpha-1-antitrypsin bound to human IgA. *Clin Chim Acta* **85**, 61-66 (1978).
271. de Beus, M. D., Chung, J. & Colon, W. Modification of cysteine 111 in Cu/Zn superoxide dismutase results in altered spectroscopic and biophysical properties. *Protein Sci* **13**, 1347-1355 (2004).
272. *Wissenschaftliche Tabellen Geigy, Teilband Haematologie und Humangenetik*. 8 edn, eds (Ciba-Geigy AG, Basel, 1979).
273. Steinert, P. M., Kim, S. Y., Chung, S. I. & Marekov, L. N. The transglutaminase 1 enzyme is variably acylated by myristate and palmitate during differentiation in epidermal keratinocytes. *J Biol Chem* **271**, 26242-26250 (1996).
274. Wilson, J. P., Raghavan, A. S., Yang, Y. Y., Charron, G. & Hang, H. C. Proteomic analysis of fatty-acylated proteins in mammalian cells with chemical reporters reveals S-acylation of histone H3 variants. *Mol Cell Proteomics* **10**, M110 001198 (2011).
275. Bach, R., Konigsberg, W. H. & Nemerson, Y. Human tissue factor contains thioester-linked palmitate and stearate on the cytoplasmic half-cystine. *Biochemistry* **27**, 4227-4231 (1988).
276. Stults, J. T. *et al.* Lung surfactant protein SP-C from human, bovine, and canine sources contains palmityl cysteine thioester linkages. *Am J Physiol* **261**, L118-125 (1991).
277. Batthyany, C. *et al.* Reversible post-translational modification of proteins by nitrated fatty acids in vivo. *J Biol Chem* **281**, 20450-20463 (2006).
278. Blatnik, M., Thorpe, S. R. & Baynes, J. W. Succination of proteins by fumarate: mechanism of inactivation of glyceraldehyde-3-phosphate dehydrogenase in diabetes. *Ann N Y Acad Sci* **1126**, 272-275 (2008).
279. Comporti, M. Lipid peroxidation and biogenic aldehydes: from the identification of 4-hydroxynonenal to further achievements in biopathology. *Free Radic Res* **28**, 623-635 (1998).
280. Esterbauer, H., Schaur, R. J. & Zollner, H. Chemistry and biochemistry of 4-hydroxynonenal, malonaldehyde and related aldehydes. *Free Radic Biol Med* **11**, 81-128 (1991).
281. Sawa, T. *et al.* Protein S-guanylation by the biological signal 8-nitroguanosine 3',5'-cyclic monophosphate. *Nat Chem Biol* **3**, 727-735 (2007).
282. Glaser, C. B., Karic, L., Huffaker, T., Chang, R. & Martin, J. Studies on the disulfide region of alpha 1-protease inhibitor. *Int J Pept Protein Res* **20**, 56-62 (1982).
283. Roepstorff, P. & Fohlman, J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed Mass Spectrom* **11**, 601 (1984).
284. Biemann, K. Mass spectrometry of peptides and proteins. *Annu Rev Biochem* **61**, 977-1010 (1992).
285. Papayannopoulos, I. A. The interpretation of collision-induced dissociation tandem mass spectra of peptides. *Mass Spectrometry Reviews* **14**, 49-73 (1995).
286. Neuhauser, N., Michalski, A., Cox, J. & Mann, M. Expert system for computer-assisted annotation of MS/MS spectra. *Mol Cell Proteomics* **11**, 1500-1509 (2012).
287. Tabb, D. L., Huang, Y., Wysocki, V. H. & Yates, J. R., 3rd. Influence of basic residue content on fragment ion peak intensities in low-energy collision-induced dissociation spectra of peptides. *Analytical chemistry* **76**, 1243-1248 (2004).
288. Ljunggren, H. G. & Kärre, K. Host resistance directed selectively against H-2-deficient lymphoma variants. Analysis of the mechanism. *J Exp Med* **162**, 1745-1759 (1985).
289. Powis, S. J. *et al.* Restoration of antigen presentation to the mutant cell line RMA-S by an MHC-linked transporter. *Nature* **354**, 528-531 (1991).

290. Ljunggren, H. G. *et al.* Empty MHC class I molecules come out in the cold. *Nature* **346**, 476-480 (1990).
291. Berger, A. E., Davis, J. E. & Cresswell, P. Monoclonal antibody to HLA-A3. *Hybridoma* **1**, 87-90 (1982).
292. Engelhard, J. *Identifizierung HLA-A*01-restringierter T-Zellepitope aus Prostatakrebs-Tumorantigenen*, Diplom thesis, University of Tübingen (2005).
293. Chen, W., Khilko, S., Fecondo, J., Margulies, D. H. & McCluskey, J. Determinant selection of major histocompatibility complex class I-restricted antigenic peptides is explained by class I-peptide affinity and is strongly influenced by nondominant anchor residues. *J Exp Med* **180**, 1471-1483 (1994).
294. Spies, P., Mueller, R., Chen, G. J. & Gygax, D. A simple approach to improve the sensitivity of enzyme-linked immunosorbent assay: using the IMAPlate 5RC96 for result readout. *Anal Biochem* **397**, 48-52 (2010).
295. Altman, J. D. *et al.* Phenotypic analysis of antigen-specific T lymphocytes. *Science* **274**, 94-96 (1996).
296. Townsend, A. *et al.* Association of class I major histocompatibility heavy and light chains induced by viral peptides. *Nature* **340**, 443-448 (1989).
297. Schumacher, T. N. *et al.* Direct binding of peptide to empty MHC class I molecules on intact cells and in vitro. *Cell* **62**, 563-567 (1990).
298. Dumrese, T. *Strategies for the identification of epitopes of cytotoxic T lymphocytes (in German)*, PhD thesis, University of Tübingen (1998).
299. Malarkannan, S., Gonzalez, F., Nguyen, V., Adair, G. & Shastri, N. Alloreactive CD8+ T cells can recognize unusual, rare, and unique processed peptide/MHC complexes. *J Immunol* **157**, 4464-4473 (1996).
300. Runck, A. M., Moriyama, H. & Storz, J. F. Evolution of duplicated beta-globin genes and the structural basis of hemoglobin isoform differentiation in Mus. *Mol Biol Evol* **26**, 2521-2532 (2009).
301. Madden, T. *The NCBI handbook, Chapter 16, The BLAST Sequence Analysis Tool.*, Vol. Chapter 16 (National Library of Medicine (US), National Center for Biotechnology Information, 2002), http://www.ncbi.nlm.nih.gov/books/NBK21097/.
302. Wang, J. R. *et al.* Imputation of single-nucleotide polymorphisms in inbred mice using local phylogeny. *Genetics* **190**, 449-458 (2012).
303. Yang, H. *et al.* Subspecific origin and haplotype diversity in the laboratory mouse. *Nat Genet* **43**, 648-655 (2011).
304. Frazer, K. A. *et al.* A sequence-based variation map of 8.27 million SNPs in inbred mouse strains. *Nature* **448**, 1050-1053 (2007).
305. Mural, R. J. *et al.* A comparison of whole-genome shotgun-derived mouse chromosome 16 and the human genome. *Science* **296**, 1661-1671 (2002).
306. Boursot, P., Auffray, J.-C., Britton-Davidian, J. & Bonhomme, F. The evolution of house mice. *Annu Rev Ecol Syst* **24**, 119-152 (1993).
307. Robertson, D. H., Hurst, J. L., Searle, J. B., Gunduz, I. & Beynon, R. J. Characterization and comparison of major urinary proteins from the house mouse, Mus musculus domesticus, and the aboriginal mouse, Mus macedonicus. *J Chem Ecol* **33**, 613-630 (2007).
308. Mudge, J. M. *et al.* Dynamic instability of the major urinary protein gene family revealed by genomic and phenotypic comparisons between C57 and 129 strain mice. *Genome Biol* **9**, R91 (2008).
309. Jiang, W. *et al.* The alpha subunit of meprin A. Molecular cloning and sequencing, differential expression in inbred mouse strains, and evidence for divergent evolution of the alpha and beta subunits. *J Biol Chem* **267**, 9185-9193 (1992).
310. Reckelhoff, J. F., Butler, P. E., Bond, J. S., Beynon, R. J. & Passmore, H. C. Mep-1, the gene regulating meprin activity, maps between Pgk-2 and Ce-2 on mouse chromosome 17. *Immunogenetics* **27**, 298-300 (1988).
311. Norman, L. P., Matters, G. L., Crisman, J. M. & Bond, J. S. Expression of meprins in health and disease. *Current topics in developmental biology* **54**, 145-166 (2003).
312. Kentsis, A. *et al.* Urine proteomics for discovery of improved diagnostic markers of Kawasaki disease. *EMBO molecular medicine* **5**, 210-220 (2013).
313. Natsch, A., Kuhn, F. & Tiercy, J. M. Lack of evidence for HLA-linked patterns of odorous carboxylic acids released from glutamine conjugates secreted in the human axilla. *J Chem Ecol* **36**, 837-846 (2010).
314. Rötzschke, O. *et al.* Exact prediction of a natural T cell epitope. *Eur J Immunol* **21**, 2891-2894 (1991).
315. Shastri, N. & Gonzalez, F. Endogenous generation and presentation of the ovalbumin peptide/Kb complex to T cells. *J Immunol* **150**, 2724-2736 (1993).
316. Spies, P., Chen, G. J. & Gygax, D. Establishment of a miniaturized enzyme-linked immunosorbent assay for human transferrin quantification using an intelligent multifunctional analytical plate. *Anal Biochem* **382**, 35-39 (2008).
317. Tirumalai, R. S. *et al.* Characterization of the low molecular weight human serum proteome. *Mol Cell Proteomics* **2**, 1096-1103 (2003).
318. Milinski, M., Griffiths, S. W., Reusch, T. B. & Boehm, T. Costly major histocompatibility complex signals produced only by reproductively active males, but not females, must be validated by a 'maleness signal' in three-spined sticklebacks. *Proc Biol Sci* **277**, 391-398 (2010).
319. Boehm, T. Sensory biology: A whiff of genome. *Nature* **496**, 304-305 (2013).
320. Köhler, G., Fischer-Lindahl, K. & Heusser, C. in *The Immune System* Vol. 2 (eds C. Steinberg and I. Lefkovits) pp. 202-208 (Karger, Basel, 1981).
321. Pearse-Pratt, R., Schellinck, H., Brown, R., Singh, P. B. & Roser, B. Soluble MHC antigens and olfactory recognition of genetic individuality: the mechanism. *Genetica* **104**, 223-230 (1998).
322. Barnstable, C. J. *et al.* Production of monoclonal antibodies to group A erythrocytes, HLA and other human cell surface antigens-new tools for genetic analysis. *Cell* **14**, 9-20 (1978).

323. Zhou, M. *et al.* An investigation into the human serum "interactome". *Electrophoresis* **25**, 1289-1298 (2004).
324. Deres, K., Beck, W., Faath, S., Jung, G. & Rammensee, H. G. MHC/peptide binding studies indicate hierarchy of anchor residues. *Cell Immunol* **151**, 158-167 (1993).
325. Stevanović, S. & Schild, H. Quantitative aspects of T cell activation--peptide generation and editing by MHC class I molecules. *Semin Immunol* **11**, 375-384 (1999).
326. Reits, E. A., Vos, J. C., Gromme, M. & Neefjes, J. The major substrates for TAP in vivo are derived from newly synthesized proteins. *Nature* **404**, 774-778 (2000).
327. Hunt, D. F. *et al.* Characterization of peptides bound to the class I MHC molecule HLA-A2.1 by mass spectrometry. *Science* **255**, 1261-1263 (1992).
328. Perreault, C., Jutras, J., Roy, D. C., Filep, J. G. & Brochu, S. Identification of an immunodominant mouse minor histocompatibility antigen (MiHA). T cell response to a single dominant MiHA causes graft-versus-host disease. *J Clin Invest* **98**, 622-628 (1996).
329. McBride, K. *et al.* The model B6(dom1) minor histocompatibility antigen is encoded by a mouse homolog of the yeast STT3 gene. *Immunogenetics* **54**, 562-569 (2002).
330. Klein, D., Tewarson, S., Figueroa, F. & Klein, J. The minimal length of the differential segment in H-2 congenic lines. *Immunogenetics* **16**, 319-328 (1982).
331. Yamazaki, K., Yamaguchi, M., Andrews, P. W., Peake, B. & Boyse, E. A. Mating Preferences of F2 Segregants of Crosses between MHC Congenic Mouse Strains. *Immunogenetics* **6**, 253-259 (1978).
332. Vincek, V., Sertic, J., Zaleska-Rutczynska, Z., Figueroa, F. & Klein, J. Characterization of H-2 congenic strains using DNA markers. *Immunogenetics* **31**, 45-51 (1990).
333. Ziegler, A., Santos, P. S., Kellermann, T. & Uchanska-Ziegler, B. Self/nonself perception, reproduction and the extended MHC. *Self Nonself* **1**, 176-191 (2010).
334. Bard, J., Yamazaki, K., Curran, M., Boyse, E. A. & Beauchamp, G. K. Effect of B2m gene disruption on MHC-determined odortypes. *Immunogenetics* **51**, 514-518 (2000).
335. Yamazaki, K. *et al.* Sensory distinction between H-2b and H-2bm1 mutant mice. *Proc Natl Acad Sci USA* **80**, 5685-5688 (1983).
336. Brown, J. L. *Some paradoxical goals of cells and organisms: the role of the MHC.* Ethical Questions in Brain and Behavior. edn, 111-124, eds D.W. Pfaff (Springer Verlag, New York, 1983).
337. Boehm, T. Quality control in self/nonself discrimination. *Cell* **125**, 845-858 (2006).
338. Church, D. M. *et al.* Lineage-specific biology revealed by a finished genome assembly of the mouse. *PLoS biology* **7**, e1000112 (2009).
339. Matsumura, M., Fremont, D. H., Peterson, P. A. & Wilson, I. A. Emerging principles for the recognition of peptide antigens by MHC class I molecules. *Science* **257**, 927-934 (1992).

8. Acknowledgements

First of all, I would like to thank Hans-Georg Rammensee for his great scientific ideas that provided the fundament for many key results of this study, most importantly the concept of the MHC-based ELISA and the proposal to search the urinary peptidome for SAV peptides. Although I could not put all his suggestions into practise, his unconventional, extremely creative mind was always inspiring. I thank him for the far-reaching freedom that he granted to me in deciding what to do. I acknowledge his large efforts to coup with bureaucracy, narrowly evading imprisonment due to temporarily unregistered collection of mouse urine! I am grateful that Hans-Georg provided a perfect financial framework for my thesis.

I greatly appreciate the work of Peter Overath, who prepared the present study together with Hans-Georg Rammensee before I joined the lab, and who solved many important technical issues associated with the detection of SIINFEKL at that early time. Peter Overath was highly valuable with the study of literature, regularly supporting me with the most interesting and latest articles concerning olfaction, and he also discovered the reports about the IMAP-ELISA, a milestone in optimising the SIINFEKL detection. Furthermore, he was a valuable partner for discussions, and he wrote the first draft our manuscript finally published in *Nature Communications*.

I am grateful for Stefan Stevanović`s valuable advice regarding all issues of MHC motifs, MHC peptide binding, synthetic peptides as well as preparative HPLC. He was also very helpful in finding MHC-related literature. Stefan`s humour transforms challenging science into an entertaining business!

I would like to thank all of my three main supervisors, Hans-Georg Rammensee, Peter Overath and Stefan Stevanović for their fairness, their personal correctness and the highly motivating working atmosphere that they created. I value Hans-Georg Rammensee`s and Peter Overath`s critical comments on the manuscript of the present thesis.

Although three supervisors are already quite a lot, I had the great pleasure to have Boris Maček as kind of a forth, special MS supervisor. It was wonderful to profit from Boris Maček`s huge knowledge about MS. Together with Stefan Jung, he designed most of the settings for operating the LTQ Orbitrap XL as well as the nano-HPLC, and they not only operated the LC-MS system but also performed the first round of MS data processing with MaxQuant version 1.0.14.3. I am also thankful to Johannes Madlung, Daniel Kowalewski and Marc Günder for technical assistance with MS measurements. I would like to thank all of these persons as well as Oliver Drews, Heiko Schuster, Alejandro Carpy and Boumediene Soufi for highly interesting and instructive discussions about MS and HPLC.

The present work greatly profited from the bioinformatic expertise of Mathias Walzer and Karsten Krug. Mathias Walzer searched the dbSNP database for non-synonymous SNVs matching our urinary

peptide data set. Karsten Krug was very helpful when I established the MaxQuant workflow in our own department. Both of them were perfect contact partners for all my bioinformatic questions.

Our work in Tübingen was crucially complemented by our collaboration with the neurobiologists Trese Leinders-Zufall and Frank Zufall from the University of Saarland. Despite some tenacious discussions about the possible impact of MHC-independent urinary peptides in olfaction at the beginning of my PhD thesis, which challenged the original thinking of many neurobiologists and behavioural scientists, the collaboration was finally very fruitful. Applying our peptides, Trese Leinders-Zufall demonstrated for example that VSNs also detect MHC-independent MHC motif peptides and that the detection and discrimination of peptides by VSNs is not dependent on an MHC motif. She performed all neurobiological experiments of our publication in *Nature Communications*, which I frequently referred to in the present thesis. Frank Zufall made major contributions in writing the manuscript for *Nature Communications*, and he also supplied us with interesting publications. The collaboration was very enlightening for me not only in terms of insights into neurobiology, olfaction and behaviour but also with regard to general strategies for science and scientific publications.

Beate Pömmerl was indispensable for the present study. She bred the B6/OVA^{+} and the B6/OVA^{+}/β_2m$^{-/-}$ mice and phenotyped them. She performed many flow cytometric experiments depicted in this thesis. Beate Pömmerl (MHC and ovalbumin stainings) and Karoline Laske (RMA-S experiment) assisted me with my own flow cytometric measurements. For many months, Beate manually collected the mouse urine together with me. Her original mind and her lovely temperament prevented the daily one hour urine collection procedures from getting boring.

I thank Hubert Kalbacher for his help with purifying synthetic SIINFEKL in his laboratory and for the possibility to work there. I would like to acknowledge Claudia Falkenburger for supplying me with several antibodies, an RMA-S starter cell culture and ready-to-use cell culture medium. Nicole Zuschke and Patricia Hrstić were precious sources for all synthetic peptides, and Jeanette Hauger was the main producer of the *in vitro* folded H2-K^{b} molecules for the SIINFEKL-H2-K^{b}-ELISA. The steady work of Franziska Löwenstein in cleaning glass ware and metabolic cages is highly appreciated, and I am also grateful for the zoo keepers caring for the well-being of the mice.

It is impossible to list all the people that were available for my questions regarding special laboratory or biological issues or that borrowed me their chemical reagents. However, I would like to thank all these people mainly from our Department of Immunology and from the Proteome Center Tübingen. My colleagues created a very pleasant working atmosphere, and I will definitely keep my time in Tübingen in very good memory.

The present study was substantially funded by the Deutsche Forschungsgemeinschaft grant Schwerpunktprogramm 1392 „Integrative Analysis of Olfaction“. Last but not least, I would like to thank the mice for providing their urine!

9. Supplementary Information

Supplementary Table 9-1 | B6/OVA$^+$/β_2m$^{-/-}$ mice express less ovalbumin on lymphocytes than B6/OVA$^+$ mice and this difference probably results from the unequal copy number of the ovalbumin transgene in these two mouse strains.

Date of exp.	Mouse strain	Ear no.	Zygosity of ovalbumin transgene	Used for anti-SIINFEKL-H2-K^b-ELISA	MFI of anti-ovalbumin staining
2009-07-02	B6/OVA$^+$	81	homo (br)	no	545.4
		82	homo (br)	no	596.7
		83	homo (br)	no	605.3
		84	homo (br)	no	561.9
		85	homo (br)	no	541.2
		86	homo (br)	no	516.6
		87	homo (br)	yes (also exp. no. 14 – 16)	580.2
		88	homo (br)	yes (also exp. no. 14 – 16)	557.3
		89	homo (br)	yes (also exp. no. 14 – 16)	667.6
		90	homo (br)	yes (exp. no. 14 – 16)	551.9
		91	homo (br)	no	532.4
		92	homo (br)	no	538.7
		93	homo (br)	no	527.7
		mean			563.3
	B6/OVA$^+$/β_2m$^{-/-}$	254	hetero (MFI)	no	317.3
		255	hetero (MFI)	no	340.4
		258	hetero (MFI)	no	312.2
		262	hetero (MFI)	no	309.7
		266	hetero (MFI)	no	341.2
		267	homo (MFI)	no	594.3
		269	hetero (MFI)	no	344.5
		mean (all B6/OVA$^+$/β_2m$^{-/-}$ mice)			365.7
		mean (heterozygous B6/OVA$^+$/β_2m$^{-/-}$ mice only)			327.6
2009-09-23	B6/OVA$^+$	95	homo (br)	no	706.9
		96	homo (br)	no	777.4
		97	homo (br)	no	766.2
		mean			750.1
	B6/OVA$^+$/β_2m$^{-/-}$	275	hetero (MFI)	no	414.5
		276	hetero (MFI)	no	420.0
		277	hetero (MFI)	no	393.2
		278	hetero (MFI)	yes	400.6
		279	hetero (MFI)	no	410.6
		280	hetero (MFI)	no	425.4
		281	hetero (MFI)	no	401.4
		282	hetero (MFI)	yes	398.8
		285	hetero (MFI)	no	392.8
		286	hetero (MFI)	no	349.4
		288	hetero (MFI)	no	385.8
		289	hetero (br)	yes	371.2
		290	hetero (br)	no	395.8
		291	hetero (MFI)	no	390.5
		292	hetero (MFI)	no	392.9
		mean			396.2

(Table continued on next page)

Supplementary Table 9-1 (continued)

Date of exp.	Mouse strain	Ear no.	Zygosity of ovalbumin transgene	Used for anti-SIINFEKL-H2-K^b-ELISA	MFI of anti-ovalbumin staining
2010-01-28	B6/OVA^+	98	homo (br)	yes	666.4
		99	homo (br)	yes	403.4
		101	homo (br)	no	527.4
		102	homo (br)	yes	690.2
		103	homo (br)	no	634.4
		105	homo (br)	no	622.4
		106	homo (br)	no	644.5
		107	homo (br)	yes	597.8
		108	homo (br)	no	658.2
		109	homo (br)	no	648.3
		110	homo (br)	yes	649.4
		111	homo (br)	no	620.4
		112	homo (br)	no	603.8
		mean			612.8
	B6/$OVA^+/\beta_2m^{-/-}$	293	hetero (br)	no	360.5
		294	hetero (br)	yes	381.6
		295	hetero (MFI)	no	377.5
		296	hetero (MFI)	no	327.3
		297	hetero (MFI)	no	367.5
		298	hetero (MFI)	no	390.5
		302	hetero (MFI)	no	331.3
		303	hetero (MFI)	no	373.4
		304	hetero (MFI)	no	341.1
		308	homo (MFI)	no	769.5
		mean (all B6/$OVA^+/\beta_2m^{-/-}$ mice)			402.0
		mean (heterozygous B6/$OVA^+/\beta_2m^{-/-}$ mice only)			361.2
2010-02-11	B6/$OVA^+/\beta_2m^{-/-}$	309	hetero (MFI)	no	680.9
		310	hetero (MFI)	no	652.4
		311	homo (MFI)	no	1239.9
		312	hetero (MFI)	no	601.2
		313	hetero (MFI)	no	642.7
		315	homo (MFI)	no	1026.1
		316	hetero (MFI)	no	536.4
		317	hetero (MFI)	no	592.5
		318	hetero (MFI)	no	566.3
		319	hetero (MFI)	no	564.2
		321	hetero (MFI)	no	542.7
		322	hetero (MFI)	no	531.6
		mean (all B6/$OVA^+/\beta_2m^{-/-}$ mice)			681.4
		mean (heterozygous B6/$OVA^+/\beta_2m^{-/-}$ mice only)			591.1

All B6/OVA^+ mice mentioned in this thesis were homozygous for the ovalbumin transgene because they were generated by repeated matings of homozygous B6/OVA^+ mice originally obtained from the Jackson Laboratories (see section "2.1. Mice"). Strikingly, the MFI of the anti-ovalbumin staining was decreased by almost factor two in most B6/$OVA^+/\beta_2m^{-/-}$ mice including those used for the anti-SIINFEKL-H2-K^b-ELISA (p = 0.0017 for exp. from 2009-07-02, p = 0.0026 for exp. from 2009-09-23, p = 0.0005 for exp. from 2010-01-28 considering all B6/$OVA^+/\beta_2m^{-/-}$ mice of the respective experiment; two-tailed, heteroskedastic Student`s t-test without Bonferroni correction; note that MFIs can only be compared within the same experiment). This can most likely be explained by the heterozygous expression of the ovalbumin transgene in these B6/$OVA^+/\beta_2m^{-/-}$ mice. Actually, the ovalbumin heterozygosity of some B6/$OVA^+/\beta_2m^{-/-}$ males (labelled with "br") was conclusively tested and confirmed by crossing them with B6/$\beta_2m^{-/-}$ females and typing the offspring for ovalbumin expression. The comparative analysis of ovalbumin expression in B6/OVA^+ and B6/$OVA^+/\beta_2m^{-/-}$ mice is not affected by the difference in β_2m expression (Supplementary Table 9-2). Indeed, a few B6/$OVA^+/\beta_2m^{-/-}$ mice displayed MFIs in the range of B6/OVA^+ mice differing from the other B6/$OVA^+/\beta_2m^{-/-}$ mice with extreme statistical significance ($p << 0.0001$, based on the calculation of MFI confidence intervals for heterozygous B6/$OVA^+/\beta_2m^{-/-}$ mice). Therefore, these B6/$OVA^+/\beta_2m^{-/-}$ mice (highlighted in green) are expected to be homozygous for the ovalbumin transgene. All mice used to generate the data for Figures 3-4 and 3-5 are printed in blue. "(br)" indicates that the zygosity of the ovalbumin transgene is derived from breeding history (B6/OVA^+ mice) and / or ovalbumin typing of the offspring. In contrast, if the zygosity is merely deduced from the mean fluorescence intensity of anti-ovalbumin staining, the tag "(MFI)" is displayed. homo = homozygous; hetero = heterozygous; Flow cytometric data depicted in this table were generated by Beate Pömmerl and kindly provided for statistical analysis.

Supplementary Table 9-2 | The measured difference in anti-ovalbumin staining of lymphocytes from B6/OVA$^+$ versus B6/OVA$^+$/β_2m$^{-/-}$ mice (see Supplementary Table 9-1) is essentially not influenced by the β_2m knockout.

Date of exp.	Mouse strain	Ear no.	MFI of anti-ovalbumin staining
2007-09-13	B6	16	44.5
		17	48.4
		23	21.4
		30	34.3
		37	29.4
		mean	35.6
	B6/β_2m$^{-/-}$	25	28.5
		38	29.4
		mean	29.0
2008-01-11	B6	1	33.2
		3	21.8
		65	36.2
		70	27.6
		73	35.6
		75	41.1
		77	37.0
		82	24.8
		83	24.8
		mean	31.3
	B6/β_2m$^{-/-}$	2	26.1
		68	27.6
		mean	26.9
2008-05-29	B6	142	31.9
	B6/β_2m$^{-/-}$	146	35.3
		149	36.3
		mean	35.8

The differences in MFI between B6 and B6/β_2m$^{-/-}$ mice are negligible and statistically not significant (p = 0.25 for exp. from 2007-09-13 and p = 0.09 for exp. from 2008-01-11, two tailed, heteroskedastic Student`s t-test). Additionally, the absolute MFIs of anti-ovalbumin staining are very low in B6 and B6/β_2m$^{-/-}$ mice as compared to B6/OVA$^+$ and B6/OVA$^+$/β_2m$^{-/-}$ mice (see Supplementary Table 9-1). Hence, although absolute MFIs cannot be accurately compared between different experiments, the impact of the β_2m knockout on the comparison of ovalbumin expression in B6/OVA$^+$ and B6/OVA$^+$/β_2m$^{-/-}$ mice is negligible. MFI = mean fluorescence intensity. Flow cytometric data depicted in this table were generated by Beate Pömmerl and kindly provided for statistical analysis.

