Structure of a Pheromone Receptor-Associated MHC Molecule with an Open and Empty Groove

Citation

Published Version
doi:10.1371/journal.pbio.0030257

Permanent link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:4456979

Terms of Use
This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

Share Your Story
The Harvard community has made this article openly available. Please share how this access benefits you. Submit a story.

Accessibility
Neurons in the murine vomeronasal organ (VNO) express a family of class Ia major histocompatibility complex (MHC) proteins (M10s) that interact with the V2R class of VNO receptors. This interaction may play a direct role in the detection of pheromonal cues that initiate reproductive and territorial behaviors. The crystal structure of M10.5, an M10 family member, is similar to that of classical MHC molecules. However, the M10.5 counterpart of the MHC peptide-biding groove is open and unoccupied, revealing the first structure of an empty class I MHC molecule. Similar to empty MHC molecules, but unlike peptide-filled MHC proteins and non-peptide–binding MHC homologs, M10.5 is thermally unstable, suggesting that its groove is normally occupied. However, M10.5 does not bind endogenous peptides when expressed in mammalian cells or when offered a mixture of class I–binding peptides. The F pocket side of the M10.5 groove is open, suggesting that ligands larger than 8–10-mer class I–binding peptides could fit by extending out of the groove. Moreover, variable residues point up from the groove helices, rather than toward the groove as in classical MHC structures. These data suggest that M10s are unlikely to provide specific recognition of class I MHC–binding peptides, but are consistent with binding to other ligands, including proteins such as the V2Rs.

Recently, it was shown that the V2R class of pheromone receptors specifically interacts with members of the mouse M1 and M10 families [7] of major histocompatibility complex (MHC) class Ib proteins [8], which do not appear to have human orthologs [9]. Classical class I MHC molecules, which exhibit high polymorphism in mice, humans, and other mammals, present peptides derived from cytoplasmic proteins to T cells during immune surveillance, and are expressed on most or all nucleated cells [10]. The less polymorphic non-classical class Ib molecules are expressed on a more limited subset of cells and are involved in a variety of functions, including presentation of hydrophobic peptides (e.g., by Qa-2), presentation of formylated peptides by H2-M3, and lipid presentation by CD1 proteins [11]. The non-classical M1 and M10 proteins are expressed exclusively in the VNO and appear to facilitate cell surface expression of V2Rs. Male mice that are genetically deficient of class I MHC–associated β2-microglobulin (β2m) light chain show no surface expression of V2R pheromone receptors in the dendritic terminals of VNO sensory neurons.

Introduction

In most mammals, chemical communication between conspecific animals is involved in initiation of reproductive and territorial behaviors. The detection of these species- and gender-specific chemical cues, also called pheromones, is thought to involve receptors of the vomeronasal organ (VNO), a small neuronal epithelium located between the nasal cavity and the palate [1]. Although neurons of the main olfactory epithelium involved in odorant detection ultimately project to cognitive regions of the brain, vomeronasal neurons send inputs via the accessory olfactory bulb to specialized centers of the hypothalamus and amygdala, where they elicit neuroendocrine responses and behaviors such as oestrous synchronization, aggression, and sex discrimination [1–9].

Pheromone receptors belong to the ubiquitous family of G protein–coupled receptors (GPCRs), but are unrelated in sequence to main olfactory epithelium receptors that detect volatile odorants [4,5]. Mouse pheromone receptors can be divided into two subtypes, V1R and V2R, each of which is expressed in the dendritic tips of bipolar neurons in spatially distinct regions of the VNO. The human orthologs of most of these genes appear to be pseudogenes [1]. Mouse V1R receptors are found in the apical VNO domain, are thought to signal through the G-protein α-subunit Gzα2b, and exhibit sequence similarity to the T2R family of bitter taste receptors [6]. V2R receptors, in contrast, are found in the basal VNO domain, are likely to signal via the Gα subunits GAB,A (γ-aminobutyric acid-B), and calcium sensing receptors. V1R and V2R receptor family members, like all G protein–coupled receptors, contain seven putative transmembrane helices, but, in addition, V2R members include a large N-terminal extracellular domain.