Supplementary Table 9-3 | MS reliability parameters for the 17 MHC motif peptides and three extended forms thereof (LNSVFDQLGSY, IDQTRVLNLGPI and SGNFIDQTRVLNLGPITR) (cf. Tables 3-1, 3-2 and 3-3) (*Nat Comm*, Theo).

H2 allele	Peptide	Theoretical monoisotopic mass	Exp. no.	Charge	LTQ-MS²		FT-Orbitrap-MS²	
					Δm in ppm	Mascot Score	Δm in ppm	Mascot Score
K^b	NKQE**F**GW**I**	1020.5029	3	2	-0.366	33	n. i.	n. i.
	*LNSV**F**DQ**L**	934.4760	1	2	0.225	49	n. i.	n. i.
	"	"	2	2	0.262	53	n. i.	n. i.
	LAPQP**F**LR**V**	1039.6179	1	2	0.260	34	0.473	32
(K^b)	*LNSV**F**DQ**L**GSY	1241.5928	1	2	-0.055	79	n. i.	n. i.
	"	"	2	2	0.053	86	0.046	6
D^b	FNIQ**N**REPL**I**	1242.6721	2	2	0.246	40	n. i.	n. i.
	KELQ**N**SIID**L**	1171.6449	1	2	0.361	54	n. i.	n. i.
	"	"	2	2	0.290	39	0.210	8
	*TRVL**N**LGP**I**	981.5971	1	2	-0.173	46	-0.056	33
	"	"	2	2	-0.061	69	0.169	44
(D^b)	*IDQTRVL**N**LGP**I**	1337.7667	1	2	-0.122	81	-0.238	36
	"	"	2	2	0.010	97	-0.119	14
	"	"	3	2	-0.411	70	-0.411	15
	SGNFIDQTRVL**N**LGP**I**TR	2000.0803	1	2 or 3	0.058	57	-0.049	15
	"	"	2	3	0.049	67	-0.183	23
K^d	*V**Y**RPDQVSI**L**	1188.6503	1	2	0.141	43	n. i.	n. i.
	"	"	2	2	0.303	49	n. i.	n. i.
	L**Y**WVDVERQ**V**	1305.6717	1	2	0.078	45	n. i.	n. i.
	L**F**KDSAFG**L**	996.5280	1	2	0.157	37	n. i.	n. i.
	"	"	2	2	-0.172	28	n. i.	n. i.
L^d	Y**S**MPPIVR**F**	1108.5739	2	2	-0.033	33	n. i.	n. i.
	***P**AVRGFS**L**	845.4759	2	2	0.757	27	n. i.	n. i.
	SSDIKER**F**	980.4927	1	2 or 3	-0.019	40	n. i.	n. i.
	L**S**SDIKER**F**	1093.5768	1	2 or 3	0.177	42	-0.059	12
	PSFVPLSK**F**	1020.5644	1	2	0.029	46	n. i.	n. i.
	"	"	2	2	-0.193	37	-0.042	5
	"	"	3	2	0.235	21	0.235	12
	***P**VESKIY**F**	981.5171	2	2	-0.133	33	n. i.	n. i.
	L**S**RQMGMV**F**	1067.5256	1	2	0.174	44	0.164	15
	"	"	2	2	0.301	47	-0.212	8
	G**P**VQGTIH**F**	954.4923	1	2	0.356	42	n. i.	n. i.

Theoretical monoisotopic masses are given for the unmodified peptide forms although methionine was sometimes oxidised in peptide parent ions used for fragmentation (cf. Supplementary Figures 9-13 and 9-19). Charges of parent ions are shown taking into account all MS² spectra which led to the identification of the peptide by MaxQuant. The mass error (Δm) of the parent peptide ions was calculated by MaxQuant subtracting the monoisotopic theoretical mass from the MaxQuant-recalibrated monoisotopic measured mass; Δm is indicated for that MS² spectrum of the peptide that obtained the best Mascot score. This Mascot score is stated in the neighbouring column. "n. i." denotes that the MaxQuant-Mascot search algorithms did not identify the peptide from the FT-Orbitrap MS² spectra. MHC anchor amino acid residues are highlighted in bold. Data referring to MS² spectra depicted in Supplementary Figures 9-1 to 9-20 are printed in blue. Asterisks indicate that the respective peptide was synthesised using $^{13}C_5{}^{15}N_1$-valine providing conclusive evidence for the peptide sequence (see section "2.4.4. Manual validation of MS2 spectra" for details).

Supplementary Table 9-4 | MS reliability parameters for identified urinary SAV peptides that are encoded in the genome of BALB/c but not B6 mice or vice versa (cf. Tables 3-4 and 3-7) (*Nat Comm*, Theo).

Gene	Peptide	Theoretical monoisotopic mass	Exp. no.	Charge	LTQ- MS²			FT-Orbitrap-MS²		
					Δm in ppm	Mascot Score	Andromeda Score	Δm in ppm	Mascot Score	Andromeda Score
Ces1c	SVFGAPLLKEGASEEETNL	1989.9895	1	2	-0.214	79	213	n. i.	n. i.	n. i.
Serpina3k	DVAETGTEAAAATGVIGGIRKAVL	2269.2278	3	3	-0.056	44(PD)	138	-0.056	15(PD)	28
	DVAETGTEAAAATGVIGGIRKAIL	2283.2434	1	2 or 3	-0.228	110	138	0.049	27	69
	II	II	2	3	0.241	34	n. p.	n. i.	n. i.	n. p.
	VAETGTEAAAATGVIGGIRKAI	2055.1324	1	3	-0.307	25	48	n. i.	n. i.	n. i.
	VAETGTEAAAATGVIGGIRKAIL	2168.2165	1	2 or 3	-0.197	97	148	-0.137	16(PD)	23
	II	II	2	3	0.272	55	n. p.	n. i.	n. i.	n. p.
	VAETGTEAAAATGVIGGIRKAILPAVHFNRP	3086.6989	1	4 or 5	0.087	29	75	0.044	21	72
	AAATGVIGGIRKAIL	1409.8718	1	2 or 3	0.004	44	118	n. i.	n. i.	n. i.
	AAATGVIGGIRKAILPAVHFNRP	2328.3543	1	3 or 4	0.170	53	134	0.012	23	45
	AATGVIGGIRKAIL	1338.8347	1	2	0.237	55	155	0.237	9	13
	AATGVIGGIRKAILPAVHFNRP	2257.3171	1	3, 4 or 5	-0.056	46	129	-0.045	14	48
	ATGVIGGIRKAIL	1267.7976	1	2 or 3	-0.116	41	110	-0.14(PD)	4(PD)	n. i.
	GVIGGIRKAIL	1095.7128	1*	2 or 3	0.625	28(PD)	89	n. i.	n. i.	n. i.
	GIRKAILPAVHFNRP	1687.9998	1	4	-0.192	35	108	n. i.	n. i.	n. i.
	IRKAILPAVHFNRP	1630.9784	1	4	-0.109	30	69	n. i.	n. i.	n. i.
	IRKAILPAVHFNRPF	1778.0468	1	3 or 4	0.059	50	123	n. i.	n. i.	n. i.
	IRKAILPAVHFNRPFL	1891.1308	1*	4	-0.221	26(PD)	80	-0.221	17(PD)	24
	KAILPAVHFNRPF	1508.8616	1	2, 3 or 4	-0.165	62	151	0.058	36	49
	II	II	2*	3	-0.22(PD)	37(PD)	n. p.	n. i.	n. i.	n. p.
	KAVLPAVcFNRPFL	1692.8844	1	3	-0.227	23(PD)	81	n. i.	n. i.	n. i.
	KAILPAVHFNRPFL	1621.9457	1	2, 3 or 4	-0.280	71	172	n. i.	n. i.	n. i.
	PAVcFNRP	1021.4474	3	2	0.877	12(PD)	89	n. i.	n. i.	n. i.
	PAVHFNRPF	1083.5614	1*	3	0.116	25(PD)	87	n. i.	n. i.	n. i.
	PAVcFNRPFL	1281.5998	1	2 or 3	-0.067	30(PD)	124	0.386	10(PD)	49
	II	II	2	2 or 3	1.18(PD)	47(PD)	n. p.	n. i.	n. i.	n. p.
	PAVHFNRPFL	1196.6455	1	2 or 3	0.149	45	151	0.238	6	40
	II	II	2	2 or 3	0.158	46	n. p.	0.42(PD)	7(PD)	n. p.

(Table continued on next page)

Supplementary Table 9-4 (continued)

Gene	Peptide	Theoretical monoisotopic mass	Exp. no.	Charge	LTQ- MS² Δm in ppm	LTQ- MS² Mascot Score	LTQ- MS² Andromeda Score	FT-Orbitrap-MS² Δm in ppm	FT-Orbitrap-MS² Mascot Score	FT-Orbitrap-MS² Andromeda Score
Serpina3k	IVIYHTSAQSIL	1343.7449	1	2	-0.020	28$^{(PD)}$	103	n. i.	n. i.	n. i.
	II	II	2	2	0.84$^{(PD)}$	35$^{(PD)}$	n. p.	n. i.	n. i.	n. p.
	FVIYHTSAQSIL	1377.7293	1	2	0.116	44	184	n. i.	n. i.	n. i.

Amino acid residues that differ between B6 and BALB/c mice are underlined. All cysteine residues occurred in a cysteinylated form, i.e. they had bound free cysteine via a disulfide bridge (cf. Supplementary Figures 9-25 and 9-27) resulting in an additional theoretical parent ion mass of 119.004099 Da included in the indicated theoretical monoisotopic masses. These modified cysteine residues are symbolised by the lower case letter "c".

SAV peptide forms not contained in the IPI *Mus musculus* version 3.64 and thus not identified during first round MS data processing with MaxQuant version 1.0.14.3 employing Mascot (cf. section "2.4. Mass spectrometric data analyses") are shaded in green. On the other hand, peptides labelled with an asterisk in the column "Exp. no." are contained in the IPI *Mus musculus* version 3.64 but were not identified or rejected in MaxQuant 1.0.14.3 data processing, e.g. because they exceeded the PEP threshold[242] for an FDR of 1% or obtained a Mascot score of < 25. Both kinds of peptides are neither included in the total number of the 639 identified peptides nor in the number of identified SAV peptides to avoid double counting of SAV peptide pairs and a distortion of the calculated proportion of SAV peptides among all urinary peptides. Nevertheless, they are reported in this table and in Table 3-7 due to their identification with either an Andromeda score of ≥ 80 in MaxQuant version 1.2.2.5 or a Mascot score of ≥ 25 in Proteome Discoverer. In all reported cases, the SAV-form of the asterisk-labelled peptides was validated by another, higher scoring peptide thus adding confidence to these borderline score peptides.

Charges of parent ions are shown taking into account all MS² spectra which led to the identification of the peptide by MaxQuant or Proteome Discoverer. The mass error (Δm) of the parent peptide ions was calculated by MaxQuant subtracting the monoisotopic theoretical mass from the MaxQuant-recalibrated monoisotopic measured mass; Δm is indicated for that MS² spectrum of the peptide that obtained the best Mascot score (best Andromeda score if no MaxQuant derived Mascot score is amenable). Only if no MaxQuant data are available for a certain peptide, mass errors and Mascot scores were taken from Proteome Discoverer designated with " $^{(PD)}$ ". The best Mascot scores and the best Andromeda scores of each peptide are indicated for the respective experiments but do not necessarily descend from the same MS² spectrum. "n. i." denotes that the MaxQuant or Proteome Discoverer search algorithms did not identify the peptide from the FT-Orbitrap MS² spectra. "n. p." indicates that no Andromeda score is available because the raw files from experiment no. 2 could not be processed with MaxQuant version 1.2.2.5 and 1.3.0.5 due to a bug in the program. Data referring to MS² spectra depicted in Supplementary Figures 9-21 to 9-32 are printed in blue.

Supplementary Table 9-5 | MS reliability parameters for unique (i.e. encoded by only one gene) urinary SAV peptides that are encoded in the genome of BALB/c and B6 mice but not in certain other laboratory or wild-derived mouse strains (cf. Table 3-4) (*Nat Comm*, Theo).

Gene	Peptide	Theoretical monoisotopic mass	Exp. no.	Charge	LTQ-MS²		FT-Orbitrap-MS²	
					Δm in ppm	Mascot Score	Δm in ppm	Mascot Score
Abca13	LSNPSGSFPAPEFDNL	1690.7839	1	2	0.435	59	n. i.	n. i.
Alb	EAHKSEIAHRY	1339.6633	1	3 or 4	-0.024	37	n. i.	n. i.
	EAHKSEIAHRYND	1568.7332	1	3	-0.139	36	n. i.	n. i.
	EAHKSEIAHRYNDL	1681.8172	1	3	-0.050	30	n. i.	n. i.
	EAHKSEIAHRYNDLGEQHFKGLVL	2790.4202	1	2, 3, 4 or 5	-0.294	33	0.070	20
	IAHRYNDLGEQHFKGLVL	2109.1120	1	3 or 4	-0.123	44	n. i.	n. i.
	AHRYNDLGEQHFKGLVL	1996.0279	1	4	-0.118	26	n. i.	n. i.
	RYNDLGEQHFKGLVL	1787.9319	1	3	-0.172	53	0.067	11
	YNDLGEQHFKGLVL	1631.8308	1	2 or 3	0.224	57	n. i.	n. i.
	NDLGEQHFKGLVL	1468.7674	1	2 or 3	-0.048	68	n. i.	n. i.
	GEQHFKGLVL	1126.6135	1	2 or 3	-0.532	34	0.234	11
	LVKHKPKATAEQLKTVMDDF	2298.2406	1	3	-0.197	49	n. i.	n. i.
Egf	LYWVDVERQV	*see Supplementary Table 9-3 and Supplementary Figure 9-11 (MHC motif peptide)*						
	YWVDVERQVL	1305.6717	2	2	0.105	38	n. i.	n. i.
	‖	‖	3	2	0.104	21	0.104	12
	WVDVERQVL	1142.6084	1	2	0.188	45	n. i.	n. i.
	‖	‖	2	2	0.160	44	-0.031	23
Kap	LLNSVFDQLGSY	1354.6769	1	2	0.148	30	n. i.	n. i.
	LNSVFDQL	*see Supplementary Table 9-3 and Supplementary Figure 9-2 (MHC motif peptide)*						
	LNSVFDQLG	991.4975	1	2	0.151	50	n. i.	n. i.
	LNSVFDQLGSY	*see Supplementary Table 9-3 and Supplementary Figure 9-4*						
	LNSVFDQLGSYR	1397.6939	1	2	0.041	88	n. i.	n. i.
	LNSVFDQLGSYRGTKAPLED	2209.1015	1	3	0.043	56	n. i.	n. i.
	NSVFDQLGSY	1128.5088	1	2	0.143	51	n. i.	n. i.
	‖	‖	2	2	-0.069	52	n. i.	n. i.
	SVFDQLGSY	1014.4658	1	2	0.036	40	n. i.	n. i.

(Table continued on next page)

Supplementary Table 9-5 (continued)

Gene	Peptide	Theoretical monoisotopic mass	Exp. no.	Charge	LTQ-MS²		FT-Orbitrap-MS²	
					Δm in ppm	Mascot Score	Δm in ppm	Mascot Score
Lrg1	SSNRLQALSPELLAPVPRLRA	2287.3124	3	4	-0.274	10	-0.199	28
	LQALSPELLAPVPR	1502.8821	1	2	0.034	70	n. i.	n. i.
	II	II	2	2	0.032	73	n. i.	n. i.
	QALSPELLAPVPRL	1502.8821	1	2	-0.013	59	n. i.	n. i.
	II	II	2	2	0.585	69	n. i.	n. i.
	SPELLAPVPRL	1190.7023	3	2	0.402	45	0.402	30
	LQGLDALGHLDLAE	1463.7620	1	2	-0.072	52	n. i.	n. i.
Prol1	GFIPSSPKFP	1075.5702	2	2	0.062	31	n. i.	n. i.
Serpina1b	VQIHIPRLSISGDYN	1710.9053	1	3	0.079	25	n. i.	n. i.

Amino acid residues with a predicted single amino acid variation (SAV) in certain laboratory or wild-derived mouse strains are underlined. Charges of parent ions are shown taking into account all MS² spectra which led to the identification of the peptide by MaxQuant. The mass error (Δm) of the parent peptide ions was calculated by MaxQuant subtracting the monoisotopic theoretical mass from the MaxQuant-recalibrated monoisotopic measured mass; Δm is indicated for that MS² spectrum of the peptide that obtained the best Mascot score. This Mascot score is stated in the neighbouring column. "n. i." denotes that the MaxQuant-Mascot search algorithms did not identify the peptide from the FT-Orbitrap MS² spectra. For each SAV reported in this table, we validated the shortest identified peptide containing this SAV by manual inspection of the MS^2 spectra. Because the identification of a peptide does not influence the discovery of its longer or shorter forms in bioinformatic MS data processing, the occurrence of a core peptide in several length variations is by itself a good validation for the correctness of the peptide sequences. The manual validation of at least one member of each group of overlapping peptide sequences adds considerable evidence for all peptides of this group. MS^2 spectra of the manually validated SAV peptides for which the Mouse Phenome Database (MPD) contains information on SNV allele distribution (cf. Tables 3-4 and 3-6) are depicted in Supplementary Figures 9-21 to 9-40. SAVs with this information in the MPD have a high reliability at the underlying genomic level, at least if the rarer SAV form was identified in more than one mouse strain. For SAVs without this MPD information, the error rate at the genomic level appears to be higher (cf. section "2.4.2. Prediction of SAV peptides") than that at the level of our identified urinary peptides. Data referring to MS² spectra depicted in Supplementary Figures 9-33 to 9-40 are printed in blue.

Supplementary Table 9-6 | MS reliability parameters for unique MUP20 derived peptides (cf. Table 3-8) (*Nat Comm*, Theo).

Peptide	Theoretical monoisotopic mass	Exp. no.	Charge	LTQ-MS²		FT-Orbitrap-MS²	
				Δm in ppm	Mascot Score	Δm in ppm	Mascot Score
EEASSMERNF	1198.4925	1	2	0.149	53	n. i.	n. i.
NVEKINGEWYTIM	1595.7654	1	2	-0.019	105	n. i.	n. i.
VEYIHVLENSL	1314.6820	1	2	-0.119	32	n. i.	n. i.
"	"	2	2	0.249	44	n. i.	n. i.
"	"	3	2	-0.395	60	-0.395	18
IHVLENSLALKF	1382.7922	1	2	0.093	33	n. i.	n. i.
IMIHLINKKDGETFQL	1899.0288	1	2, 3 or 4	0.042	47	n. i.	n. i.
LINKKDGETFQL	1404.7613	1	2	-0.122	45	n. i.	n. i.

Theoretical monoisotopic masses are given for the unmodified peptide forms although methionine was sometimes oxidised in peptide parent ions used for fragmentation. Charges of parent ions are shown taking into account all MS² spectra which led to the identification of the peptide by MaxQuant. The mass error (Δm) of the parent peptide ions was calculated by MaxQuant subtracting the monoisotopic theoretical mass from the MaxQuant-recalibrated monoisotopic measured mass; Δm is indicated for that MS² spectrum of the peptide that obtained the best Mascot score. This Mascot score is stated in the neighbouring column. The MS^2 spectra of the indicated MUP20 peptides have not been controlled manually. However, because of several overlapping peptide sequences, the applied false discovery rate (FDR) of 1% in target-decoy database searches and the removal of all peptides with Mascot scores below 25, the number of incorrectly identified MUP20 peptides is anticipated to be zero. "n. i." denotes that the MaxQuant-Mascot search algorithms did not identify the peptide from the FT-Orbitrap MS² spectra.

Supplementary Table 9-7 | MS reliability parameters for peptides derived from the meprin A α protein sequence (*Nat Comm*, Theo).

Peptide	Theoretical monoisotopic mass	Exp. no.	Charge	LTQ-MS²		FT-Orbitrap-MS²	
				Δm in ppm	Mascot Score	Δm in ppm	Mascot Score
NLFQGDILLPR	1284.7190	1	2	0.066	47	n. i.	n. i.
FQGDILLPR	1057.5920	1	2	0.003	63	0.160	37
‖	‖	2	2	-0.029	48	-0.029	7
ADNLELNAKGAILHAFE	1824.9370	2	3	0.031	45	n. i.	n. i.
ADNLELNAKGAILHAFEM	1955.9775	2	3	0.081	26	n. i.	n. i.
ELNAKGAILHAFE	1411.7460	1	2 or 3	-0.117	46	n. i.	n. i.
‖	‖	2	2 or 3	-0.104	63	n. i.	n. i.
ELNAKGAILHAFEM	1542.7864	2	2	-0.089	37	n. i.	n. i.
NAKGAILHAFEM	1300.6598	1	2 or 3	0.339	47	n. i.	n. i.
AKGAILHAFEM	1186.6169	2	2	0.099	59	n. i.	n. i.
GAILHAFEM	1003.4797	2	2	-0.328	47	n. i.	n. i.
NTIIGQLPDF	1116.5815	1	2	0.051	36	n. i.	n. i.
‖	‖	2	2	0.003	50	n. i.	n. i.
NTIIGQLPDFS	1203.6136	1	2	0.235	32	n. i.	n. i.
NTIIGQLPDFSA	1274.6507	1	2	0.232	42	n. i.	n. i.
NTIIGQLPDFSAID	1502.7617	1	2	0.043	40	n. i.	n. i.
‖	‖	2	2	0.286	54	n. i.	n. i.
YNSEGYGVGVTLYPNGRITSNSGFL	2664.2820	1	3	-0.170	38	n. i.	n. i.
VTLYPNGRITSNSGFL	1737.9050	1	2	-0.028	54	n. i.	n. i.
TLYPNGRITSNSGFL	1638.8366	1	2	0.225	54	n. i.	n. i.
PNGRITSNSGFLGL	1431.7470	2	2	0.036	49	n. i.	n. i.
HLYSGDNDAILEWPVENRQAI	2439.1819	2	3	-0.171	51	n. i.	n. i.
YSGDNDAILEWPVENRQAI	2189.0389	3	3	-0.195	30	-0.195	4
YSGDNDAILEWPVENRQAIM	2320.0794	2	3	0.195	34	n. i.	n. i.
AILEWPVENRQAI	1537.8253	1	2	0.095	56	n. i.	n. i.

Theoretical monoisotopic masses are given for the unmodified peptide forms although methionine was sometimes oxidised in peptide parent ions used for fragmentation. Charges of parent ions are shown taking into account all MS² spectra which led to the identification of the peptide by MaxQuant. The mass error (Δm) of the parent peptide ions was calculated by MaxQuant subtracting the monoisotopic theoretical mass from the MaxQuant-recalibrated monoisotopic measured mass; Δm is indicated for that MS² spectrum of the peptide that obtained the best Mascot score. This Mascot score is stated in the neighbouring column. The MS^2 spectra of the indicated meprin A α peptides have not been controlled manually. However, because of the many overlapping sequences, the applied false discovery rate (FDR) of 1% in target-decoy database searches and the removal of all peptides with Mascot scores below 25, the number of incorrectly identified meprin A α peptides is anticipated to be zero or very low leaving the percentage of meprin A α peptides among all identified urinary peptides essentially unchanged. “n. i.” denotes that the MaxQuant-Mascot search algorithms did not identify the peptide from the FT-Orbitrap MS² spectra.

Supplementary Note to MS^2 spectra

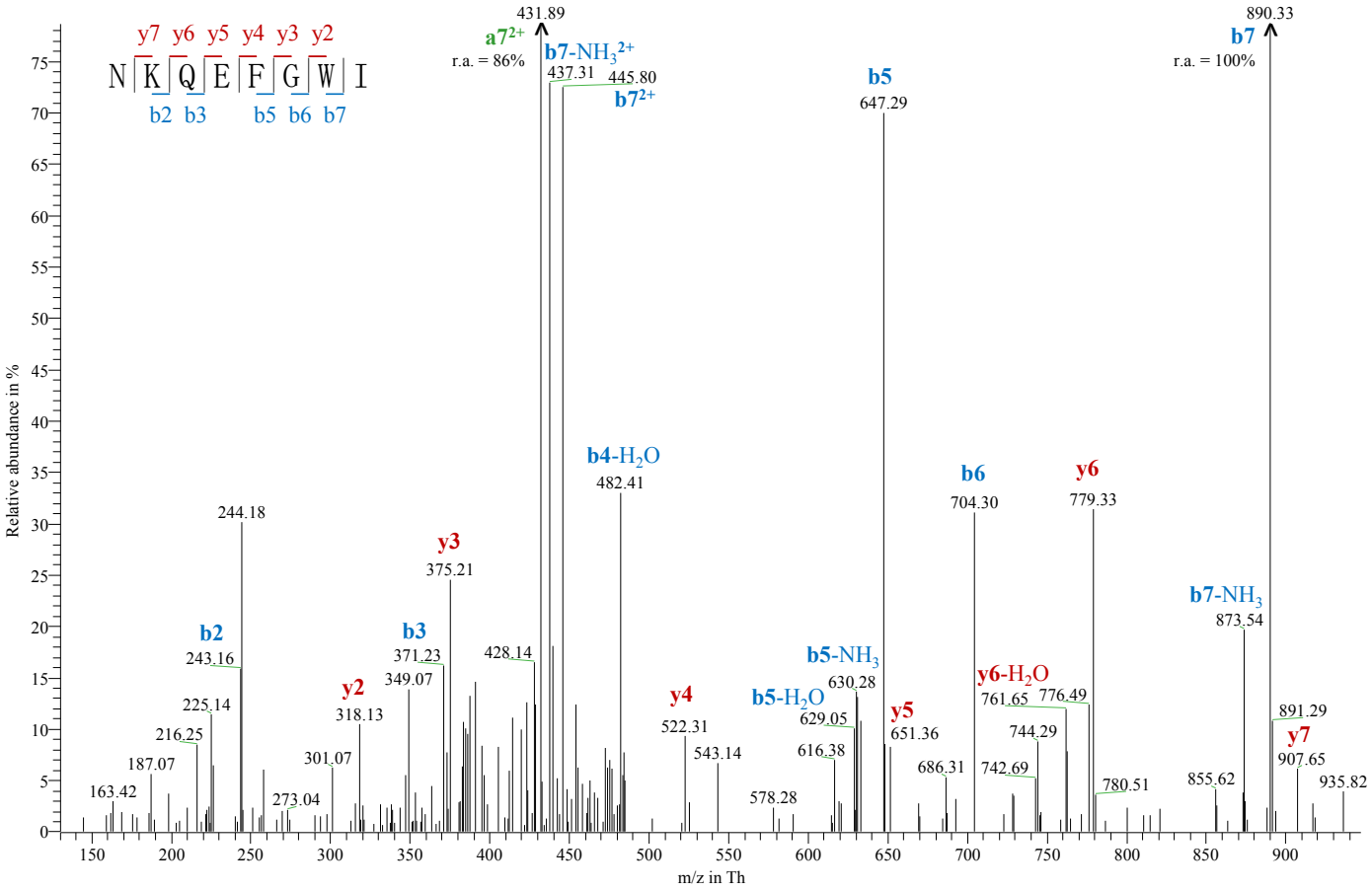
I selected the MS^2 spectra of the following peptides for display (*Nat Comm*, Theo):

- Supplementary Figures 9-1 to 9-20: *The 17 MHC motif peptides and three extended forms thereof (LNSVFDQLGSY, IDQTRVLNLGPI and SGNFIDQTRVLNLGPITR) identified in mouse urine.* Spectra of MHC motif peptides are given in the same order as in Supplementary Table 9-3.

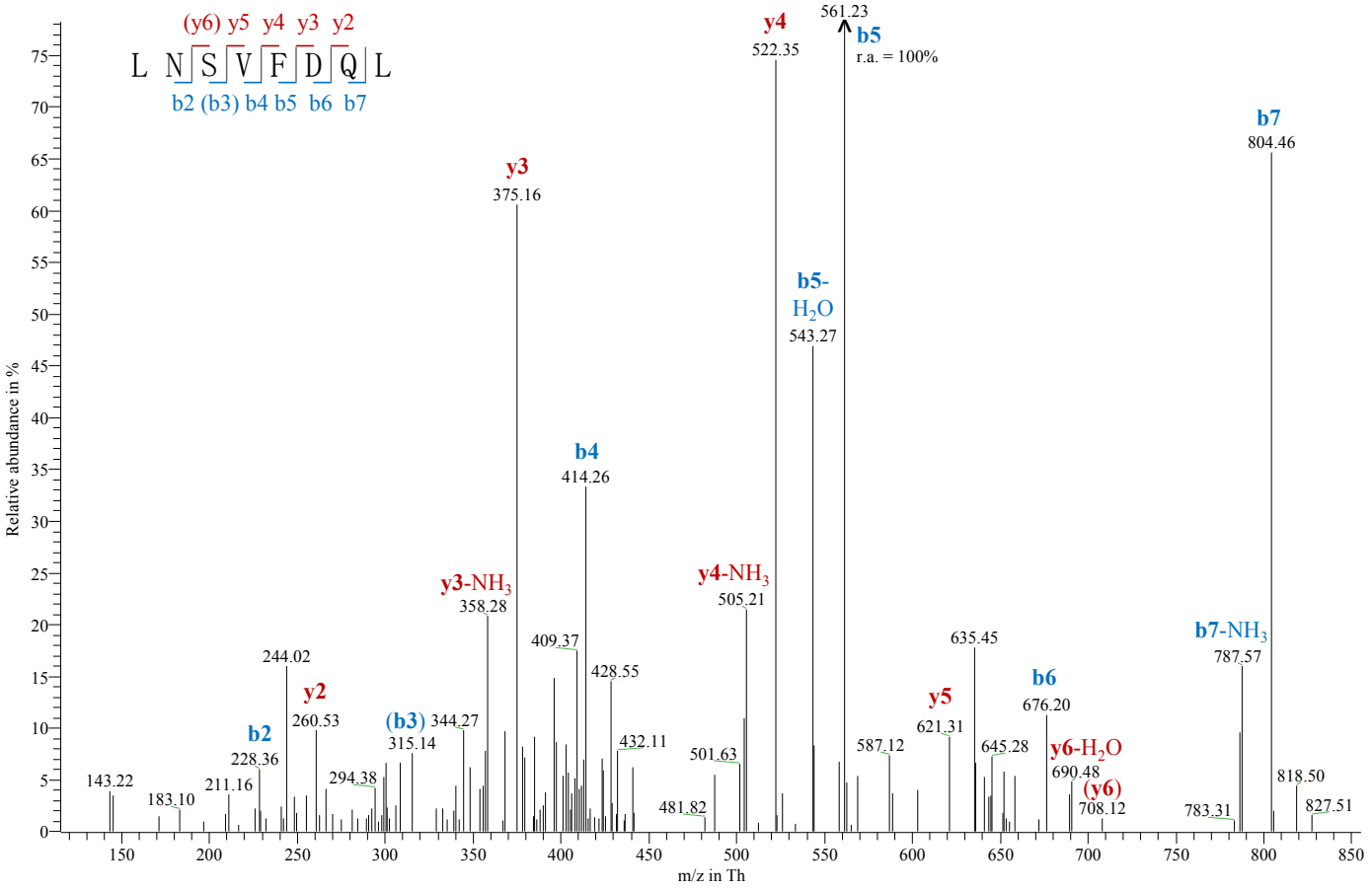
- Supplementary Figures 9-21 to 9-32: *Selected urinary SAV peptides that are encoded in the genome of BALB/c but not B6 mice or vice versa.* Each SAV is substantiated by at least one MS^2 spectrum. Where possible, I choose peptides for MS^2 spectra depiction that represent identified SAV peptide pairs of equal length in B6 and BALB/c mice. Spectra of SAV peptides are given in the same order as in Tables 3-4 and 3-7.

- Supplementary Figures 9-33 to 9-40: *Selected unique (i.e. encoded by only one gene) urinary SAV peptides that are encoded in the genome of BALB/c and B6 mice but not in certain other laboratory or wild-derived mouse strains.* Each SAV for which the Mouse Phenome Database (MPD) contains information on SNV allele distribution (cf. Tables 3-4 and 3-6) is substantiated by at least one MS^2 spectrum. For each SAV, we depict the MS^2 spectrum of the shortest identified peptide containing this SAV (cf. Supplementary Table 9-5). Spectra of SAV peptides are given in the same order as in Supplementary Table 9-5.

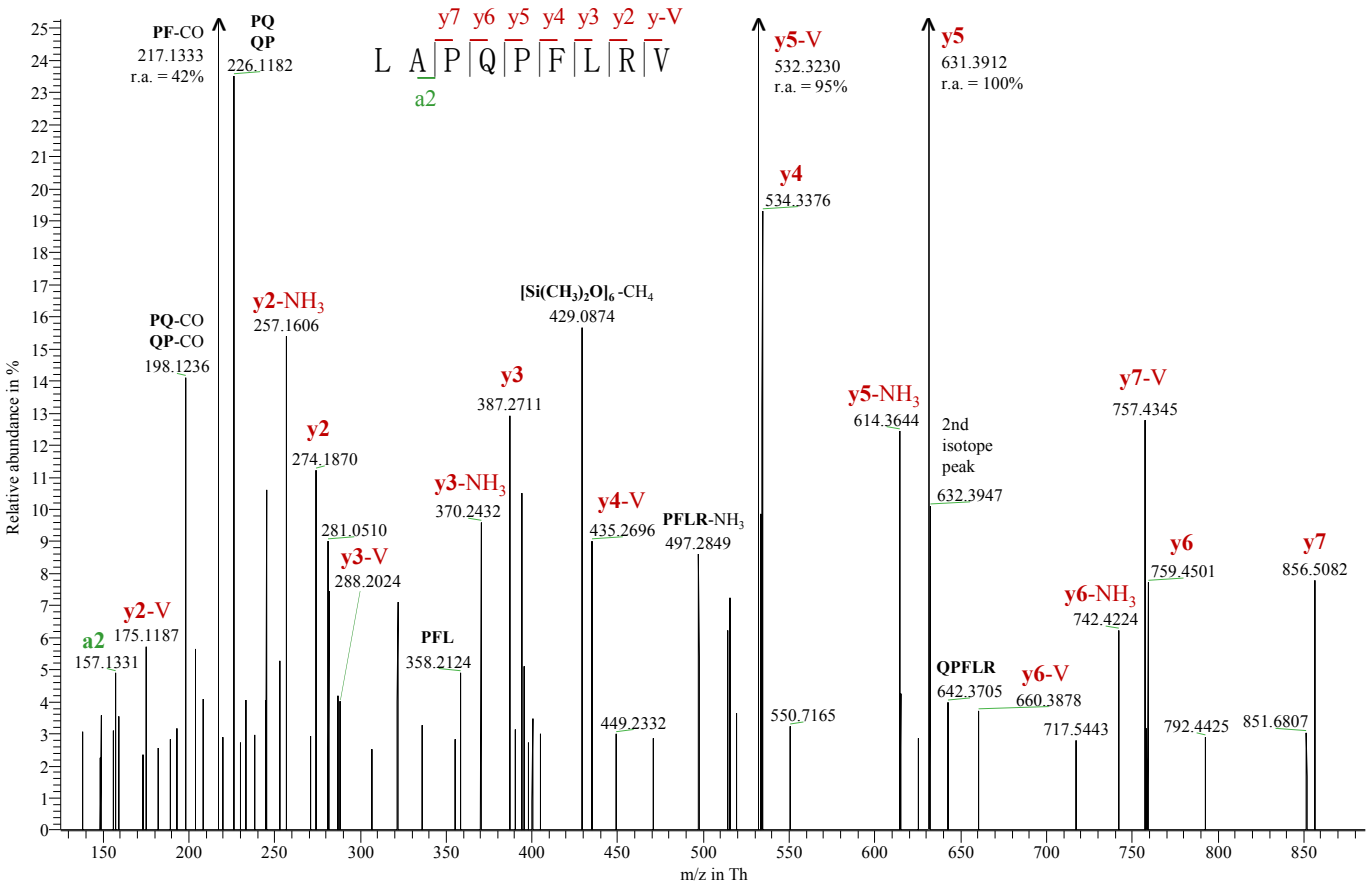
For every selected peptide, the MS^2 spectrum with the best Mascot score (best Andromeda score if no MaxQuant derived Mascot score is amenable) is shown. Only if a Fourier transform (FT)-Orbitrap-MS^2 spectrum with a Mascot score above 30 was available, I decided to show this spectrum instead of the respective higher scoring but lower mass accuracy linear trap quadrupole (LTQ)-MS^2 spectrum. Fragment ions were annotated manually allowing 0.5 Th mass deviation for MS^2 spectra obtained in the LTQ and 10 ppm for those from FT-Orbitrap measurements. In doing so, I took into account that the LTQ resolves isotope peaks only occasionally and that the monoisotopic peak does not constitute the main signal for very large fragment ions. Due to the difference in mass accuracy, we decided to annotate internal fragment ions only in FT-Orbitrap-MS^2 spectra but not in LTQ-MS^2 spectra in order to reduce the number of false positive annotations − with the exception of IVIYHTSAQSIL and FVIYHTSAQSIL where prove for internal fragments was achieved by synthetic counterparts (Supplementary Figures 9-29 to 9-32). If not mentioned otherwise, all annotated fragment ions have a single positive charge due to a bound H^+ ion. In FT-Orbitrap-MS^2 spectra, the ion $[Si(CH_3)_2O]_6H^+$-CH_4 (m/z = 429.0887) is occasionally observed originating from the loss of methane from $[Si(CH_3)_2O]_6H^+$ (polycyclodimethylsiloxane, PCM-6, m/z = 445.1200) used as lock mass for internal calibration[241]. Arrows at the top of peaks indicate that the respective signal is not shown in full high to enable a better visualisation of lower intensity peaks. In such cases the measured relative abundance (r.a.) is displayed next to the signals. Fragment ions given in parentheses refer to signals raised only slightly above background noise (*Nat Comm*, Theo).



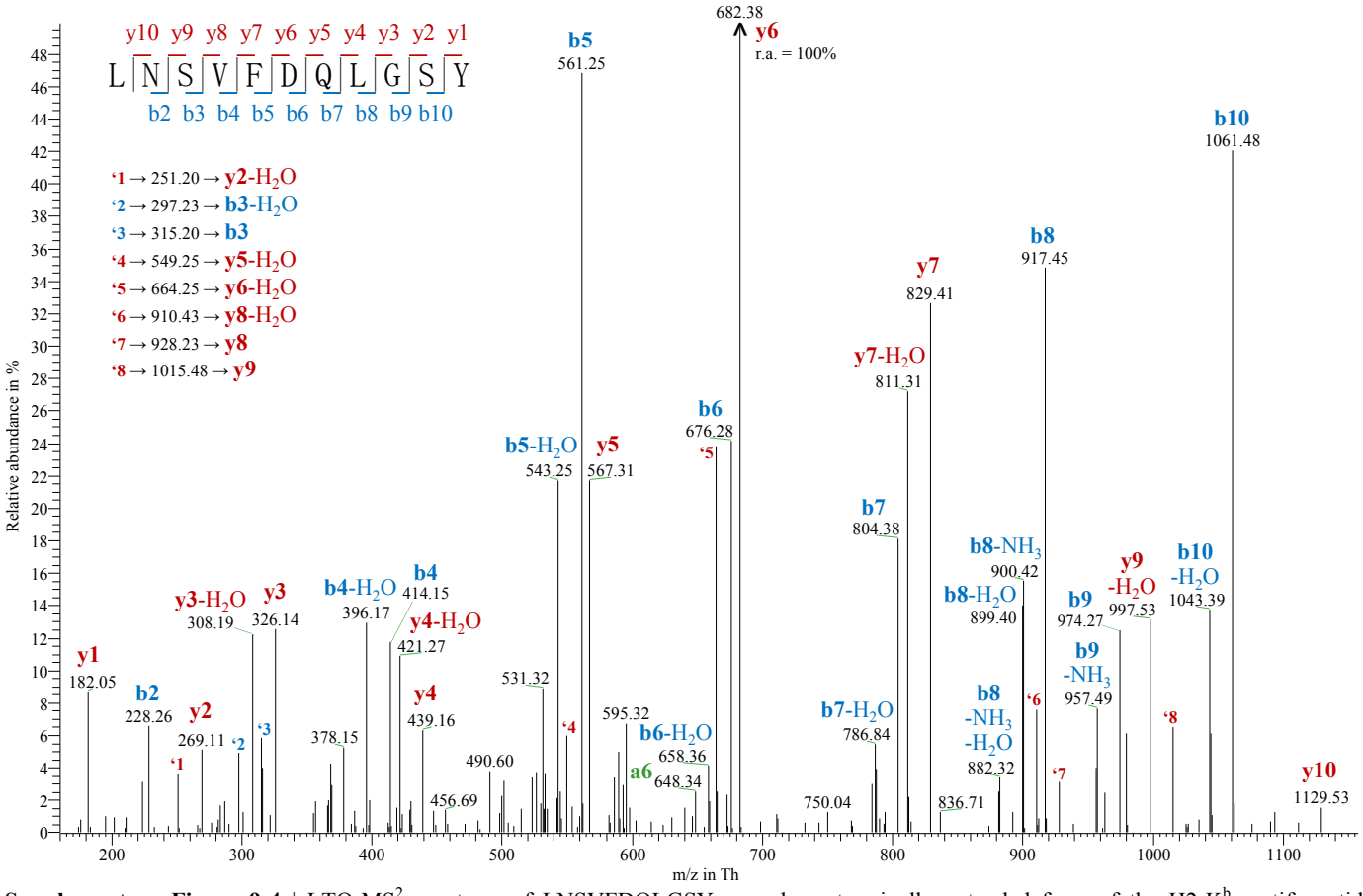
Supplementary Figure 9-1 | LTQ-MS^2 spectrum of NKQEFGWI, an H2-K^b motif peptide. See also Supplementary Note (*Nat Comm*, Theo).



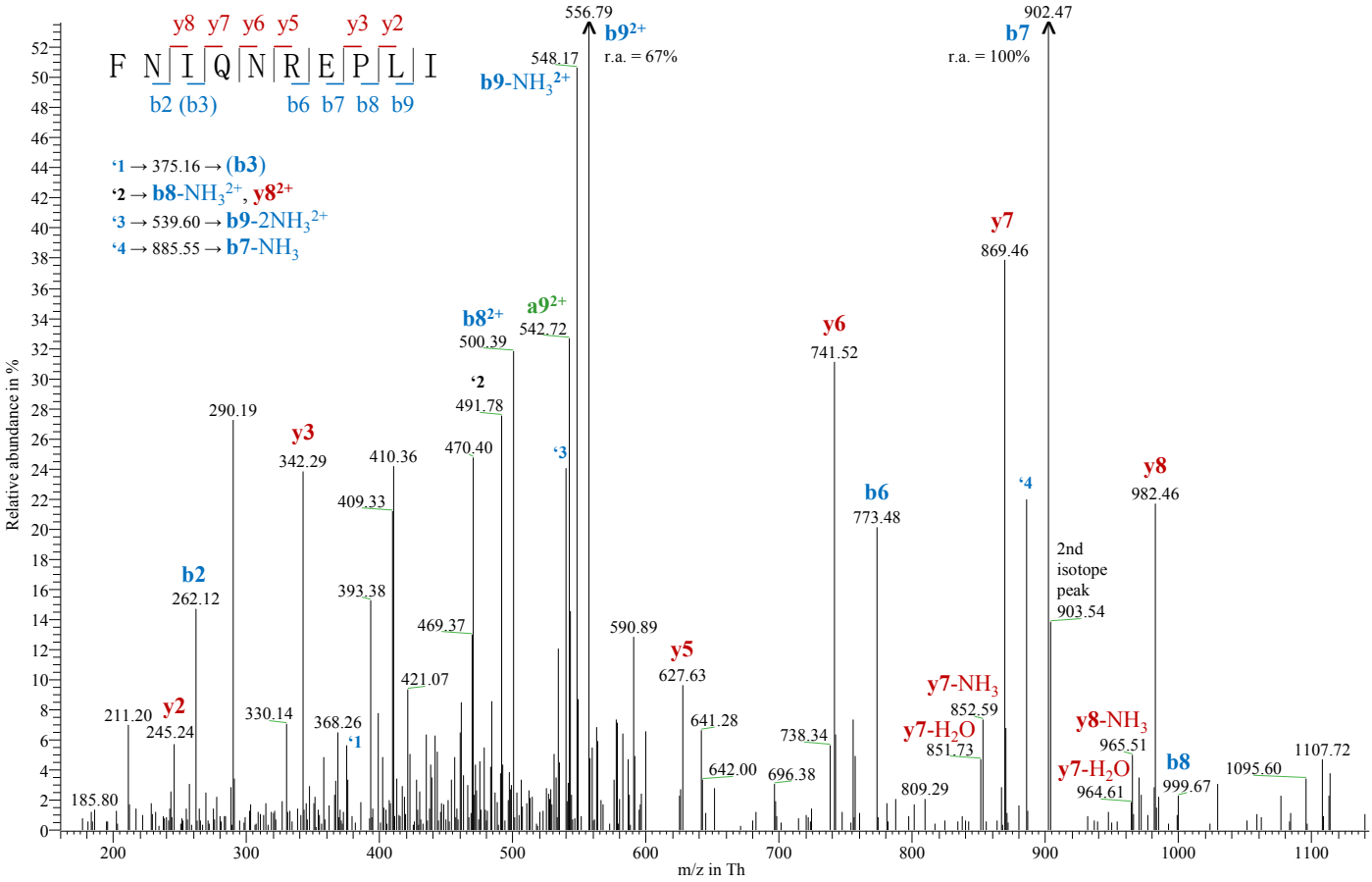
Supplementary Figure 9-2 | LTQ-MS^2 spectrum of LNSVFDQL, an H2-K^b motif peptide. We obtained final evidence for this peptide by comparing MS^2 spectra as well as nano-HPLC elution times with those of the synthetic $^{13}C_5{}^{15}N_1$-valine labelled synthetic counterpart. See also Supplementary Note (*Nat Comm*, Theo).



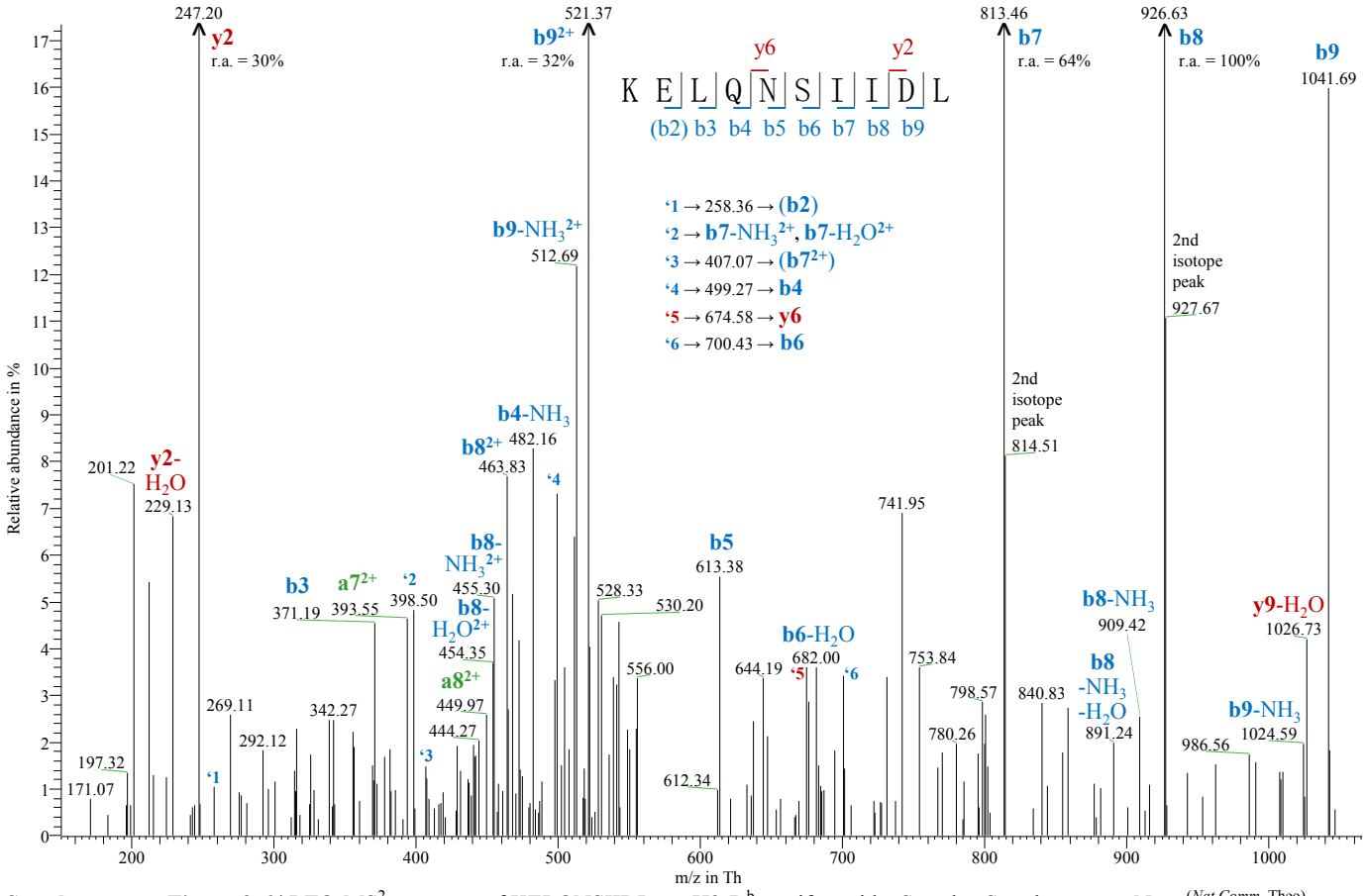
Supplementary Figure 9-3 | FT-Orbitrap-MS^2 spectrum of LAPQPFLRV, an H2-K^b motif peptide. See also Supplementary Note (*Nat Comm*, Theo).



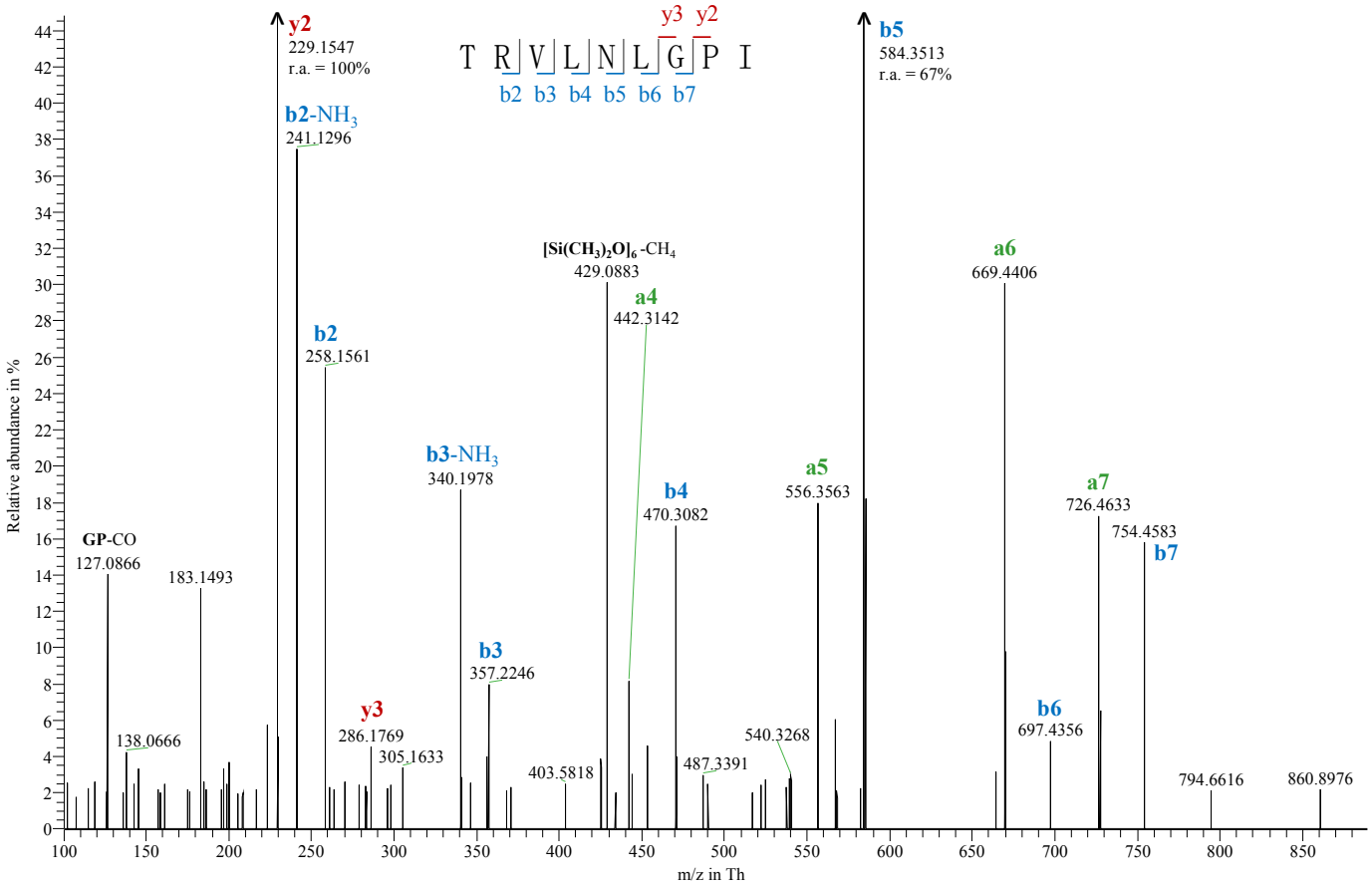
Supplementary Figure 9-4 | LTQ-MS^2 spectrum of LNSVFDQLGSY, a carboxy-terminally extended form of the H2-K^b motif peptide LNSVFDQL. We obtained final evidence for this peptide by comparing MS^2 spectra as well as nano-HPLC elution times with those of the synthetic $^{13}C_5{}^{15}N_1$-valine labelled synthetic counterpart. See also Supplementary Note (*Nat Comm*, Theo).