Received April 6, 2005; Accepted May 18, 2005; Published July 12, 2005
DOI: 10.1371/journal.pbio.0030257

Copyright: © 2005 Olson et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abbreviations: β2m, β2-microglobulin; CD, circular dichroism; CHO, Chinese hamster ovary; ER, endoplasmic reticulum; MALDI-TOF, matrix-assisted laser desorption ionization time-of-flight; MHC, major histocompatibility complex; MUPS, major urinary proteins; NCS, non-crystallographic symmetry; VNO, vomeronasal organ

Academic Editor: Andrej Sali, University of California, San Francisco, United States of America

*To whom correspondence should be addressed. E-mail: bjorkman@caltech.edu
neurons and lack aggressive behavior toward intruder males [8]. Facilitation of V2R surface expression appears to involve M10 binding to a V2R because immunopurification of VNO receptors identified a multimolecular complex formed between M10, β2m, and a V2R protein. A given neuron generally expresses only one of the roughly 150 different subtypes of V2R receptor and one of the nine known VNO-specific MHC class Ib proteins (six M10 and three M1 family genes) [8,12]. Genetic analysis of individual neurons suggests specificity in binding between different V2R and M10 proteins [8].

Homologous receptors to the V2R proteins, such as the GABAB receptors and the umami and sweet taste receptors, also require accessory proteins for effective surface expression [13–15]. However, the nature of M10s as MHC molecules together with the association between the MHC and mating behavior in rodents [16] offers the intriguing possibility that M10 proteins play a direct role in mate or species recognition. Mice tend to choose mates with disparate MHC haplotypes [17], and pregnant mice have a higher incidence of spontaneous abortion if exposed to the scent of a male with a different MHC haplotype (the Bruce effect) [18]. These functions, as well as other species-specific indicators, may be mediated via M10-V2R complexes, perhaps by binding peptides or other accessory molecules in the canonical peptide-binding groove that exists in classical class I MHC structures [10]. Peptide binding during heavy and light chain assembly in the endoplasmic reticulum (ER), as occurs for classical class I MHC molecules, seems unlikely for M10 proteins because in situ hybridization assays demonstrate that TAP1 and TAP2, which are required for transport of cytoplasmic-derived peptides into the ER, are not expressed in the VNO [8]. However, a recent study shows that V2R-containing neurons are activated by nonamer class I MHC–binding peptides [19], suggesting involvement of exogenously acquired MHC–binding peptides, and perhaps M10 proteins, in individual mate or species recognition.

To determine the peptide-binding capability and structural characteristics of V2R receptor-associated MHC molecules, we expressed, characterized, and solved the crystal structure of the ectodomain of M10.5, an M10 family member. The M10.5 structure reveals an open conformation of the molecule. Thermal stability studies suggest that the M10.5 peptide-binding groove, which represents the first example, to our knowledge, of the structure of an empty class I MHC molecule (approximately 50% amino acid identity) for which a structure is available [24], thus we have used H-2Dd for comparisons. As in other class I structures, the first 180 residues of the M10.5 heavy chain form the α1–α2 domain superdomain, which is composed of an eight-stranded anti-parallel β-sheet platform topped by two anti-parallel α-helices. The following approximately 90 residues form the α3 domain, which resembles an immunoglobulin constant region domain (Figure 1A and 1B). The non-

**Results**

The M10.5 Structure

A soluble form of M10.5 was expressed together with human β2m in baculovirus-infected insect cells, and soluble M10.5-β2m complexes were purified from cell supernatants. Similar efforts to express soluble M10.5 together with mouse β2m in insect cells or Chinese hamster ovary (CHO) cells were unsuccessful (data not shown), perhaps related to the observation that mouse class I MHC molecules form stronger heterodimeric complexes with human, as compared with mouse, β2m [20,21]. M10.5 was crystallized in space group P212121 with five molecules in the asymmetric unit. Most crystals diffracted weakly to 3.5–4.0 Å resolution, but an incomplete dataset from a rare crystal that diffracted beyond 3.5 Å was combined with a more complete dataset to 4.0 Å (Table 1). The structure was solved by molecular replacement using the mouse class Ib MHC molecule Qa-2 [22] as a search model. Refinement using non-crystallographic symmetry (NCS) restraints yielded a final model (Rcryst = 26.6%; Rfree = 30.5%) (Table 1). Although data to 3.0 Å were included in the refinement, the high-resolution data are incomplete, thus the effective resolution of the structure is 3.2 Å. Two loops comprising residues 145–150 (α2 domain) and 194–197 (α3 domain) are missing in electron-density maps and are not included in our model.