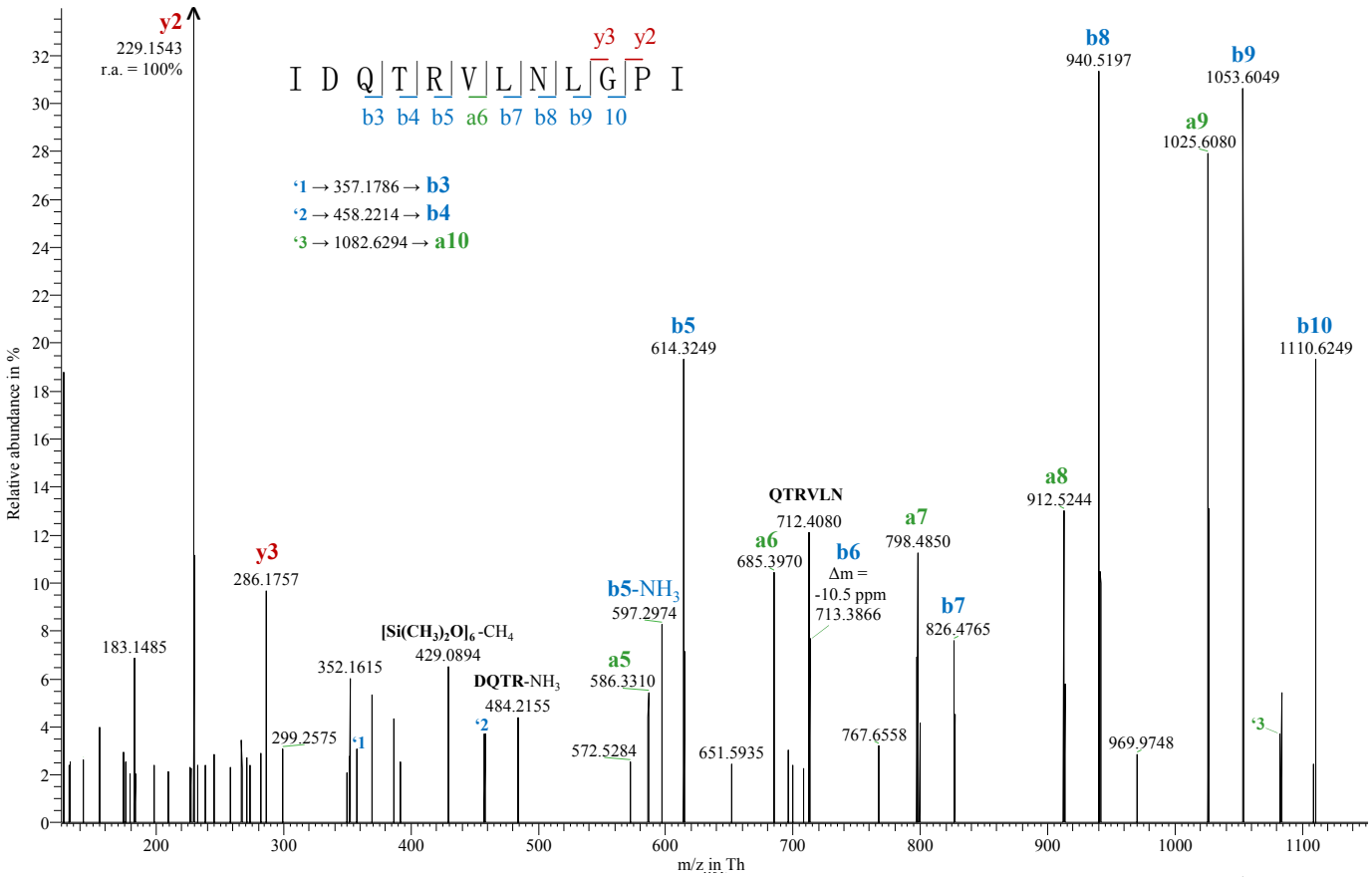
Supplementary Figure 9-5 | LTQ-MS^2 spectrum of FNIQNREPLI, an H2-D^b motif peptide. See also Supplementary Note (*Nat Comm*, Theo).



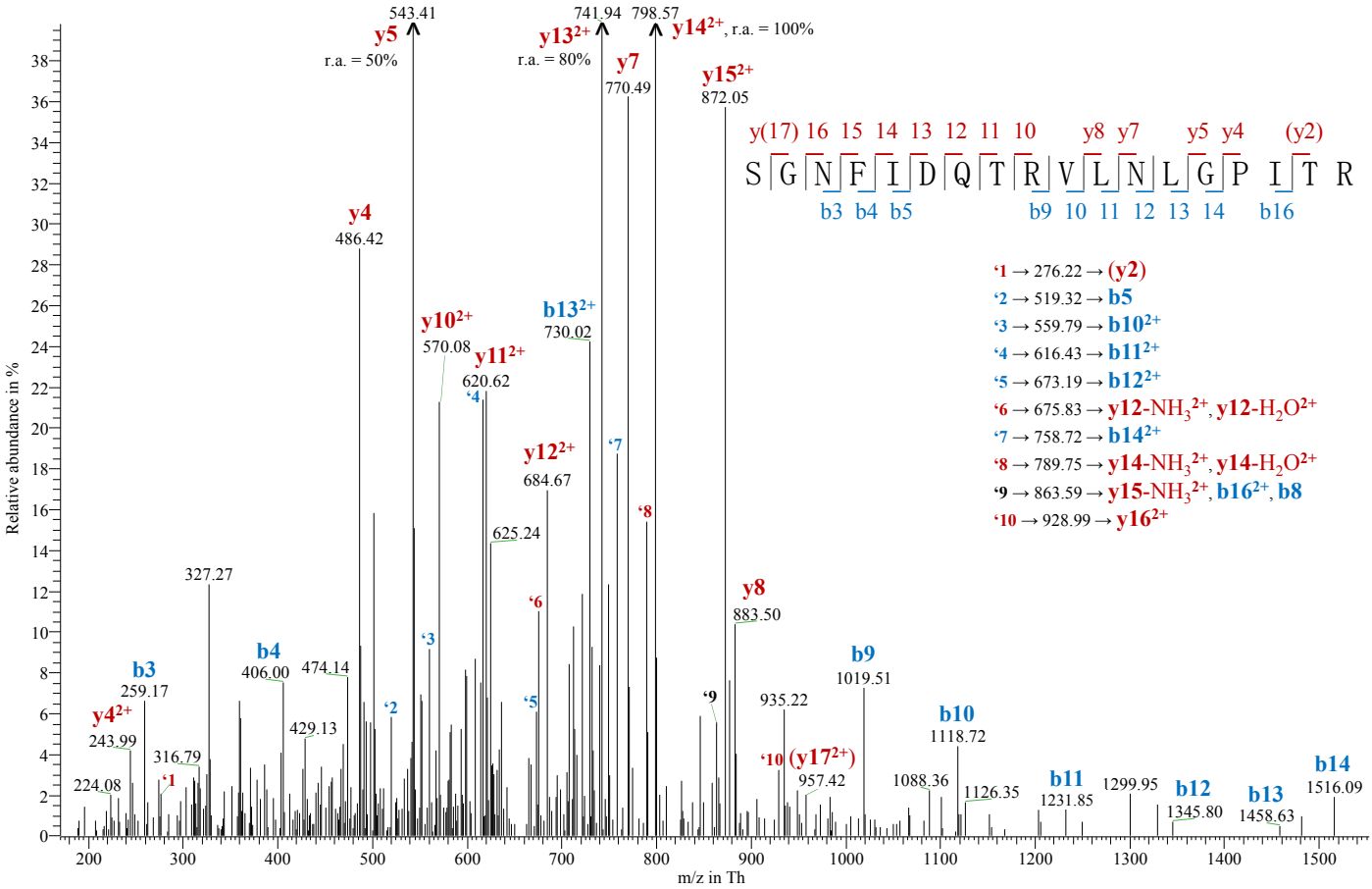
Supplementary Figure 9-6 | LTQ-MS^2 spectrum of KELQNSIIDL, an H2-D^b motif peptide. See also Supplementary Note (*Nat Comm*, Theo).



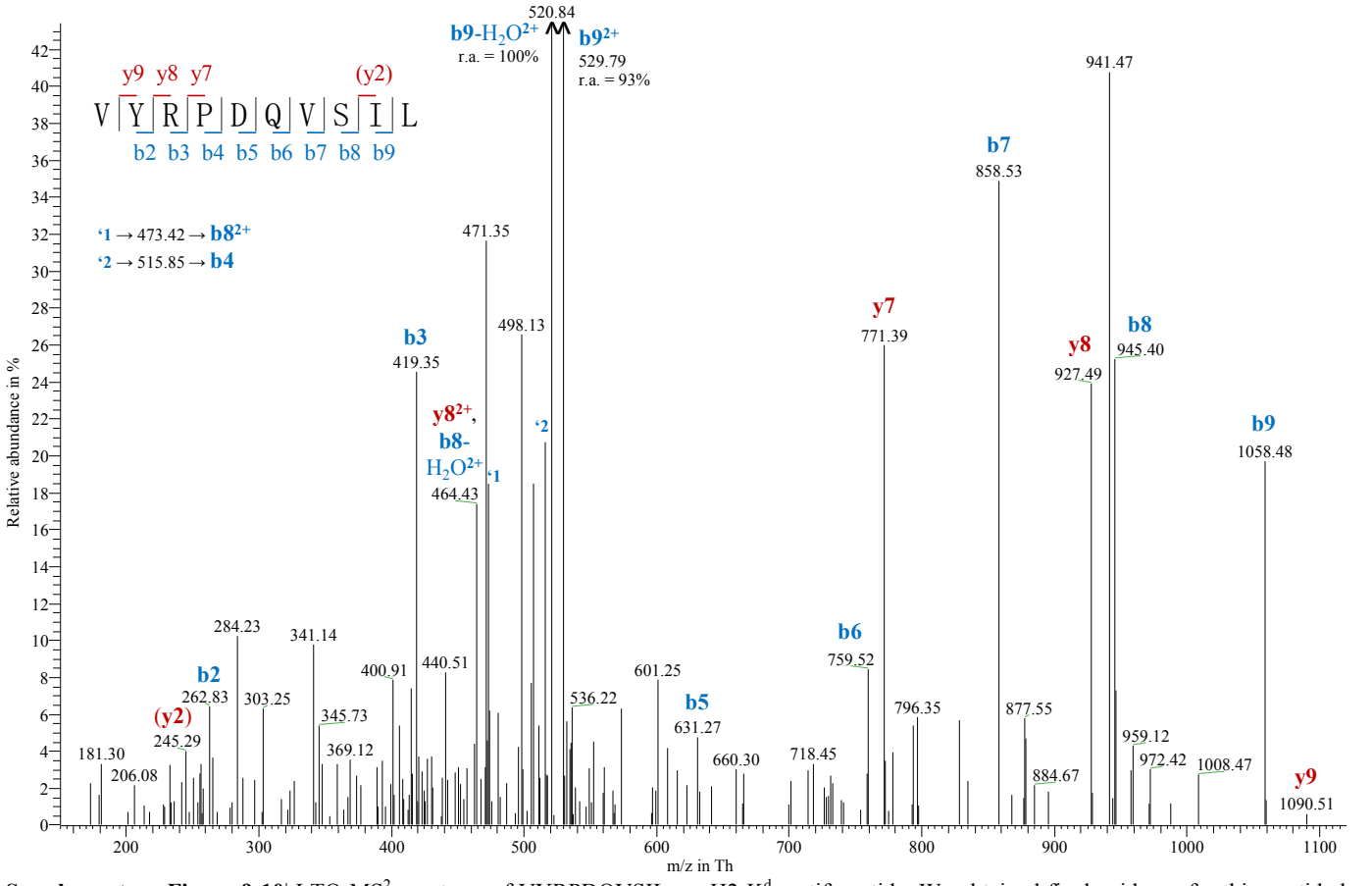
Supplementary Figure 9-7 | FT-Orbitrap-MS^2 spectrum of TRVLNLGPI, an H2-D^b motif peptide. We obtained final evidence for this peptide by comparing MS^2 spectra as well as nano-HPLC elution times with those of the synthetic $^{13}C_5{}^{15}N_1$-valine labelled synthetic counterpart. See also Supplementary Note (*Nat Comm*, Theo).



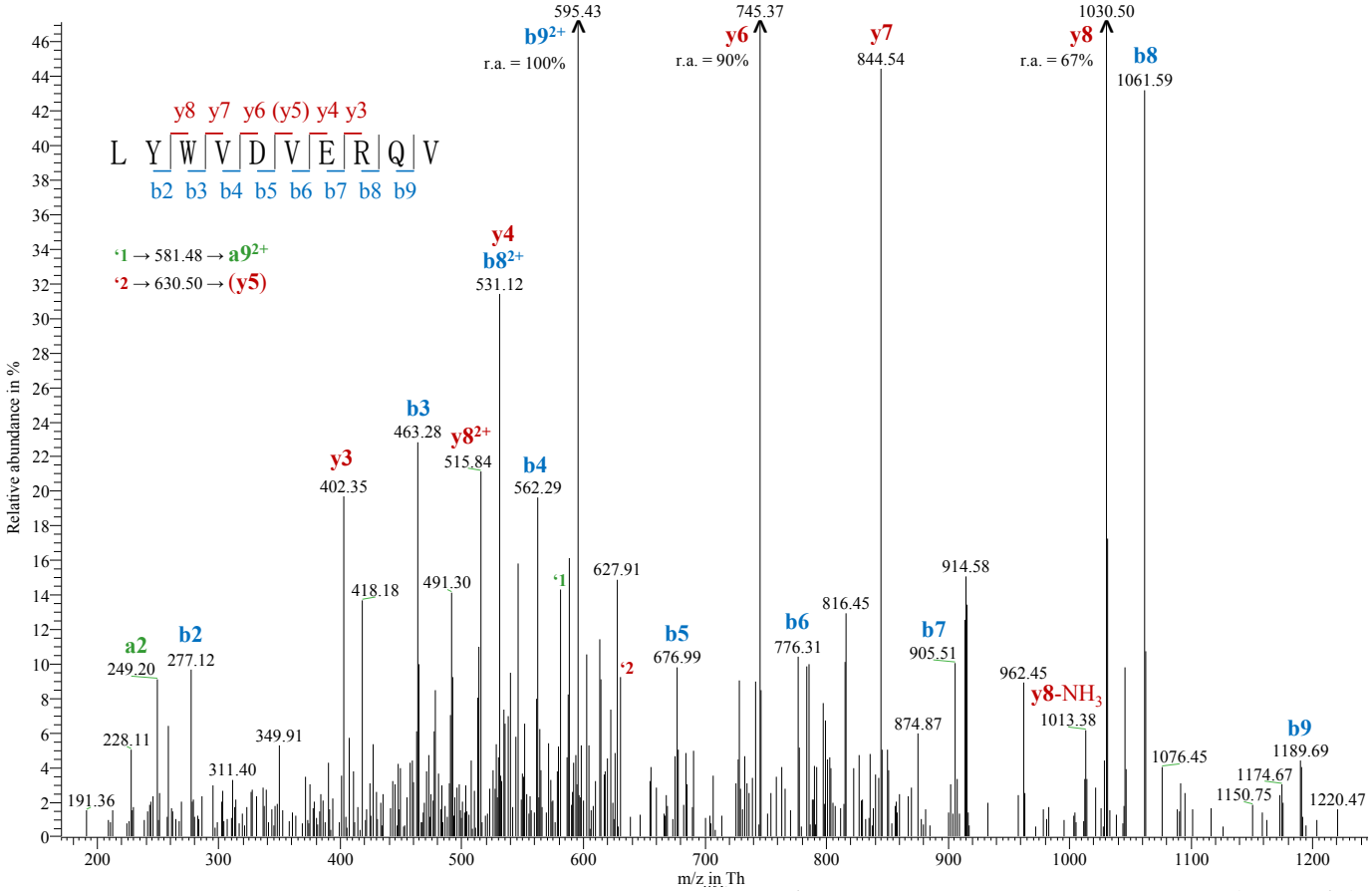
Supplementary Figure 9-8| FT-Orbitrap-MS^2 spectrum of IDQTRVLNLGPI, an amino-terminally extended form of the H2-D^b motif peptide TRVLNLGPI. We obtained final evidence for this peptide by comparing MS^2 spectra as well as nano-HPLC elution times with those of the synthetic $^{13}C_5{}^{15}N_1$-valine labelled synthetic counterpart. See also Supplementary Note (*Nat Comm*, Theo).



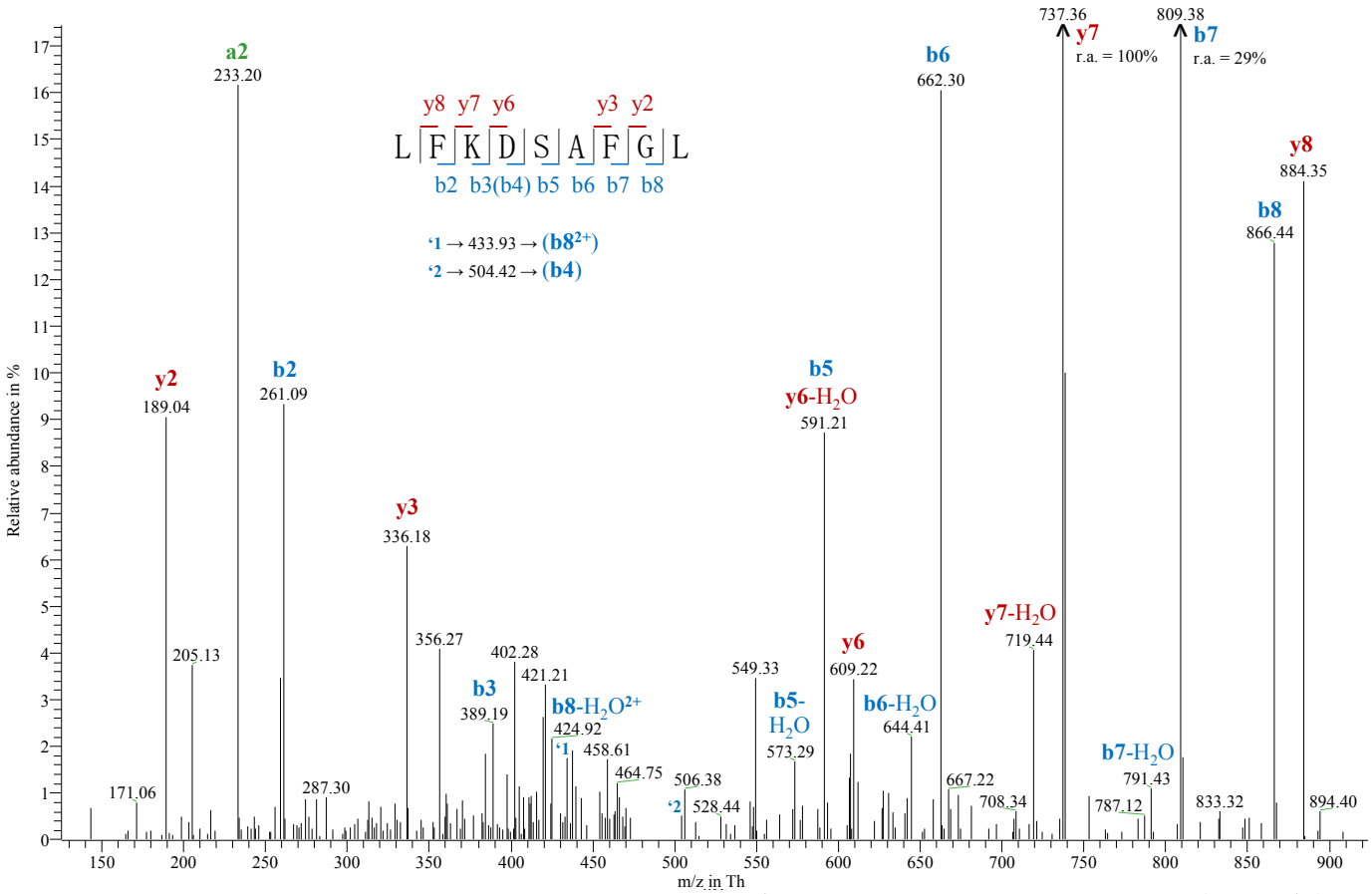
Supplementary Figure 9-9 | LTQ-MS^2 spectrum of SGNFIDQTRVLNLGPITR, an amino- and carboxy-terminally extended form of the H2-D^b motif peptide TRVLNLGPI. See also Supplementary Note (Nat Comm, Theo).



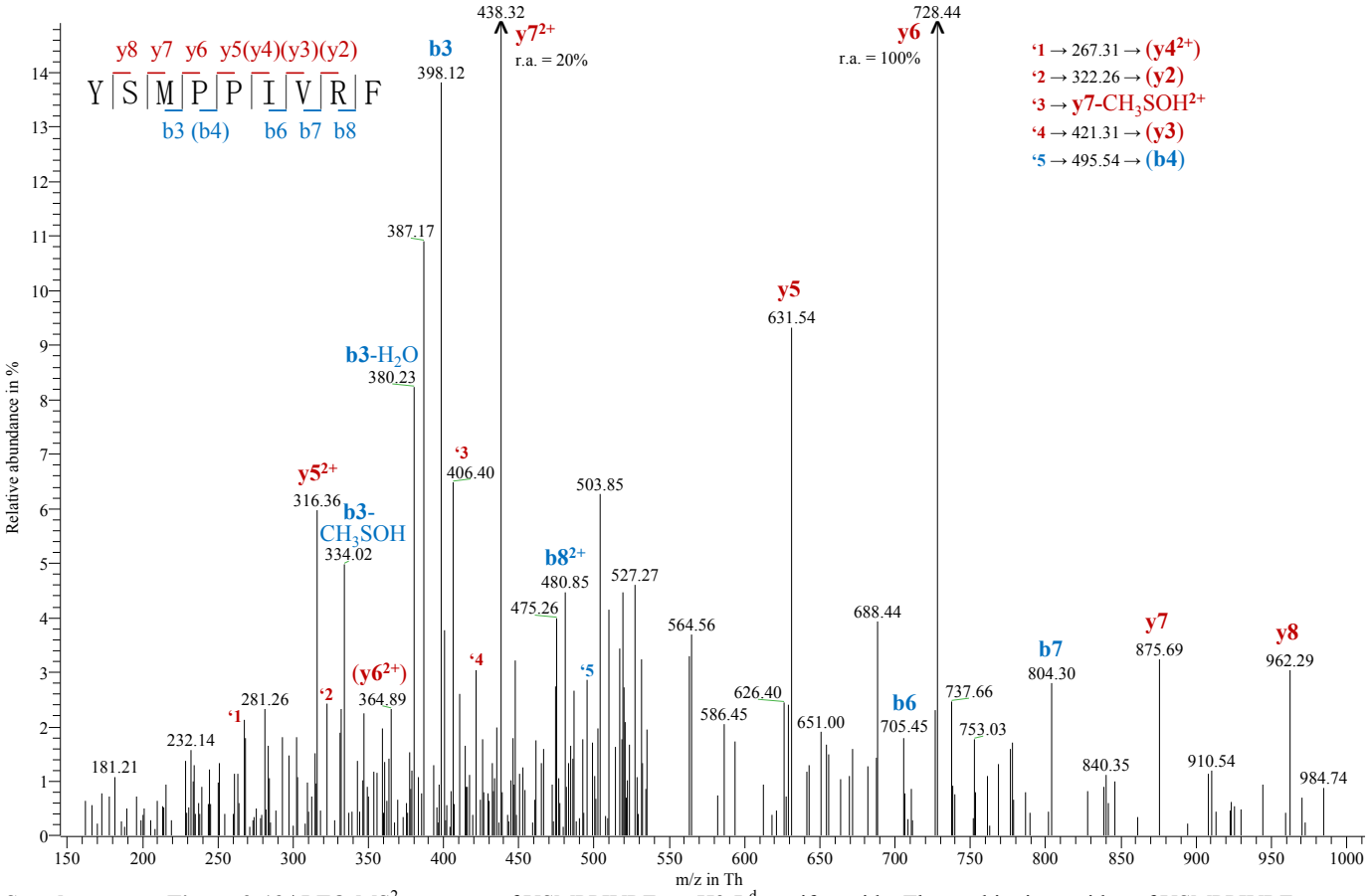
Supplementary Figure 9-10| LTQ-MS^2 spectrum of VYRPDQVSIL, an H2-K^d motif peptide. We obtained final evidence for this peptide by comparing MS^2 spectra as well as nano-HPLC elution times with those of the synthetic $^{13}C_5{}^{15}N_1$-valine labelled synthetic counterpart. See also Supplementary Note (Nat Comm, Theo).



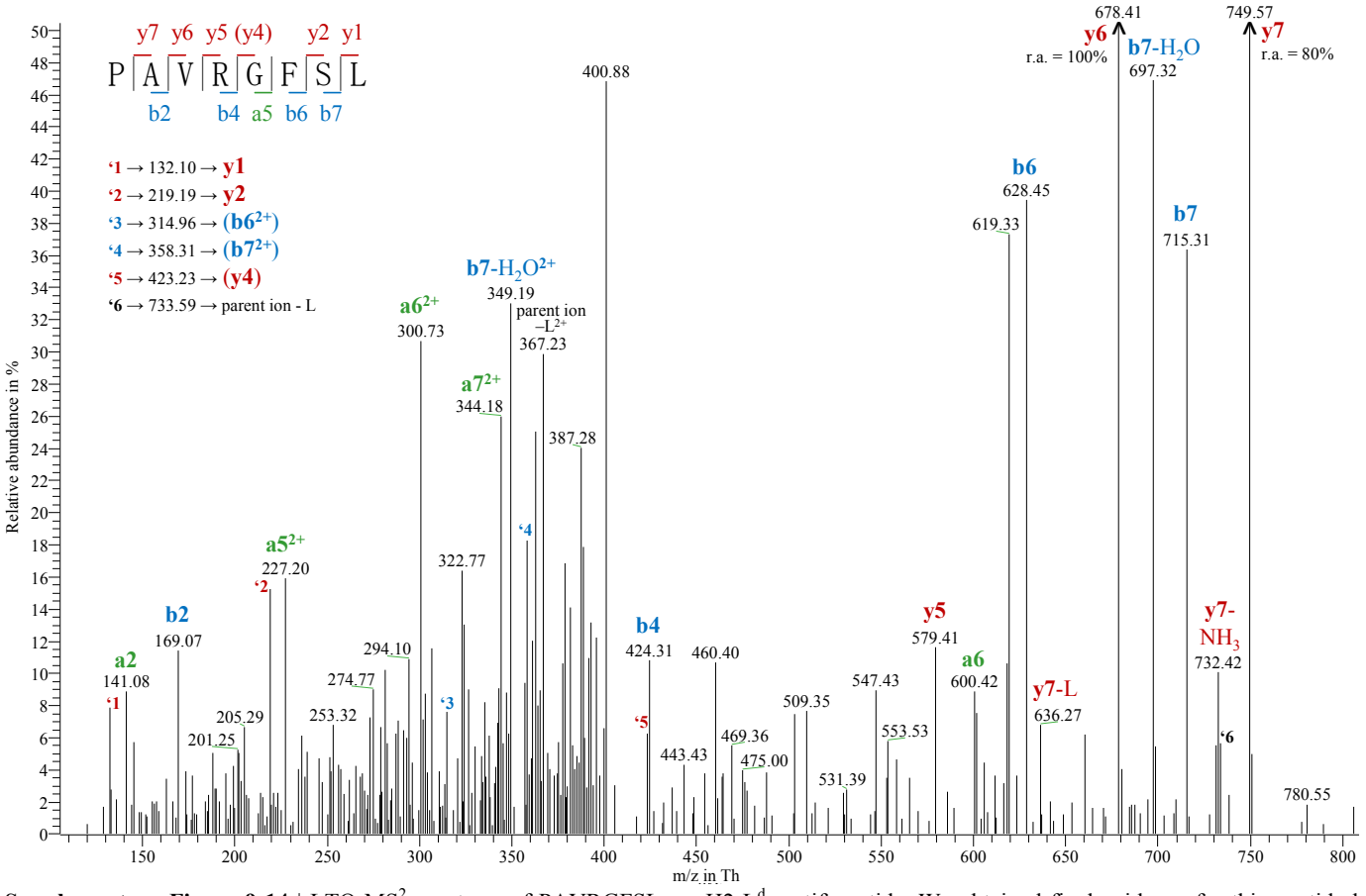
Supplementary Figure 9-11 | LTQ-MS^2 spectrum of LYWVDVERQV, an H2-K^d motif peptide. See also Supplementary Note (Nat Comm, Theo).



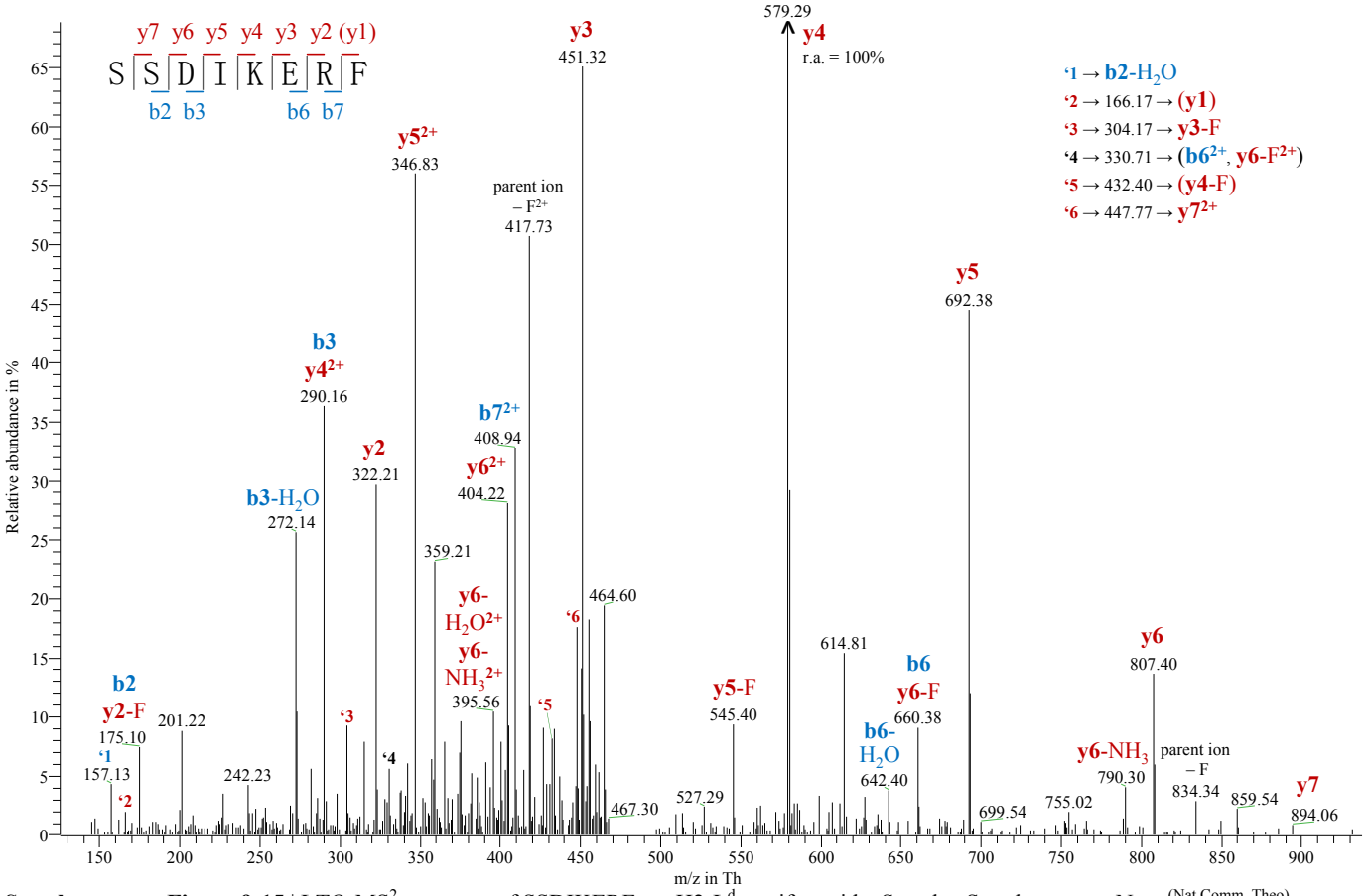
Supplementary Figure 9-12 | LTQ-MS^2 spectrum of LFKDSAFGL, an H2-K^d motif peptide. See also Supplementary Note (Nat Comm, Theo).



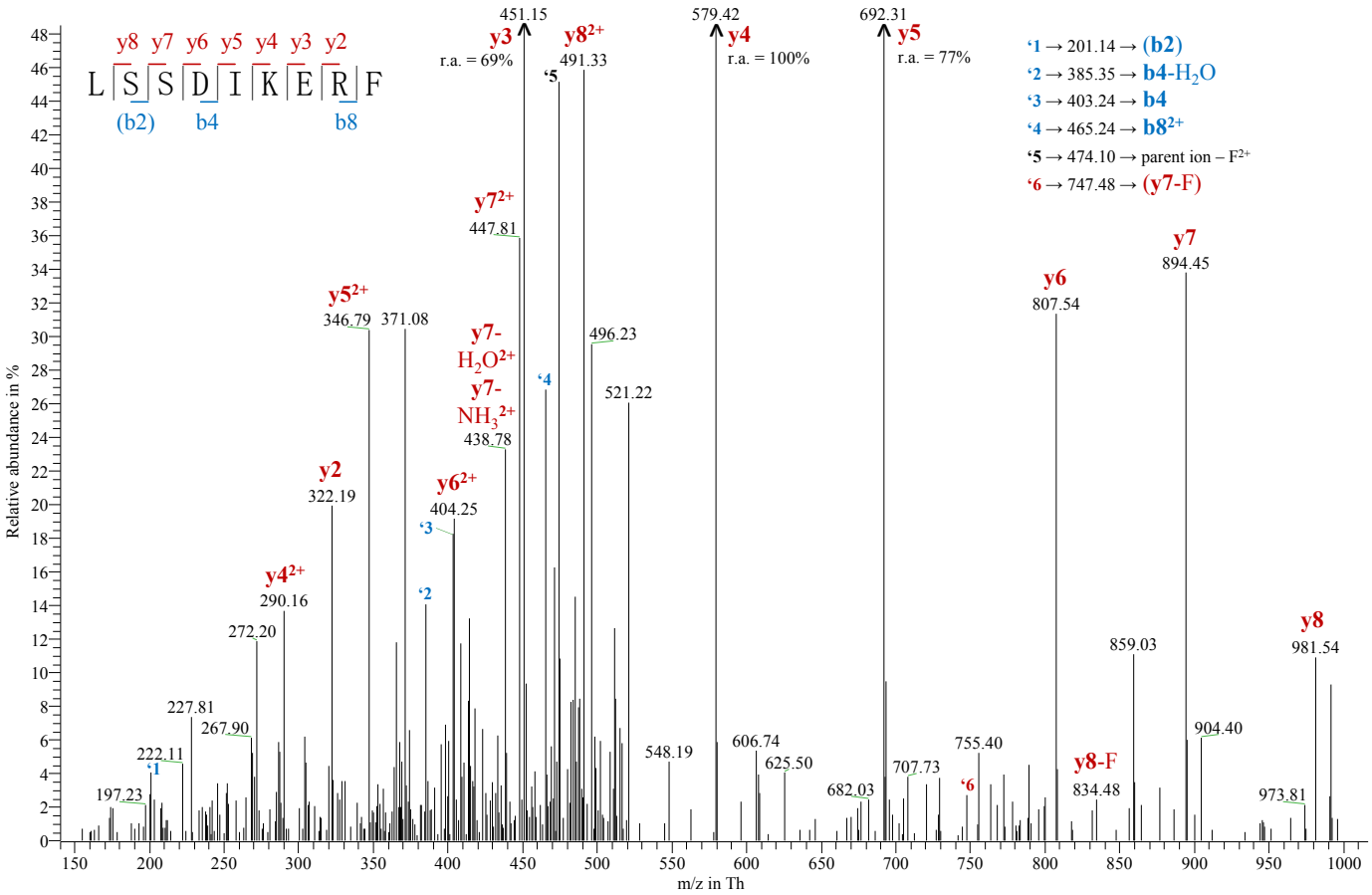
Supplementary Figure 9-13 | LTQ-MS^2 spectrum of YSMPPIVRF, an H2-L^d motif peptide. The methionine residue of YSMPPIVRF occurred in the oxidised form, i.e. as methionine sulfoxide, in the parent ions of this MS^2 spectrum. See also Supplementary Note (Nat Comm, Theo).



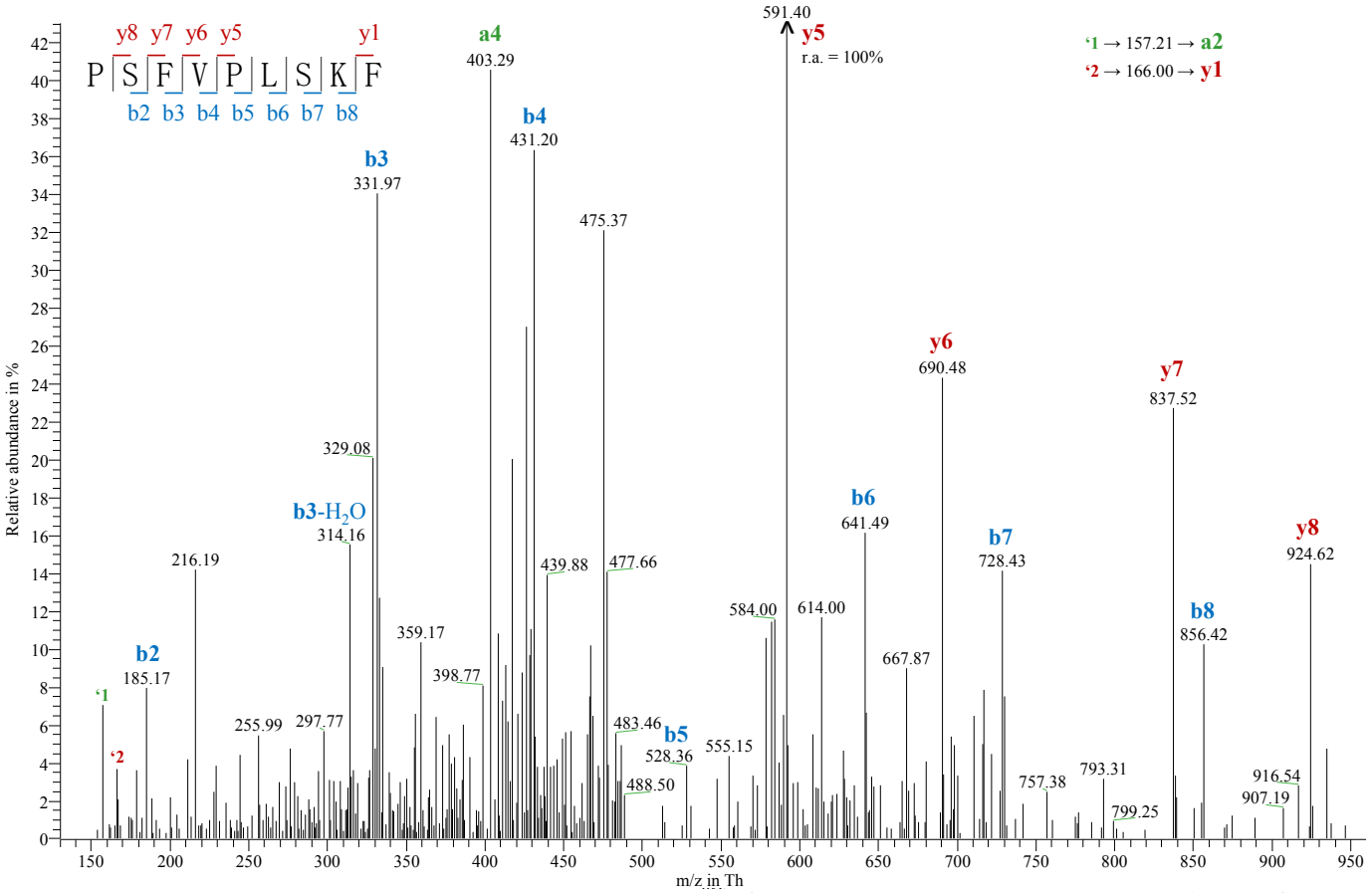
Supplementary Figure 9-14 | LTQ-MS^2 spectrum of PAVRGFSL, an H2-L^d motif peptide. We obtained final evidence for this peptide by comparing MS^2 spectra as well as nano-HPLC elution times with those of the synthetic $^{13}C_5{}^{15}N_1$-valine labelled synthetic counterpart. See also Supplementary Note (Nat Comm, Theo).



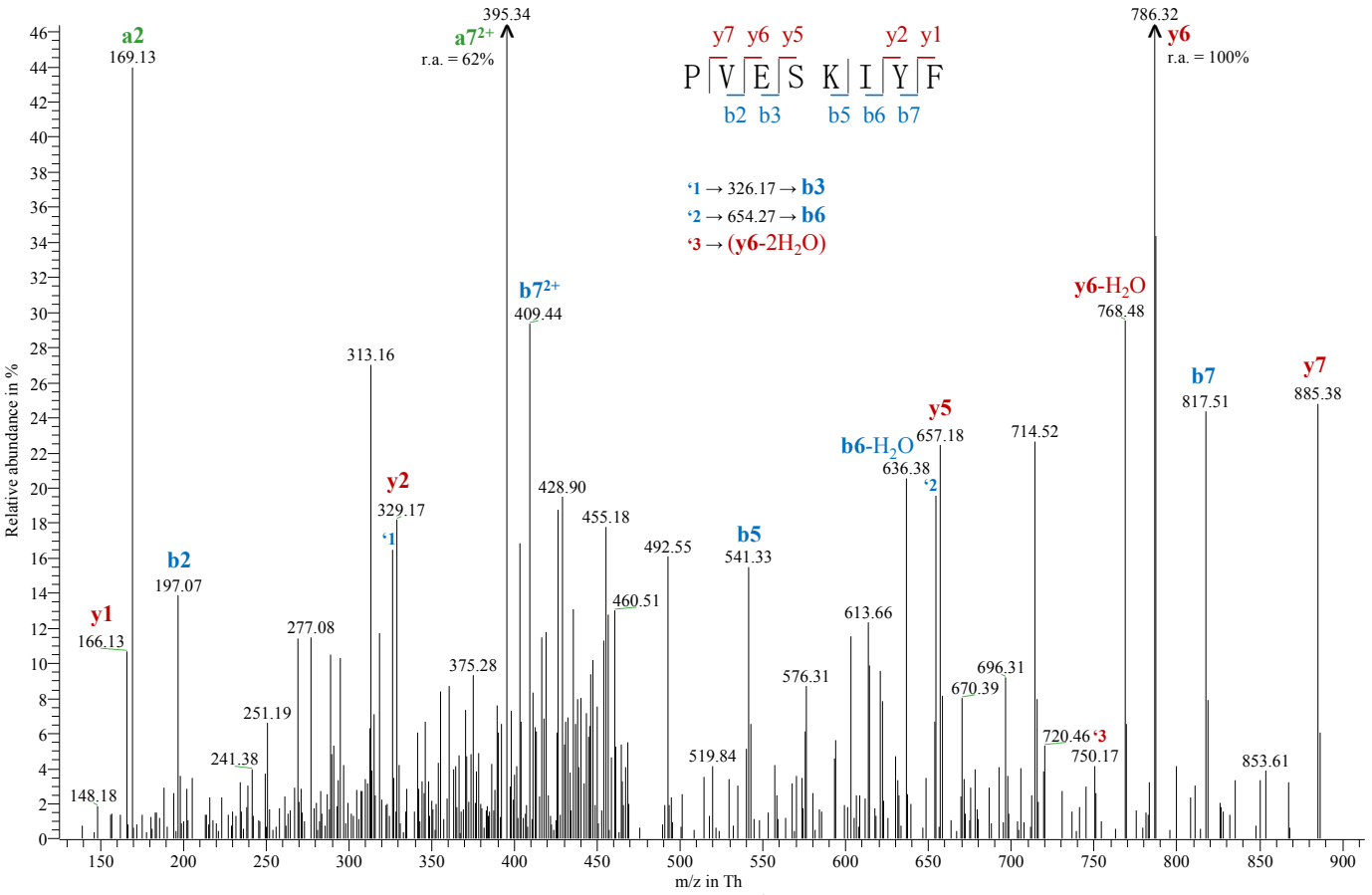
Supplementary Figure 9-15 | LTQ-MS^2 spectrum of SSDIKERF, an H2-L^d motif peptide. See also Supplementary Note (Nat Comm, Theo).



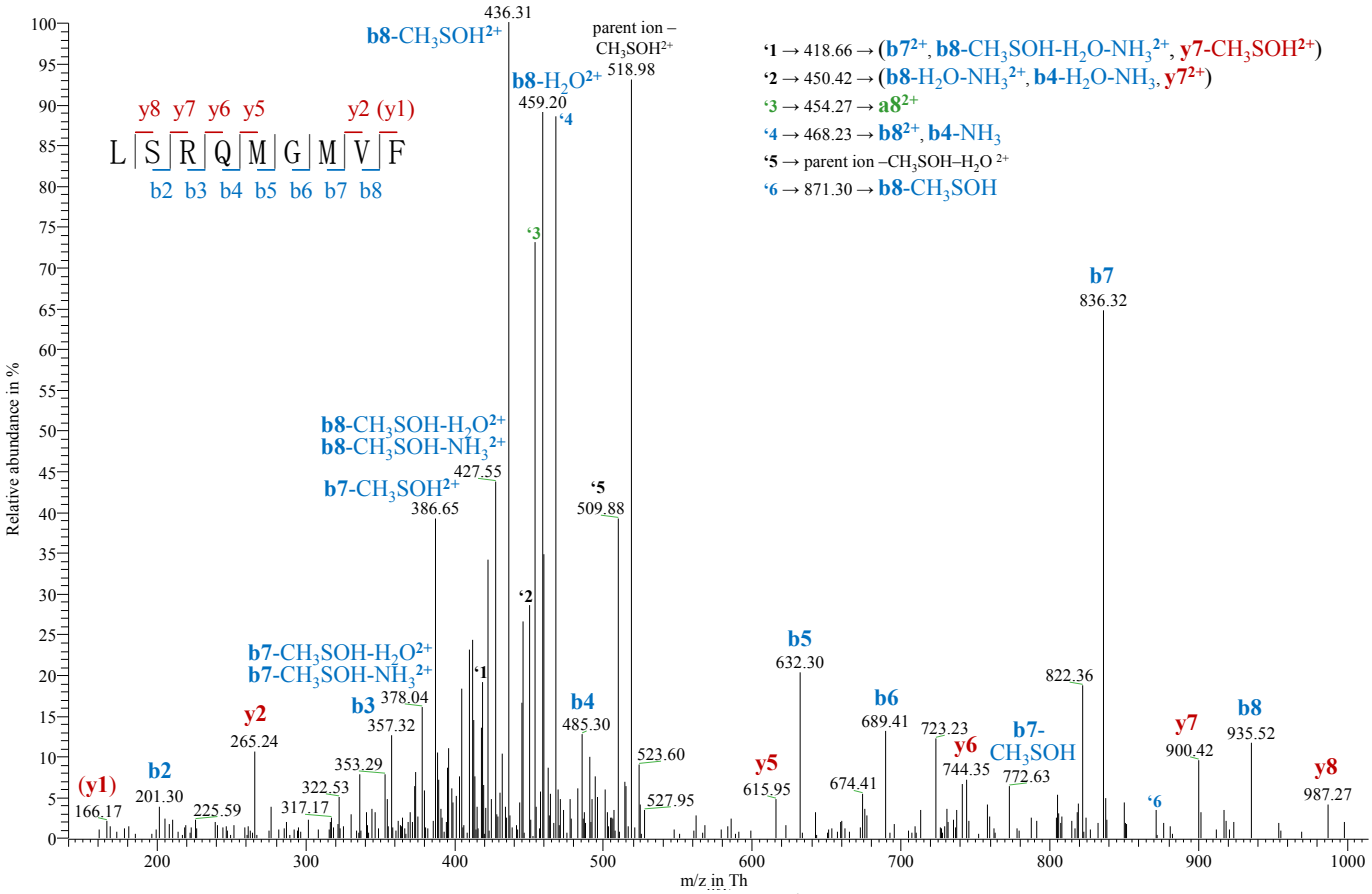
Supplementary Figure 9-16 | LTQ-MS^2 spectrum of LSSDIKERF, an H2-L^d motif peptide. See also Supplementary Note (Nat Comm, Theo).



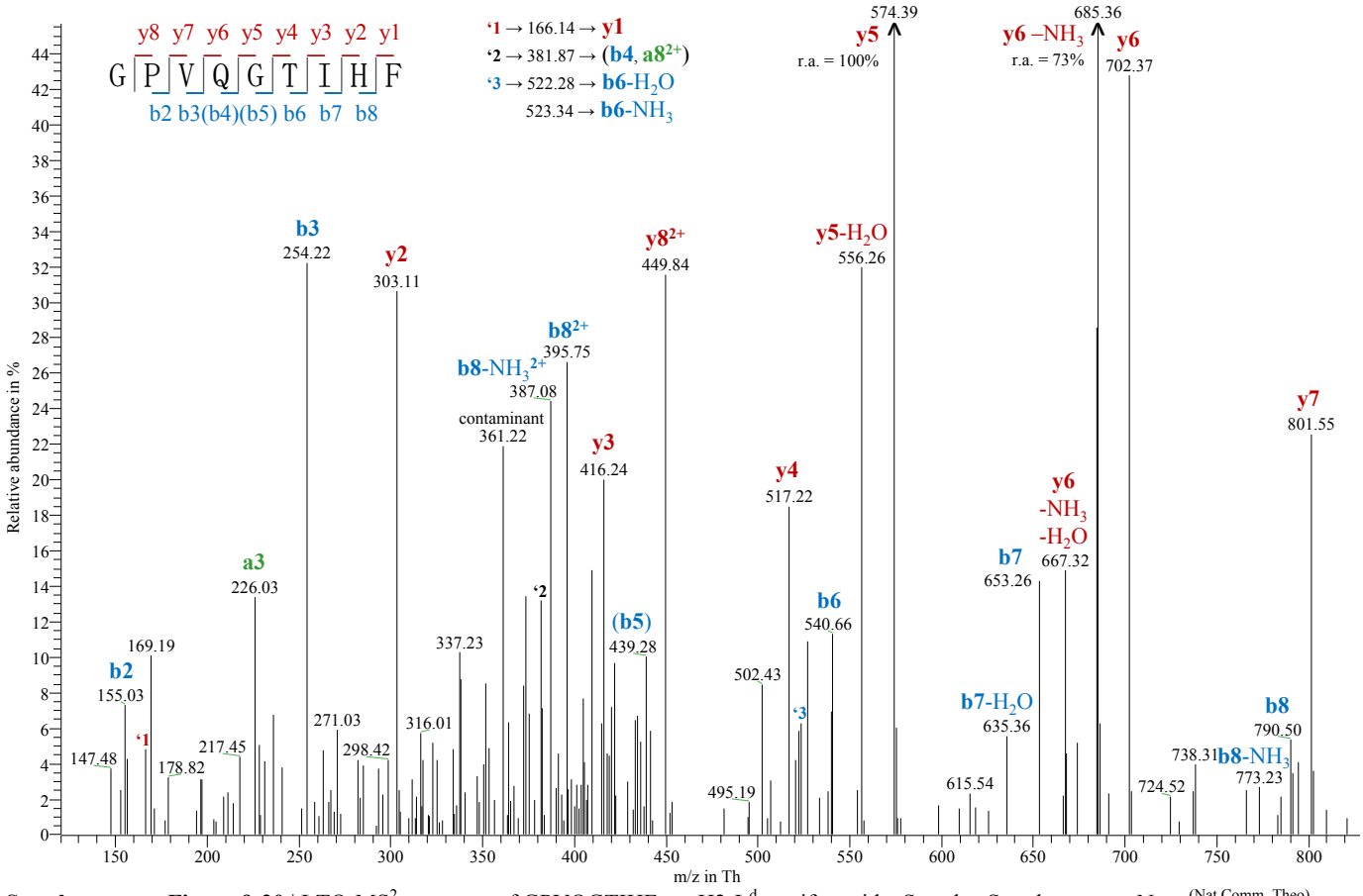
Supplementary Figure 9-17 | LTQ-MS^2 spectrum of PSFVPLSKF, an H2-L^d motif peptide. See also Supplementary Note (Nat Comm, Theo).



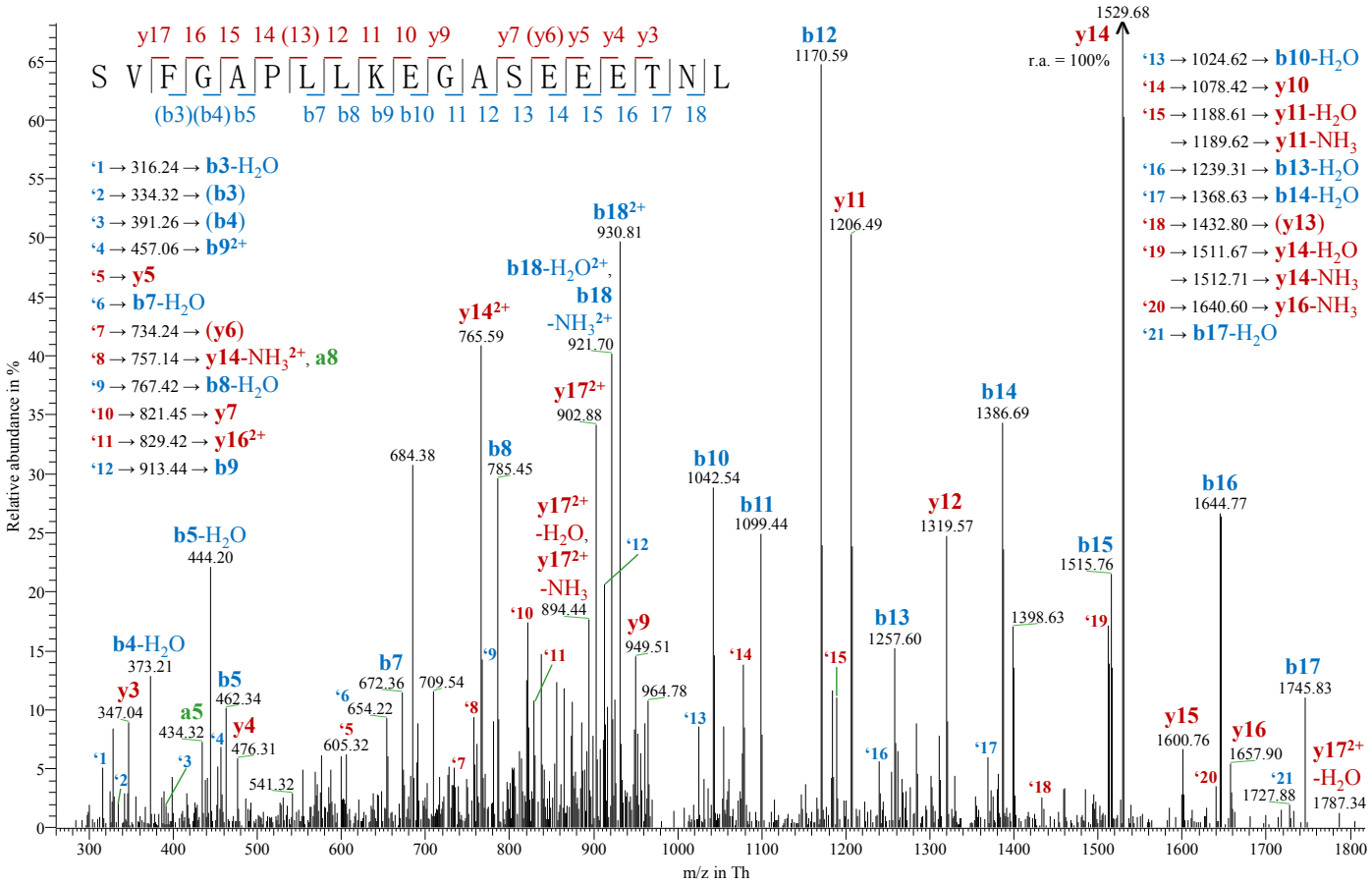
Supplementary Figure 9-18 | LTQ-MS^2 spectrum of PVESKIYF, an H2-L^d motif peptide. We obtained final evidence for this peptide by comparing MS^2 spectra as well as nano-HPLC elution times with those of the synthetic $^{13}C_5{}^{15}N_1$-valine labelled synthetic counterpart. See also Supplementary Note (Nat Comm, Theo).



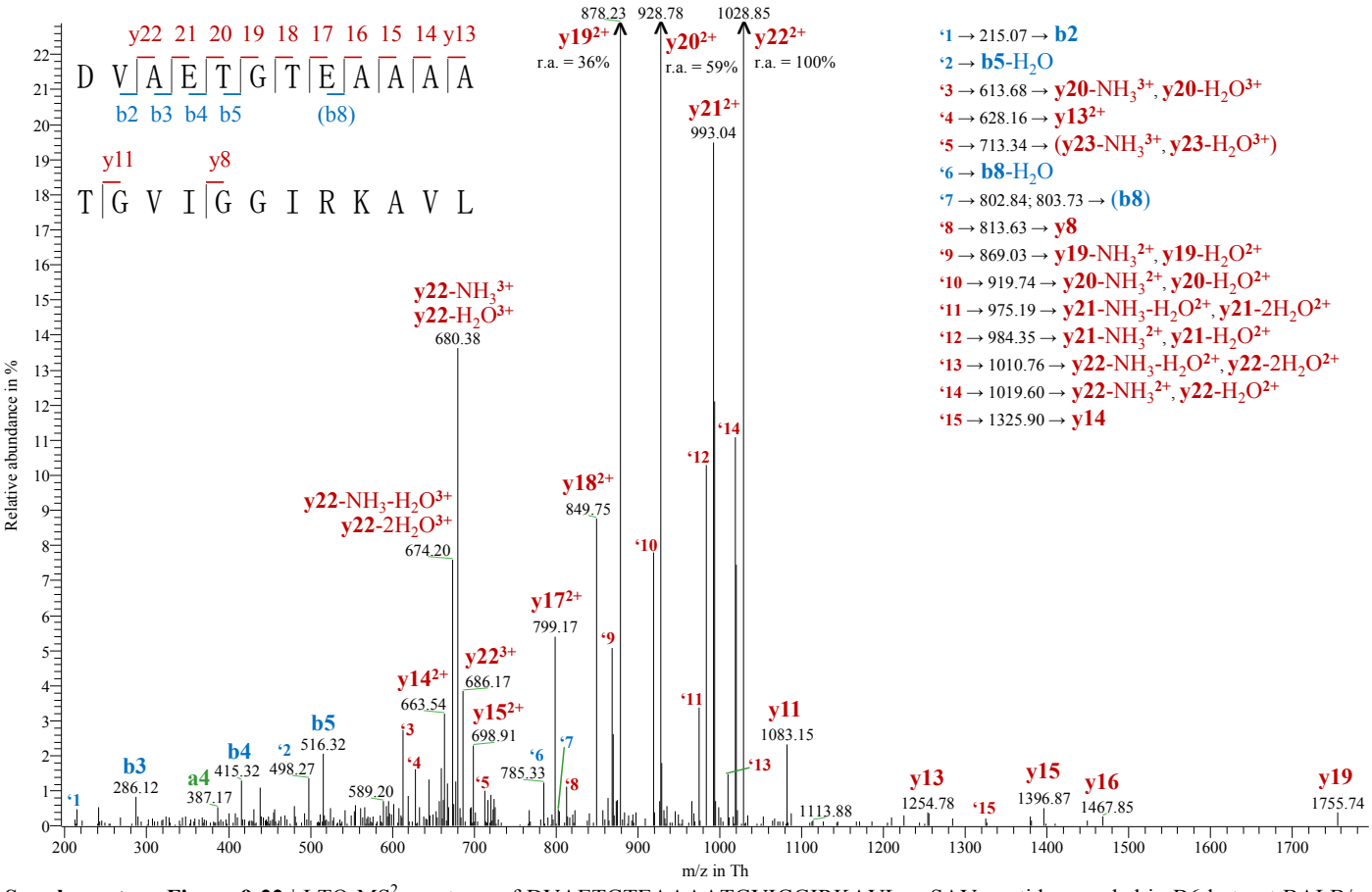
Supplementary Figure 9-19 | LTQ-MS^2 spectrum of LSRQMGMVF, an H2-L^d motif peptide. Both methionine residues of LSRQMGMVF occurred in the oxidised form, i.e. as methionine sulfoxide, in the parent ions of this MS^2 spectrum. See also Supplementary Note (Nat Comm, Theo).



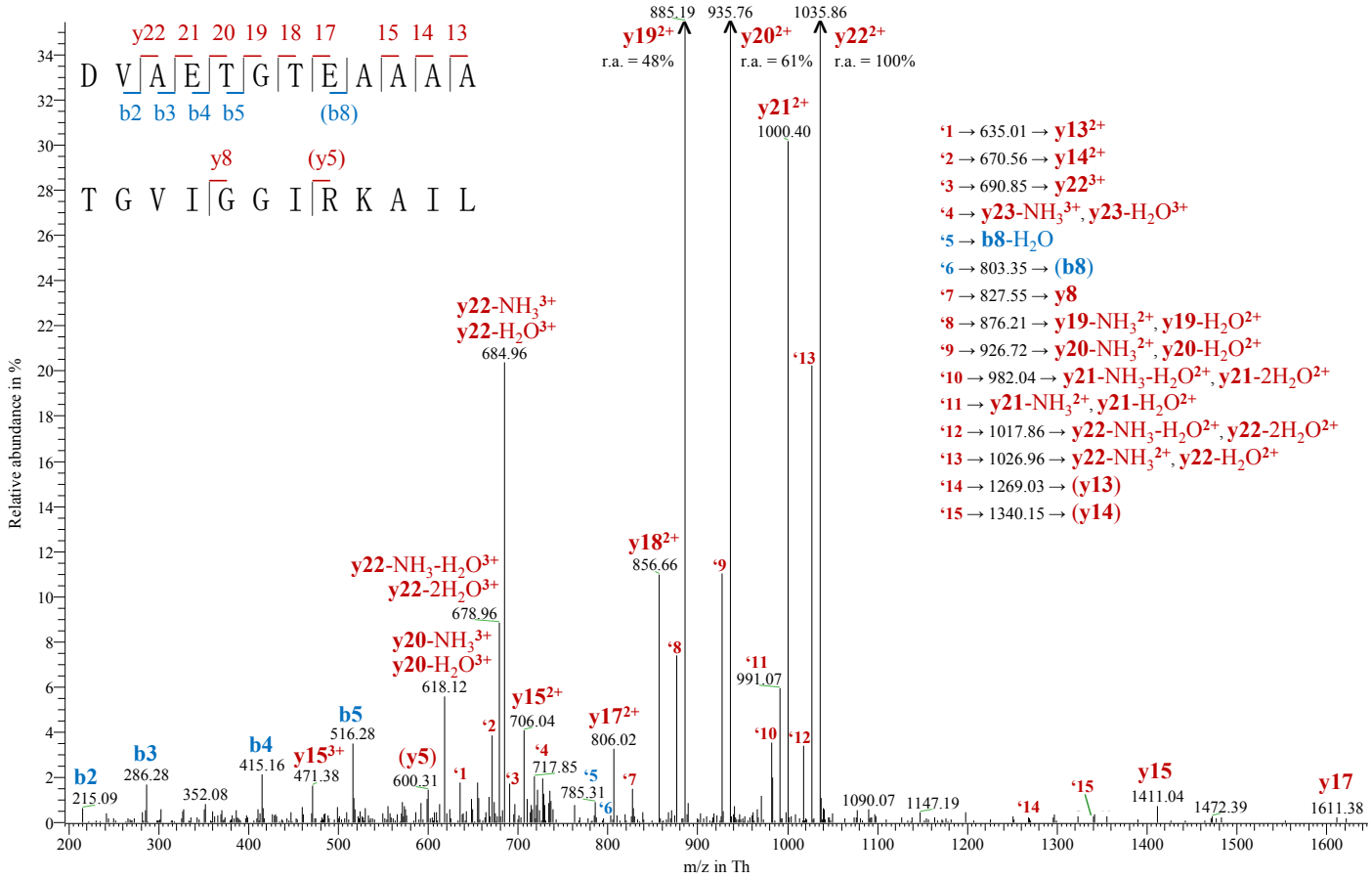
Supplementary Figure 9-20 | LTQ-MS^2 spectrum of GPVQGTIHF, an H2-L^d motif peptide. See also Supplementary Note (Nat Comm, Theo).



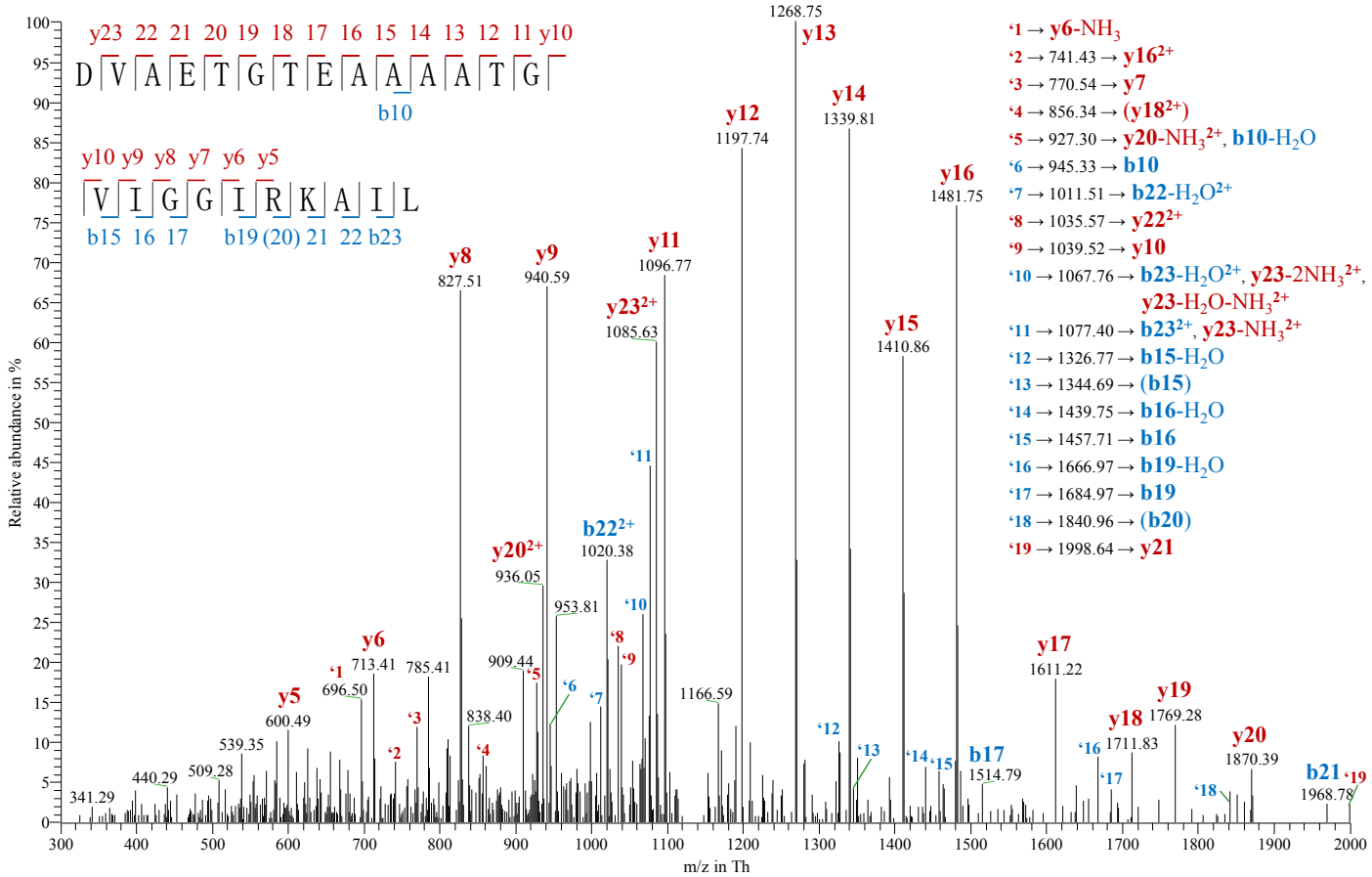
Supplementary Figure 9-21 | LTQ-MS^2 spectrum of SVFGAPLLKEGASEEETNL, a SAV peptide encoded in BALB/c but not B6 mice. See also Supplementary Note (Nat Comm, Theo).



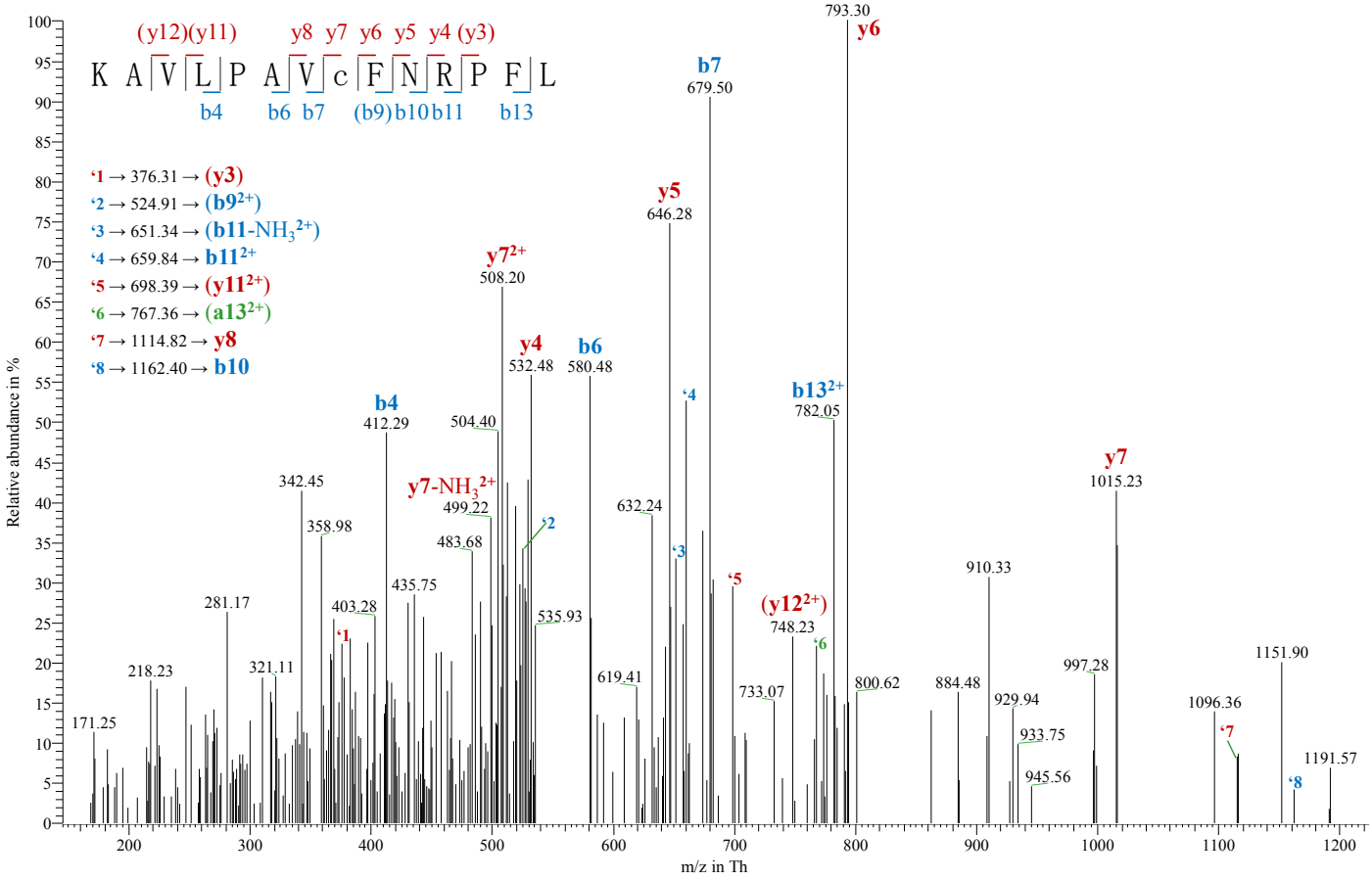
Supplementary Figure 9-22 | LTQ-MS^2 spectrum of DVAETGTEAAAATGVIGGIRKAVL, a SAV peptide encoded in B6 but not BALB/c mice. The parent ions of this MS^2 spectrum were triply charged. See also Supplementary Note (Nat Comm, Theo).



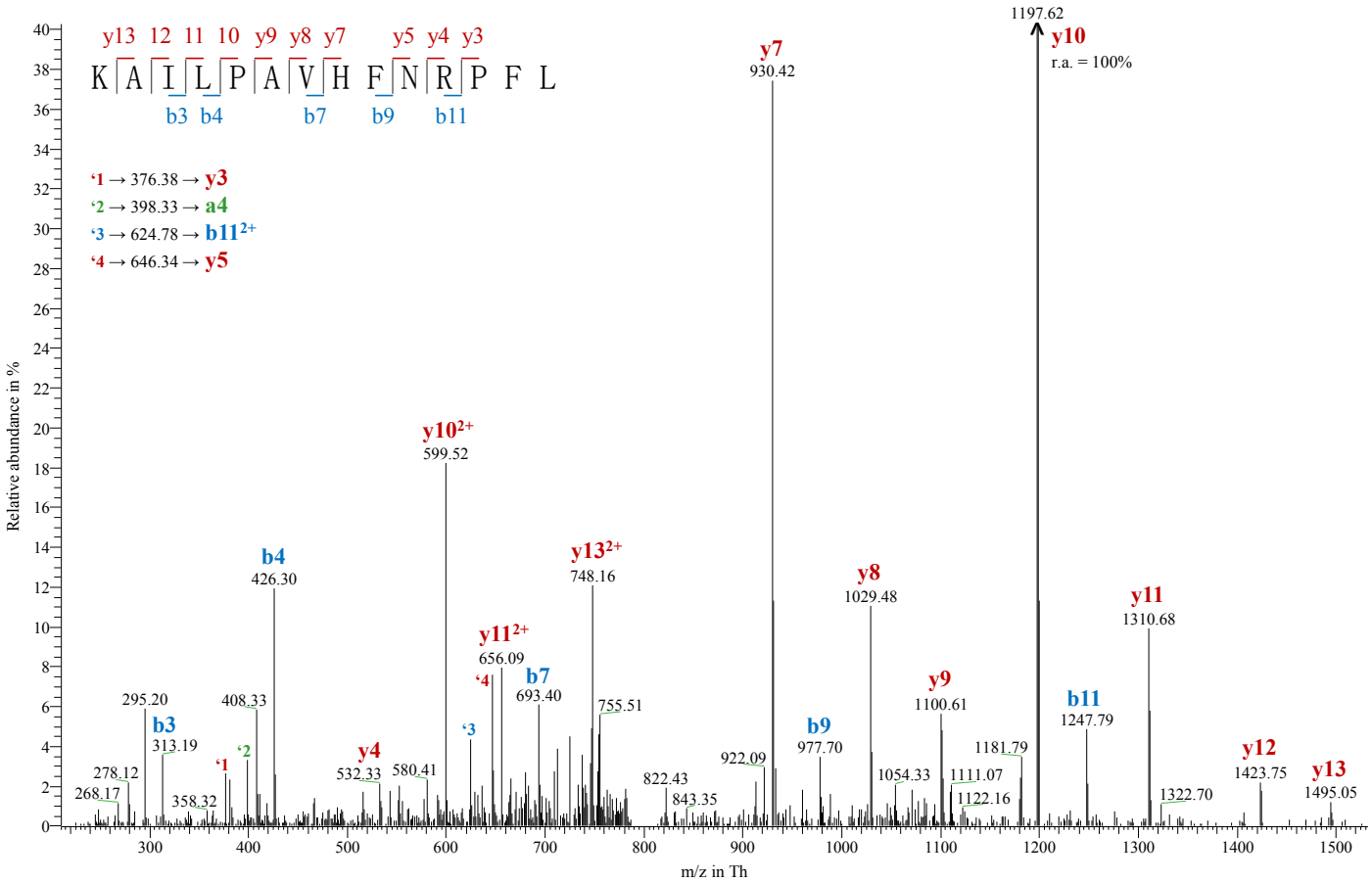
Supplementary Figure 9-23 | LTQ-MS^2 spectrum of DVAETGTEAAAATGVIGGIRKAIL, a SAV peptide encoded in BALB/c but not B6 mice. The parent ions of this MS^2 spectrum were triply charged. Although the MS^2 spectrum of doubly charged DVAETGTEAAAATGVIGGIRKAIL (see Supplementary Figure 9-24) resulted in a higher Andromeda score than the MS^2 spectrum of the triply charged precursor, this spectrum is shown in addition to demonstrate the remarkable similarity of MS^2 fragmentation patterns of triply charged DVAETGTEAAAATGVIGGIRKAVL and DVAETGTEAAAATGVIGGIRKAIL. This congruence further augments the confidence of identified peptide sequences. See also Supplementary Note (Nat Comm, Theo).



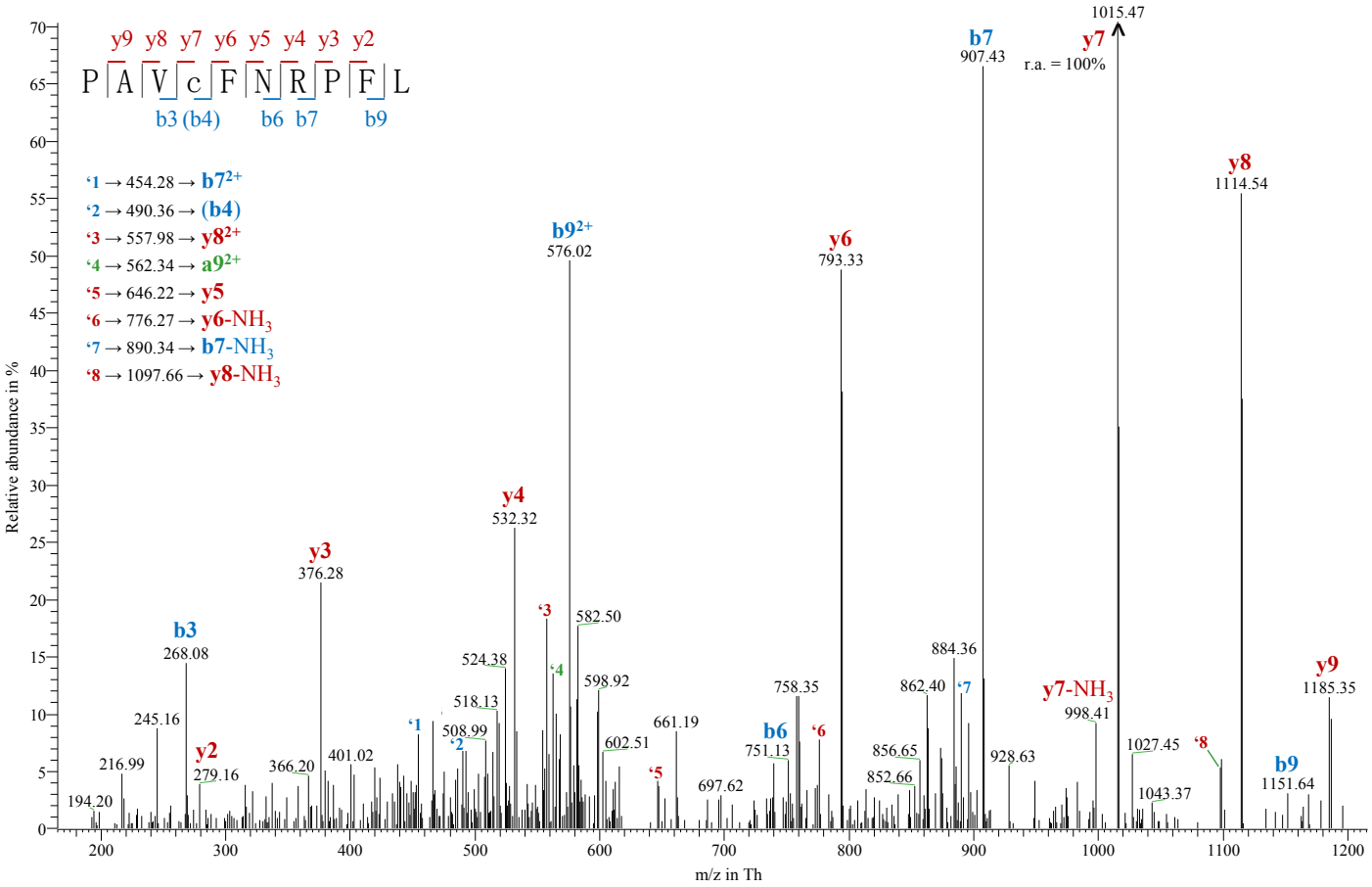
Supplementary Figure 9-24 | LTQ-MS^2 spectrum of DVAETGTEAAAATGVIGGIRKAIL, a SAV peptide encoded in BALB/c but not B6 mice. The parent ions of this MS^2 spectrum were doubly charged. Although MS^2 spectra of doubly and triply (see Supplementary Figure 9-23) charged DVAETGTEAAAATGVIGGIRKAIL look quite different, they both descend from this peptide which is affirmed by the identical elution profiles of the two precursor forms in nano-HPLC. See also Supplementary Note (Nat Comm, Theo).



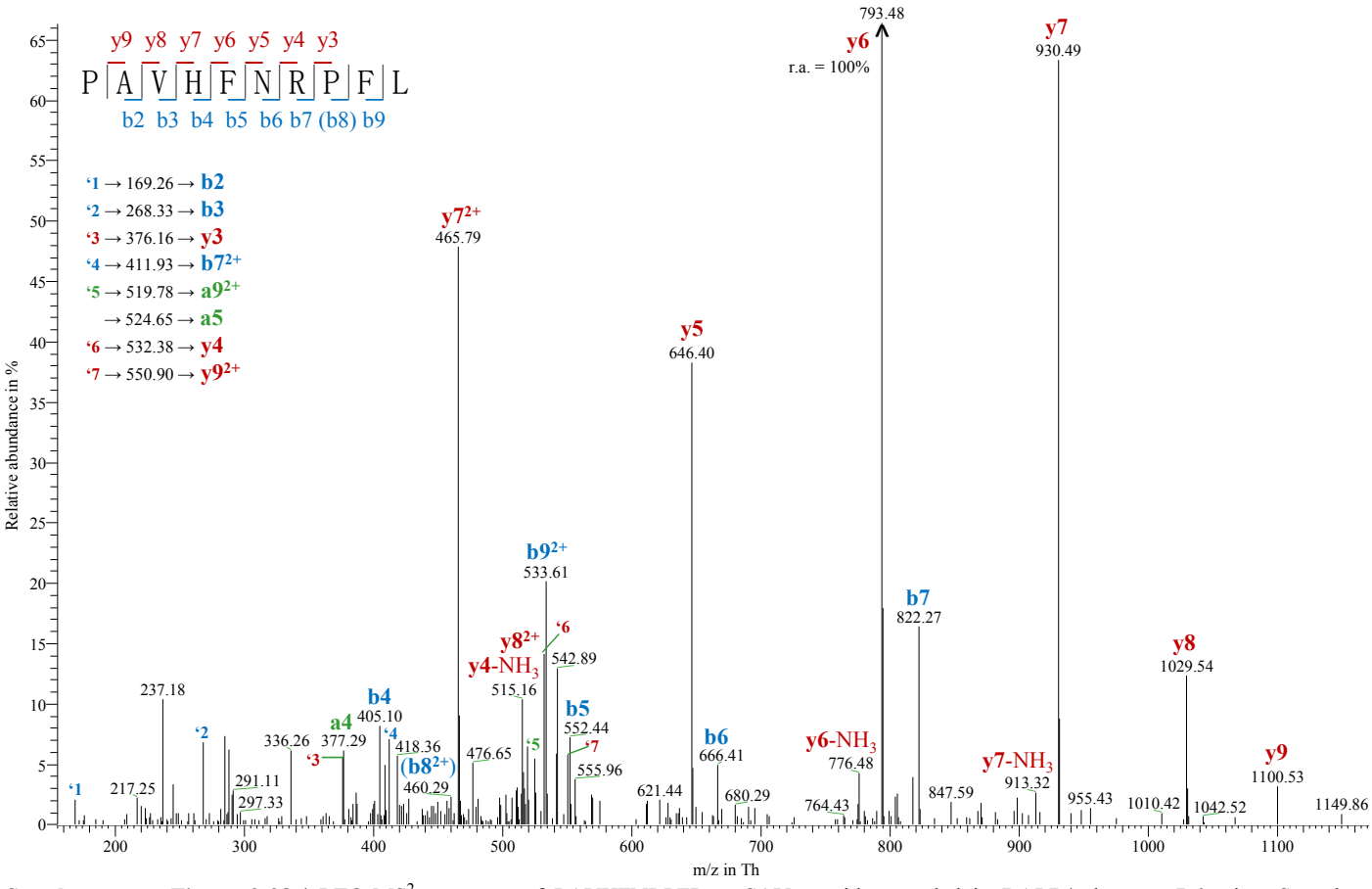
Supplementary Figure 9-25 | LTQ-MS^2 spectrum of KAVLPAVC^{Cys}FNRPFL, a SAV peptide encoded in B6 but not BALB/c mice. The cysteine residue occurred in a cysteinylated form, i.e. it had bound free cysteine via a disulfide bridge. See also Supplementary Note (Nat Comm, Theo).



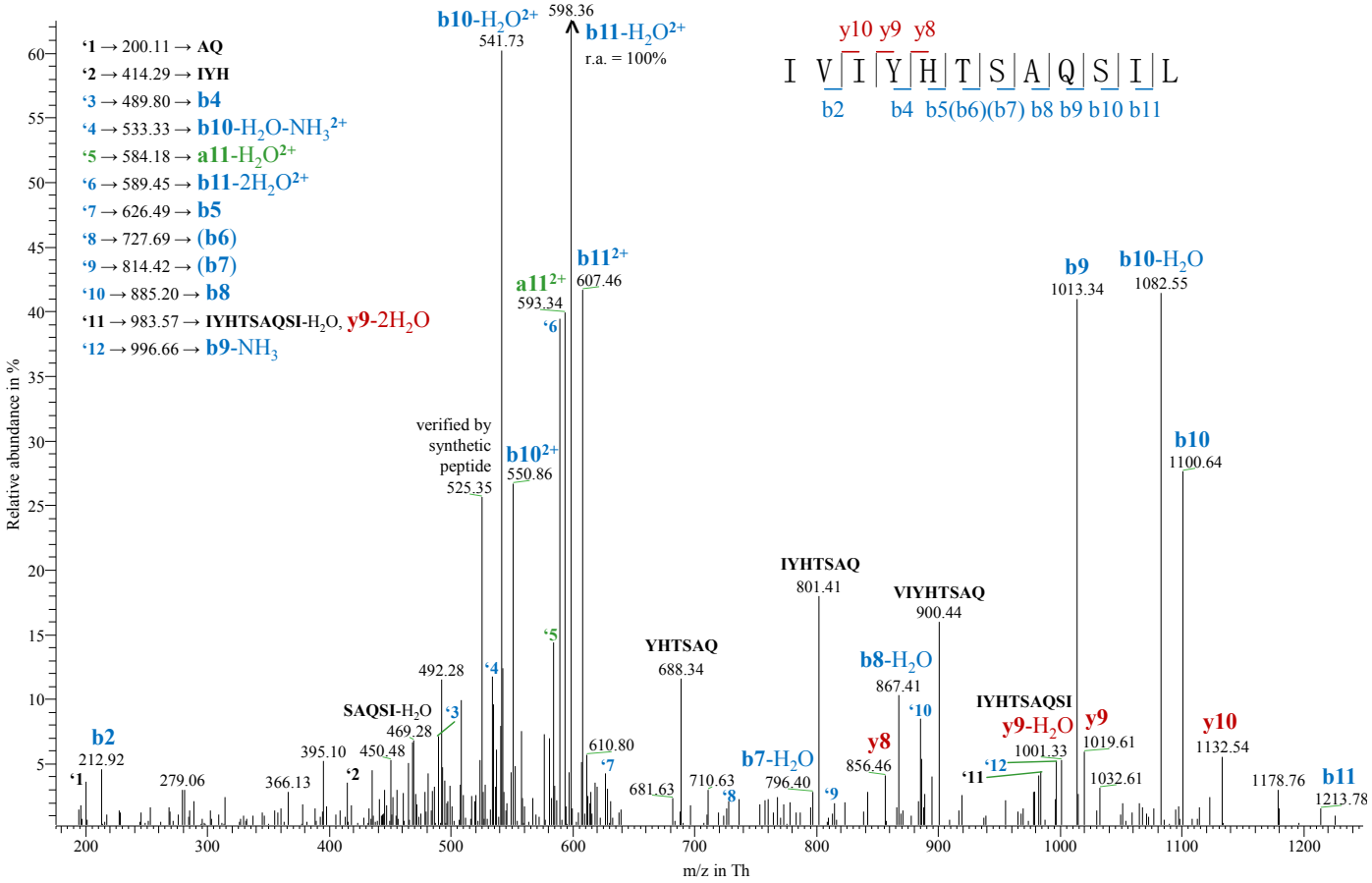
Supplementary Figure 9-26 | LTQ-MS^2 spectrum of KAILPAVHFNRPFL, a SAV peptide encoded in BALB/c but not B6 mice. Note that this MS^2-spectrum of KAILPAVHFNRPFL derives from a doubly charged precursor whereas the depicted MS^2 spectrum of KAVLPAVC^{Cys}FNRPFL (Supplementary Figure 9-25) descends from triply charged parent ions. The MS^2 spectra of triply charged KAILPAVHFNRPFL parent ions differed from those of the doubly charged form, too (compare MS^2 spectra of triply and doubly charged DVAETGTEAAAATGVIGGIRKAIL, Supplementary Figures 9-23 and 9-24 respectively). See also Supplementary Note (Nat Comm, Theo).



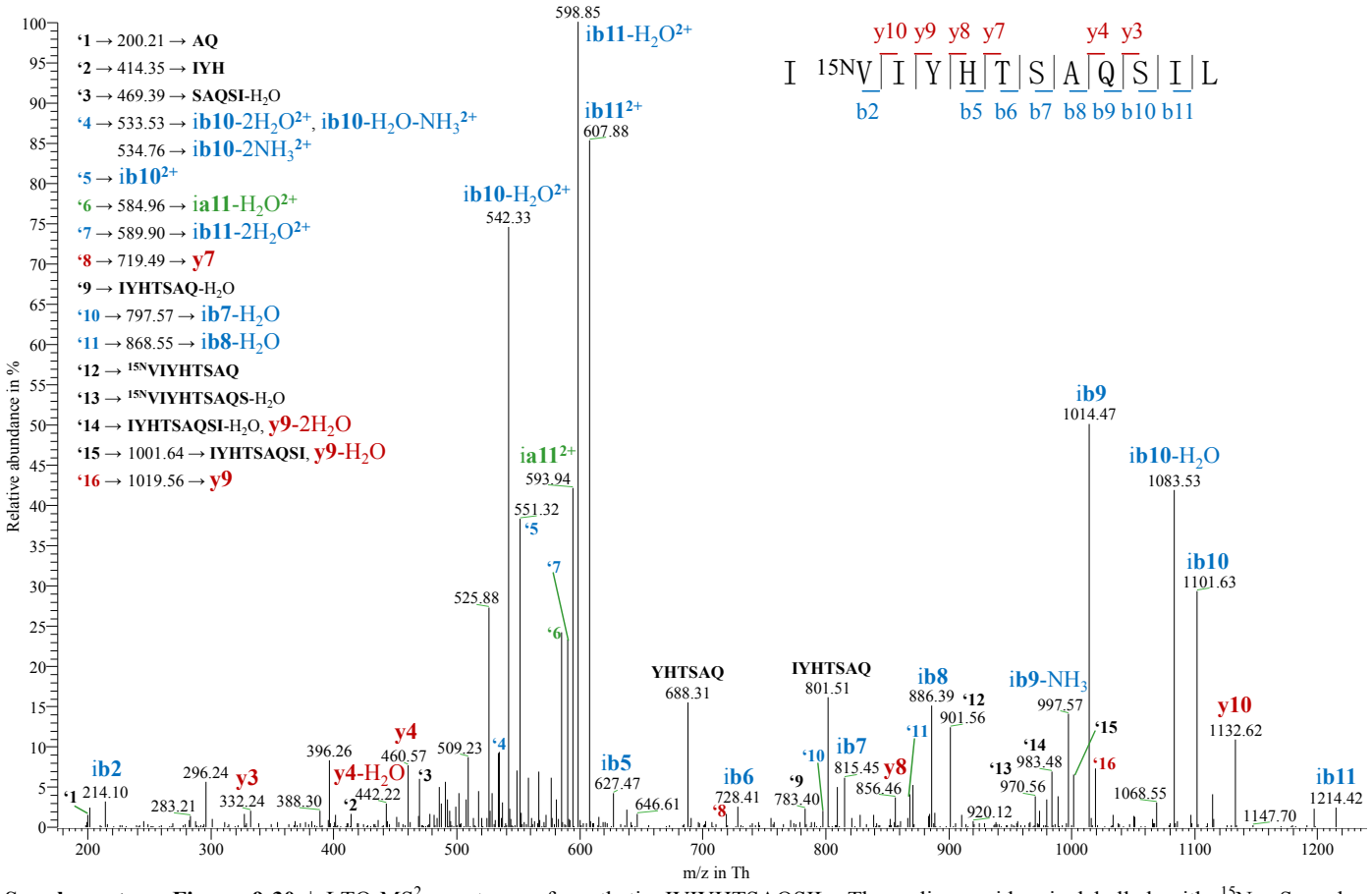
Supplementary Figure 9-27 | LTQ-MS^2 spectrum of PAVC^{Cys}FNRPFL, a SAV peptide encoded in B6 but not BALB/c mice. The cysteine residue occurred in a cysteinylated form, i.e. it had bound free cysteine via a disulfide bridge. See also Supplementary Note (Nat Comm, Theo).



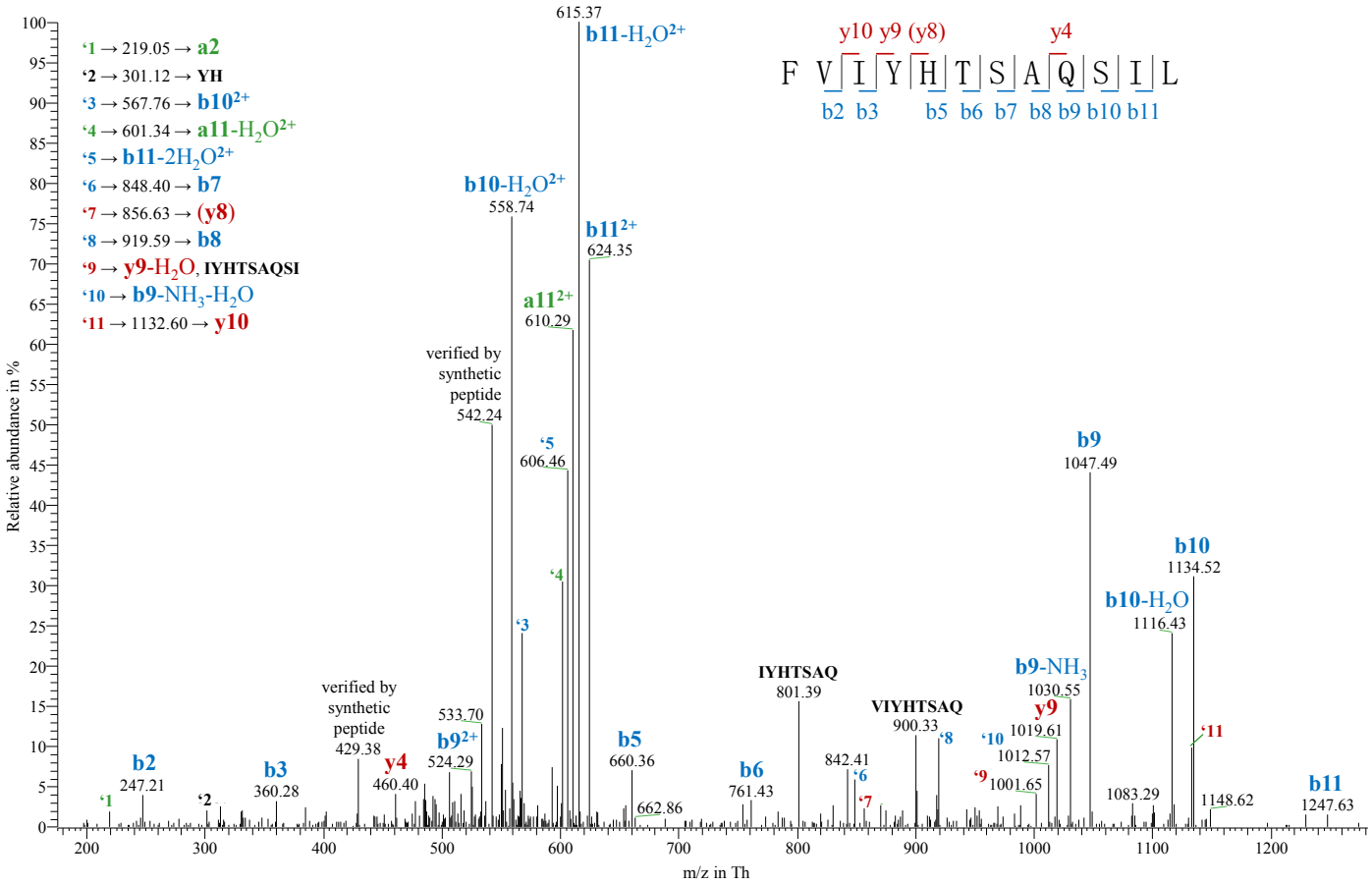
Supplementary Figure 9-28 | LTQ-MS^2 spectrum of PAVHFNRPFL, a SAV peptide encoded in BALB/c but not B6 mice. See also Supplementary Note (Nat Comm, Theo).



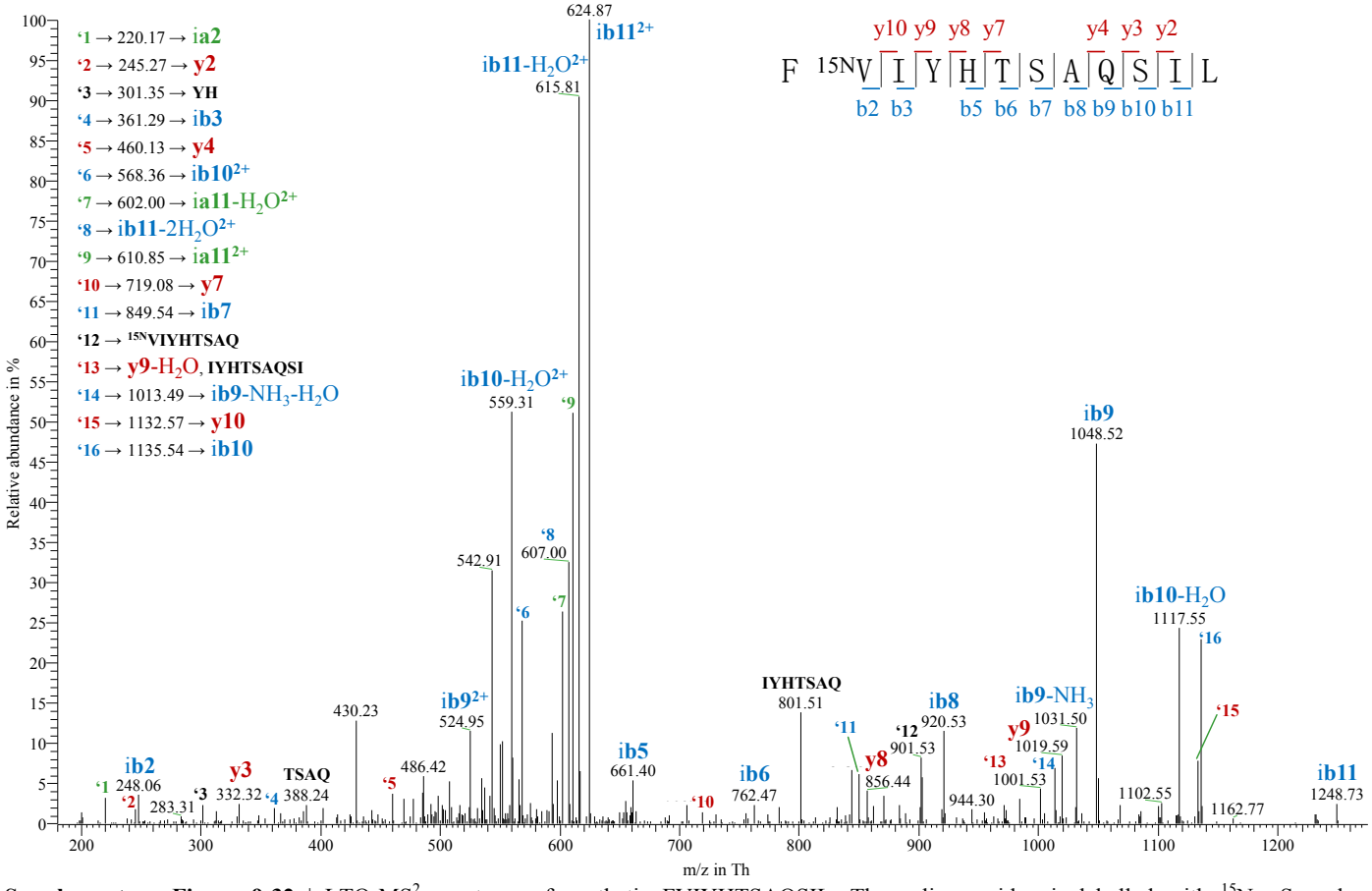
Supplementary Figure 9-29 | LTQ-MS^2 spectrum of IVIYHTSAQSIL, a SAV peptide encoded in B6 but not BALB/c mice. See also Supplementary Note (Nat Comm, Theo).



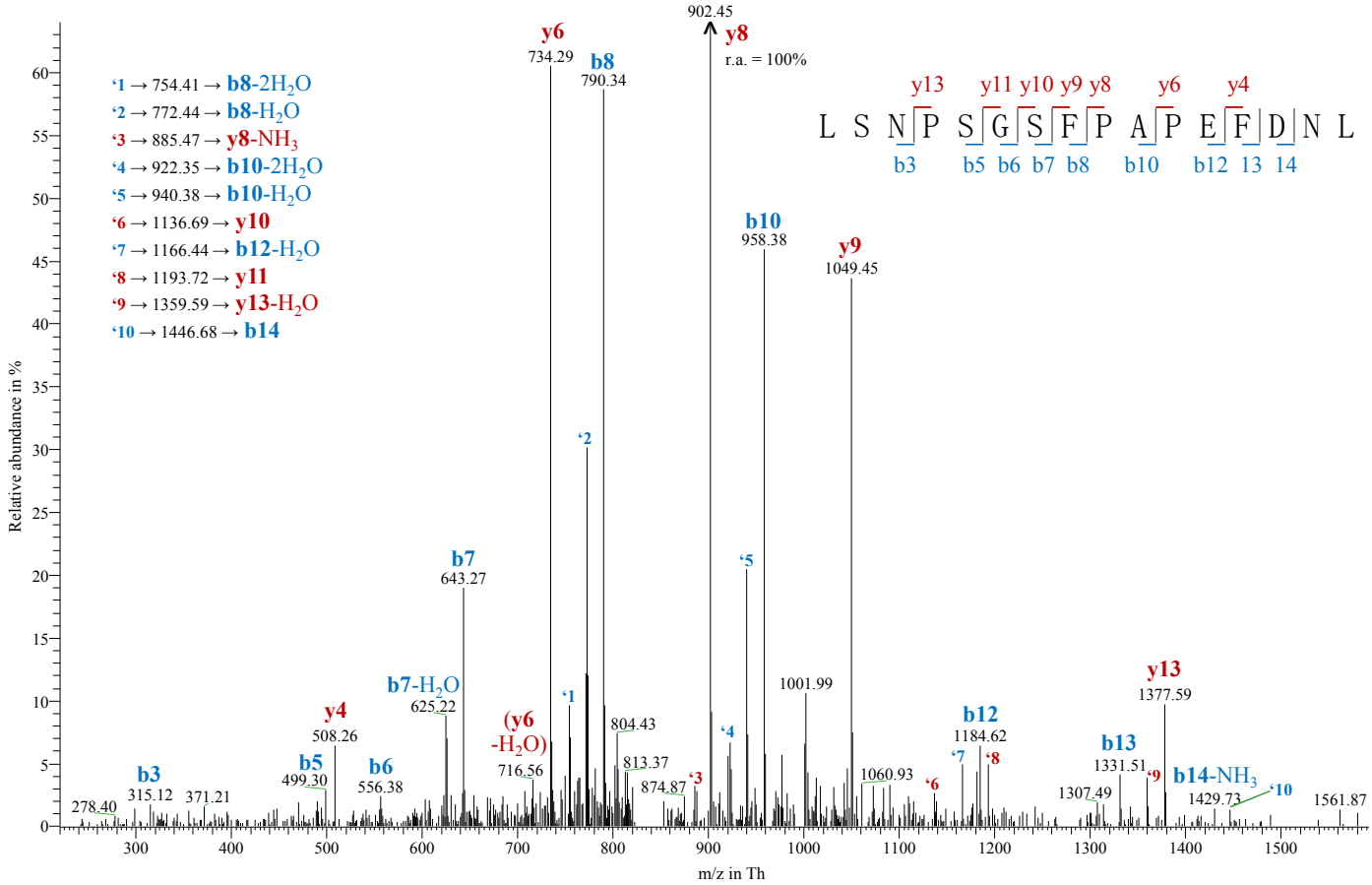
Supplementary Figure 9-30 | LTQ-MS^2 spectrum of synthetic IVIYHTSAQSIL. The valine residue is labelled with $^{15}N_1$. See also Supplementary Note (Nat Comm, Theo).



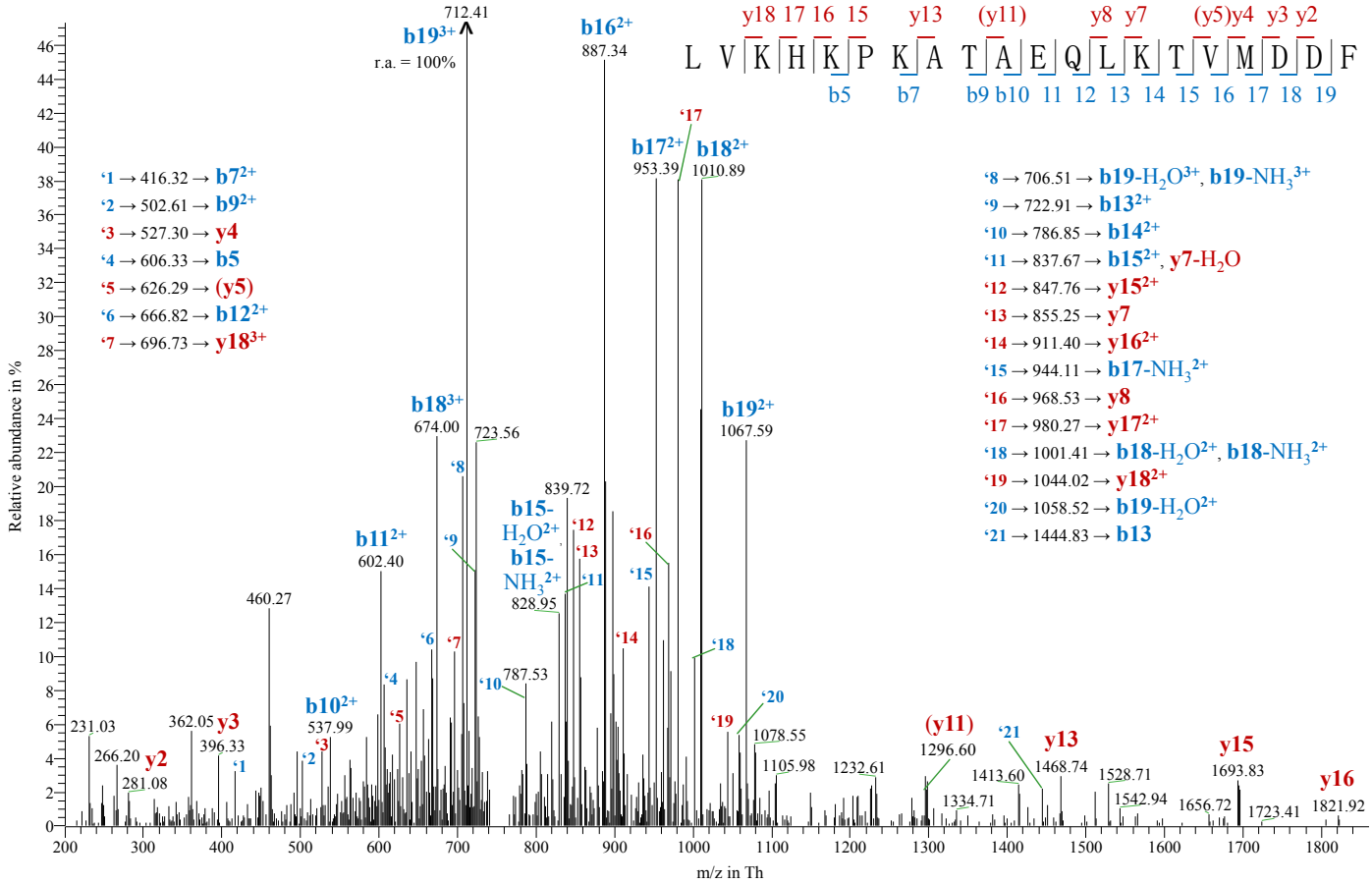
Supplementary Figure 9-31 | LTQ-MS^2 spectrum of FVIYHTSAQSIL, a SAV peptide encoded in BALB/c but not B6 mice. See also Supplementary Note (Nat Comm, Theo).



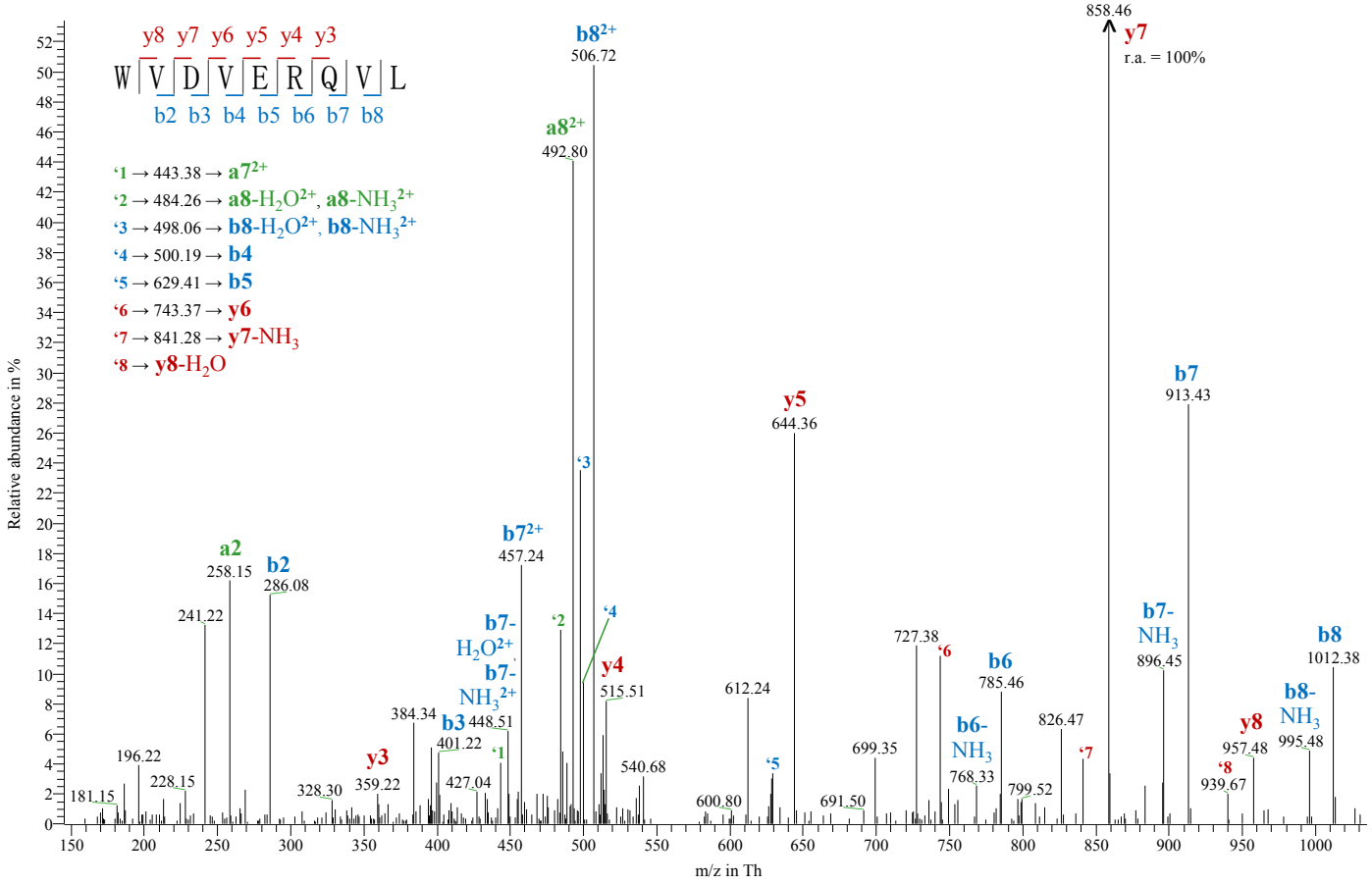
Supplementary Figure 9-32 | LTQ-MS^2 spectrum of synthetic FVIYHTSAQSIL. The valine residue is labelled with $^{15}N_1$. See also Supplementary Note (Nat Comm, Theo).



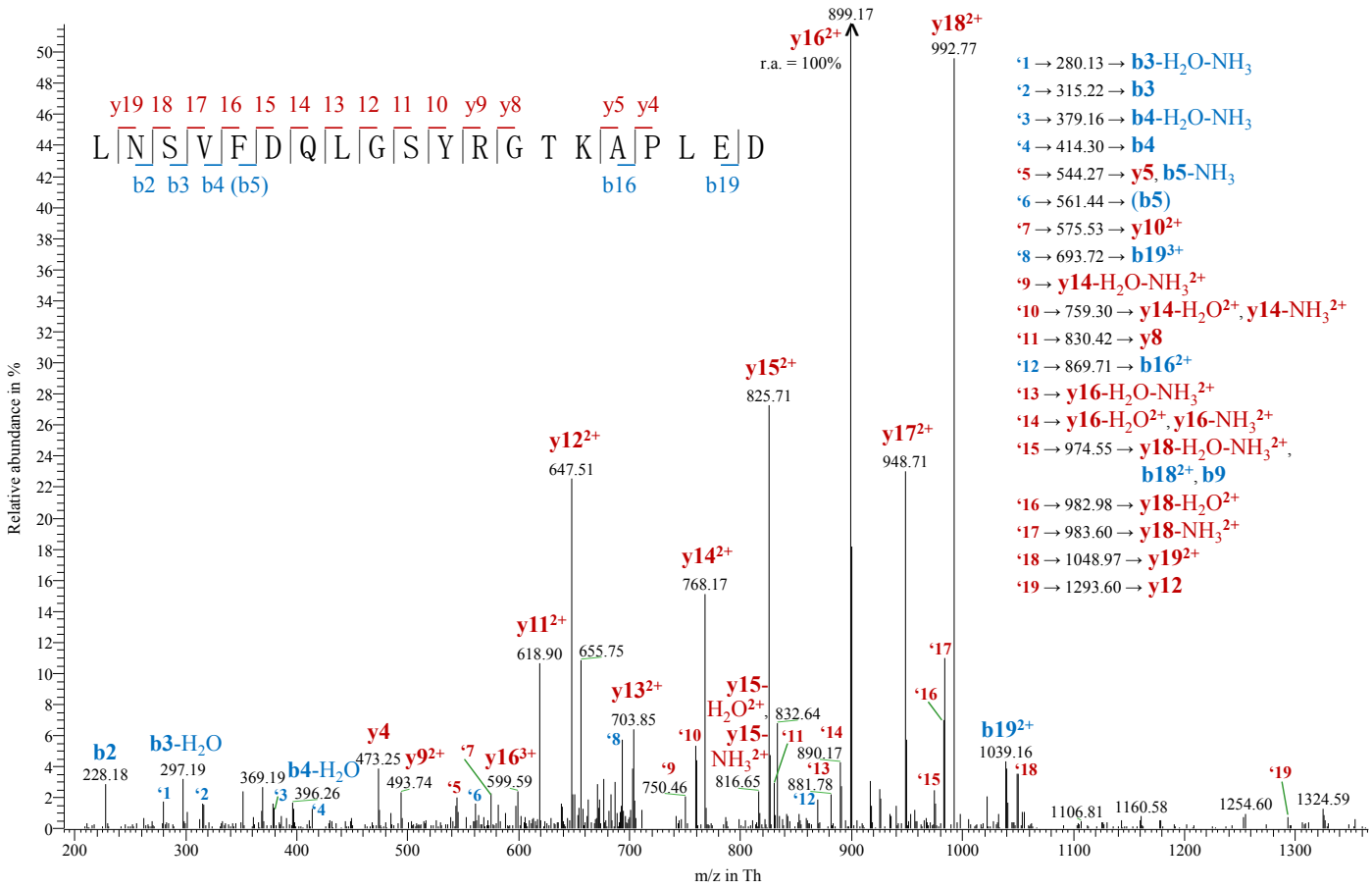
Supplementary Figure 9-33 | LTQ-MS^2 spectrum of LSNPSGSFPAPEFDNL, a SAV peptide of BALB/c and B6 mice derived from the *Abca13* gene. The peptide is encoded as LFNPSGSFPAPEFDNL in *Mus musculus musculus* (cf. Tables 3-4 and 3-6). See also Supplementary Note (Nat Comm, Theo).



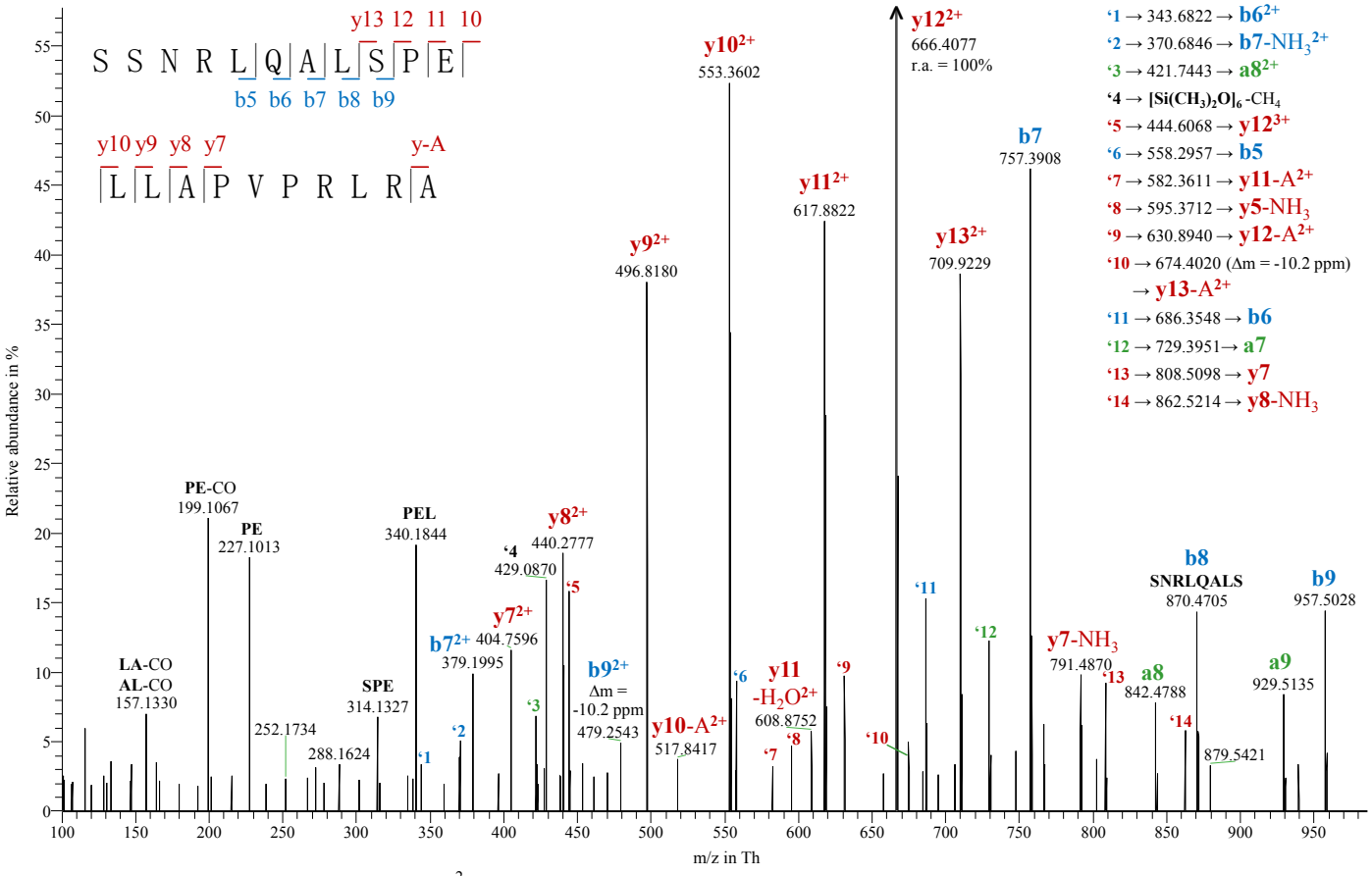
Supplementary Figure 9-34 | LTQ-MS^2 spectrum of LVKHKPKATAEQLKTVMDDF, a SAV peptide of BALB/c and B6 mice derived from the *Alb* gene. The peptide is encoded as LVKHKPKATSEQLKTVMDDF in *Mus musculus* subspecies *musculus* and *castaneus* (cf. Tables 3-4 and 3-6). See also Supplementary Note (Nat Comm, Theo).



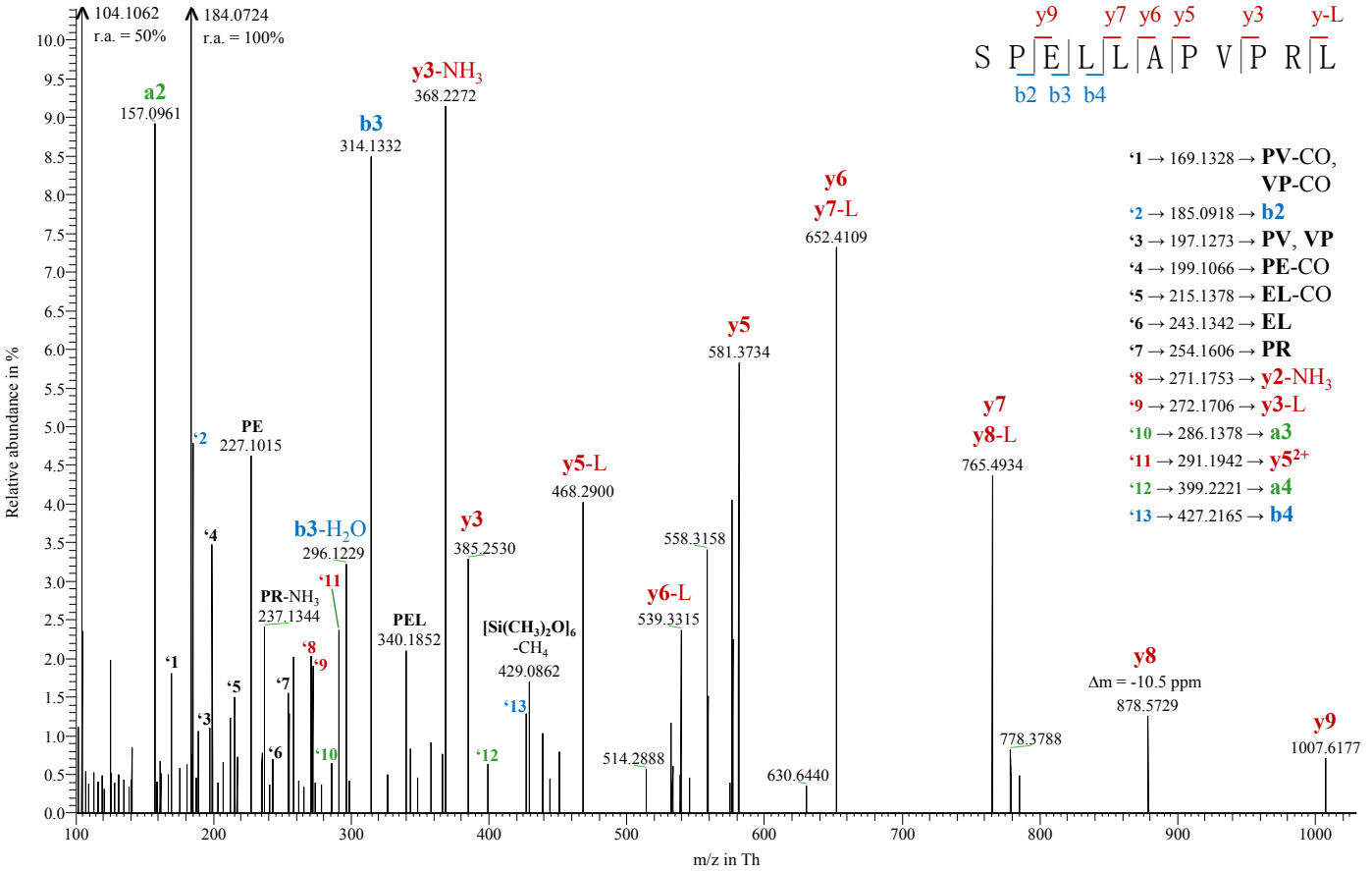
Supplementary Figure 9-35 | LTQ-MS^2 spectrum of WVDVERQVL, a SAV peptide of BALB/c and B6 mice derived from the *Egf* gene. The peptide is encoded as WVDVERQLL in *Mus musculus* subspecies *domesticus*, *musculus* and *castaneus* as well as *Mus spretus* (cf. Tables 3-4 and 3-6). See also Supplementary Note (Nat Comm, Theo).



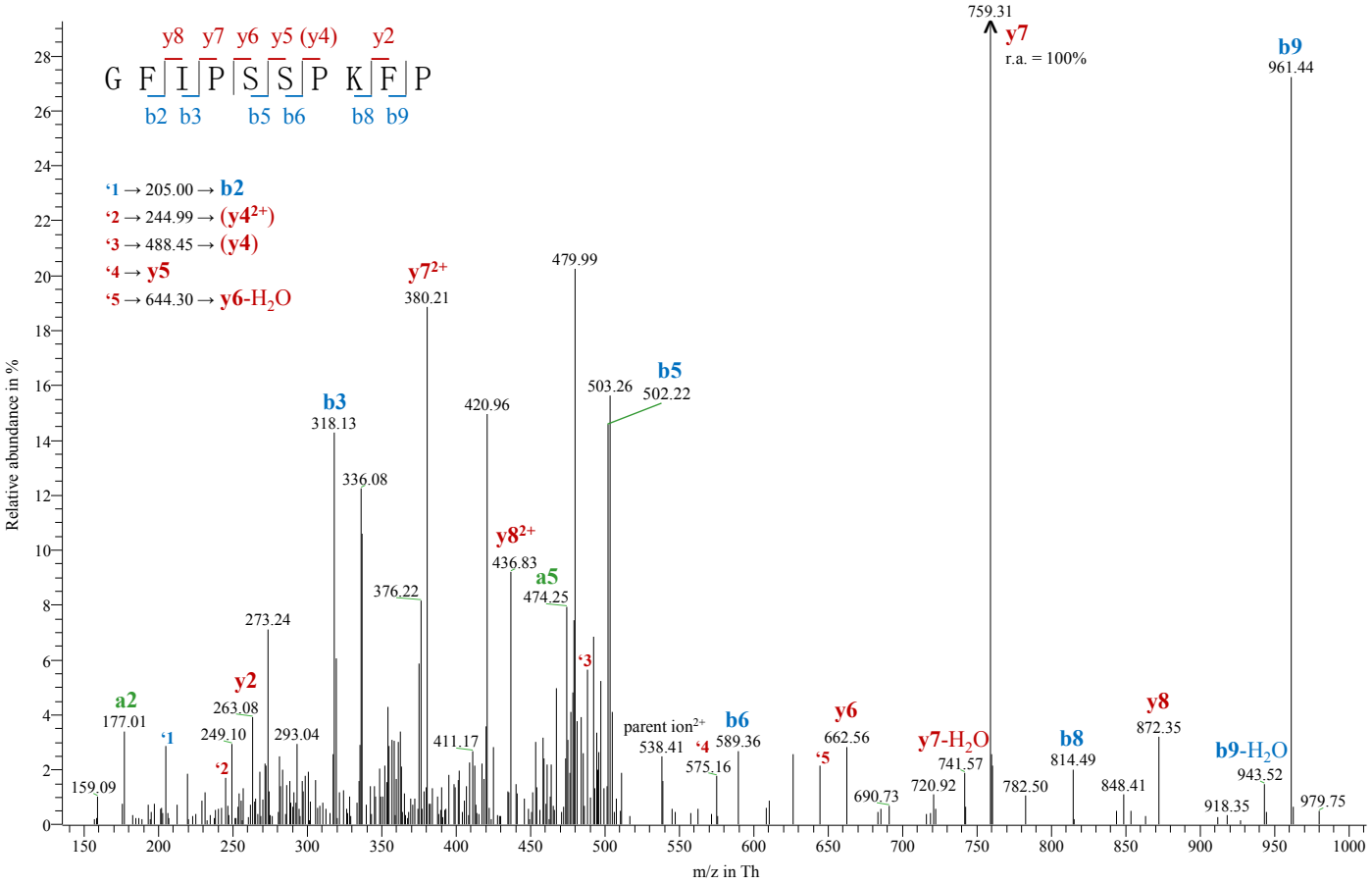
Supplementary Figure 9-36 | LTQ-MS^2 spectrum of LNSVFDQLGSYRGTKAPLED, a SAV peptide of BALB/c and B6 mice derived from the *Kap* gene. The peptide is encoded as LNSVFDRLGSYRGTKAPLED in *Mus musculus musculus* and as LNSVFDQLGSYRGTKSPLED in *Mus musculus castaneus* and at least two laboratory inbred mouse strains (NZL/LtJ and NZO/HlLtJ; cf. Tables 3-4 and 3-6). See also Supplementary Note (Nat Comm, Theo).



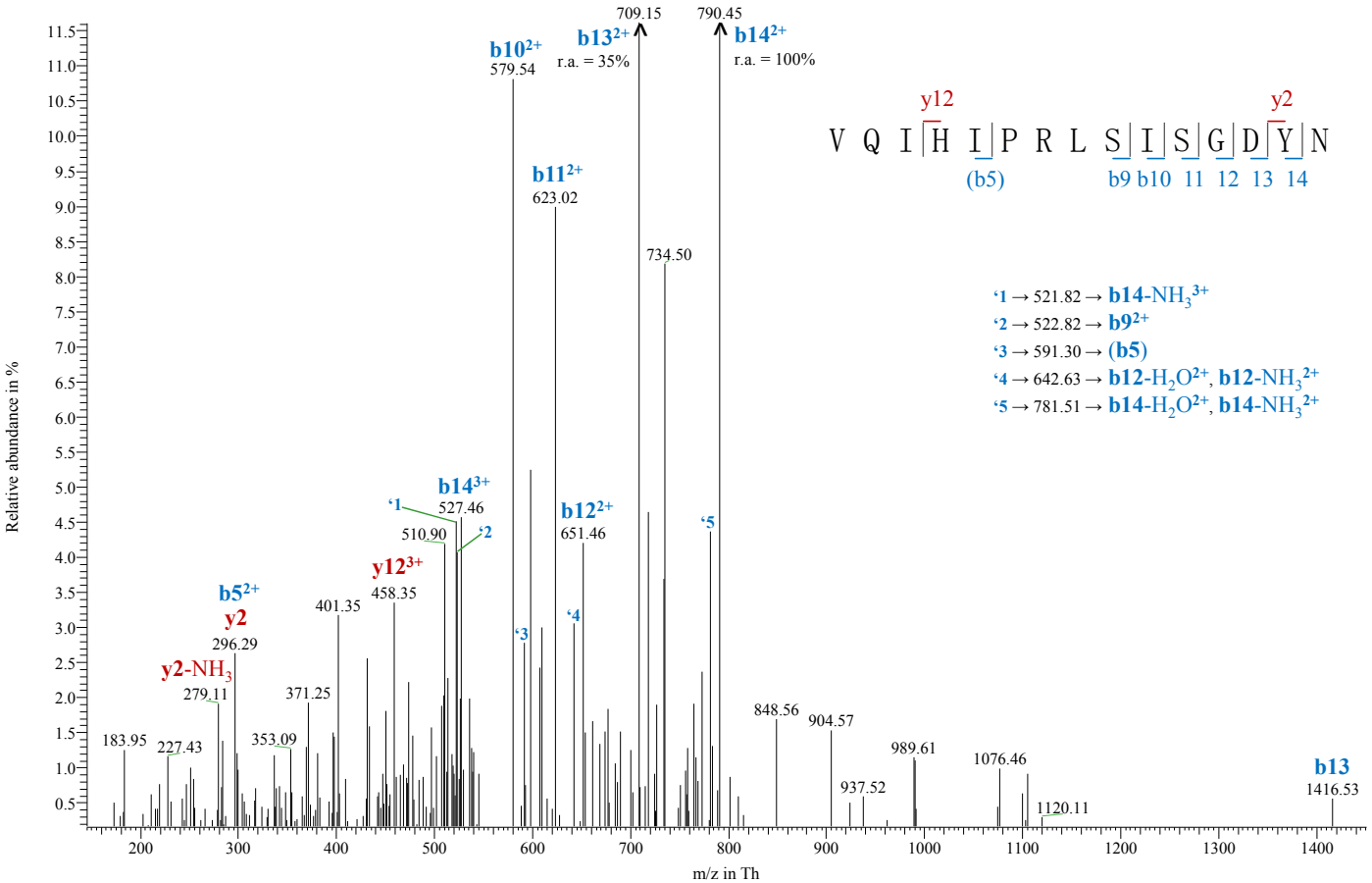
Supplementary Figure 9-37 | FT-Orbitrap-MS^2 spectrum of SSNRLQALSPELLAPVPRLRA, a SAV peptide of BALB/c and B6 mice derived from the *Lrg1* gene. The peptide is encoded as SSNRLQALFPELLAPVPRLRV in *Mus musculus musculus* and as SSNRLQALSPELLAPVPRLRV in *Mus musculus castaneus* and *Mus spretus* (cf. Tables 3-4 and 3-6). The upper half of the m/z range of this MS^2 spectrum is not depicted to enable a better visualisation of the lower mass range. However, only one of the ten peaks contained in the omitted part reached a relative abundance of > 5% (m/z = 1643.5082; r.a. = 14%). See also Supplementary Note (Nat Comm, Theo).



Supplementary Figure 9-38 | FT-Orbitrap-MS^2 spectrum of SPELLAPVPRL, a SAV peptide of BALB/c and B6 mice derived from the *Lrg1* gene. The peptide is encoded as FPELLAPVPRL in *Mus musculus musculus* (cf. Tables 3-4 and 3-6). See also Supplementary Note (Nat Comm, Theo).



Supplementary Figure 9-39 | LTQ-MS^2 spectrum of GFIPSSPKFP, a SAV peptide of BALB/c and B6 mice derived from the *Prol1* gene. The peptide is encoded as SFIPSSPKFP in *Mus musculus castaneus* and several laboratory inbred mouse strains (e.g. NONcNZO10/LtJ, NZB/BlNJ, NZL/LtJ, NZM2410/J and NZO/HlLtJ; cf. Tables 3-4 and 3-6). See also Supplementary Note (Nat Comm, Theo).



Supplementary Figure 9-40 | LTQ-MS^2 spectrum of VQIHIPRLSISGDYN, a SAV peptide of BALB/c and B6 mice derived from the *Serpina1b* gene. The peptide is encoded as VQIHIPRLSISGNYN in *Mus musculus musculus* and several laboratory inbred mouse strains (e.g. A/J, C3H/HeJ and NZW/LacJ; cf. Tables 3-4 and 3-6). See also Supplementary Note (Nat Comm, Theo).