The overall structure of M10.5 resembles the structures of classical class I MHC molecules. A BLAST search [23] identifies mouse H-2Dd as the most closely related classical class I MHC molecule (approximately 50% amino acid identity) for which a structure is available [24], thus we have used H-2Dd for comparisons. As in other class I structures, the first 180 residues of the M10.5 heavy chain form the α1–α2 domain superdomain, which is composed of an eight-stranded anti-parallel β-sheet platform topped by two anti-parallel α-helices. The following approximately 90 residues form the α3 domain, which resembles an immunoglobulin constant region domain (Figure 1A and 1B). The non-

**Table 1. Crystallographic Data and Refinement Statistics**

<table>
<thead>
<tr>
<th>Category</th>
<th>Characteristic</th>
<th>Strict NCS</th>
<th>Restrained NCS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data collection</td>
<td>Wavelength (Å)</td>
<td>1.00879, 1.11587</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Resolution (Å)</td>
<td>100–3.0 (3.31–3.2)</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Unique reflections</td>
<td>40,549</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Total reflections</td>
<td>262,067</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Mean redundancy</td>
<td>63 (5.2)</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Completeness (%)</td>
<td>95.5 (83.9)</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Rcryst (%)</td>
<td>15.4 (43.7)</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Rfree (%)</td>
<td>9.2 (3.2)</td>
<td>-</td>
</tr>
<tr>
<td>Refinement statistics</td>
<td>Resolution (Å)</td>
<td>50–3.0b</td>
<td>50–3.0b</td>
</tr>
<tr>
<td></td>
<td>Reflections in working set</td>
<td>44,543</td>
<td>44,543</td>
</tr>
<tr>
<td></td>
<td>Reflections in test set</td>
<td>2,355</td>
<td>2,355</td>
</tr>
<tr>
<td></td>
<td>Rcryst (%)</td>
<td>30.9</td>
<td>26.6</td>
</tr>
<tr>
<td></td>
<td>Rfree (%)</td>
<td>30.7</td>
<td>30.3</td>
</tr>
<tr>
<td></td>
<td>Atoms (%)</td>
<td>2,733</td>
<td>13,720</td>
</tr>
<tr>
<td></td>
<td>RMS deviations from ideality</td>
<td>RMS deviations bonds (Å)</td>
<td>0.009</td>
</tr>
<tr>
<td></td>
<td>RMS deviations angles (°)</td>
<td>1.84</td>
<td>1.72</td>
</tr>
<tr>
<td></td>
<td>Ramachandran plot (%)</td>
<td>83.8</td>
<td>83.5</td>
</tr>
<tr>
<td></td>
<td>Additionally allowed</td>
<td>14.6</td>
<td>13.7</td>
</tr>
<tr>
<td></td>
<td>Generously allowed</td>
<td>1.3</td>
<td>2.4</td>
</tr>
<tr>
<td></td>
<td>Disallowed</td>
<td>0.3</td>
<td>0.4</td>
</tr>
</tbody>
</table>

[^1]: Refinement (R) = Σ|Fobs| – |Fcalc|/Σ|Fobs|, where i is the i-th observation of the intensity of unique reflection hkl and |Fobs| is the mean intensity of reflection hkl. Numbers in parenthesis indicate the highest resolution shell.

[^2]: Refinements between 3.2 and 3.0 Å resolution were included in the refinement, but due to insufficient completeness, we are reporting this as a 3.2 Å structure.

[^3]: Rcryst (f) = Σ(|Fobs| – |Fcalc|)/Σ|Fobs|, where |Fobs| and |Fcalc| are the observed and calculated structure factor amplitudes, respectively, Rcryst is the crystallographic R-factor calculated for 5% of the data withdrawn from refinement. RMS, root mean square.

DOI: 10.1371/journal.pbio.0030257.s001
covalently attached β2m light chain, also resembling an immunoglobulin constant region, contacts both the underside of the α1–α2 platform and one β-sheet of the α3 domain in an orientation consistent with previously solved mouse and human class I MHC structures, indicating that pairing human β2m with the mouse M10.5 heavy chain does not disrupt the overall M10.5 structure.

Ordered N-linked carbohydrate is observed attached to Asn223 within the loop joining the third and fourth β-strand in the α3 domain. The carbohydrate occludes the counterpart of the region in class I MHC molecules that is the major determinant for binding the T cell co-receptor CD8 [25], likely preventing M10.5 from participating in CD8+ T cell-mediated immunological responses. Ordered carbohydrate is not observed at either of the two other predicted N-linked glycosylation sites (Asn62 and Asn198), but the quality of the electron-density map is poor in these regions.

The α1–α2 Platform of M10.5 Contains an Open, but Apparently Empty, Groove

A large groove between the α1 and α2 domain helices forms the peptide-binding site in classical class I MHC molecules [10]. Structural studies of class I MHC molecules and homologs show a correlation between the degree of separation of the α1–α2 domain helices and the ability to bind peptides or other small molecules [26]. The helices are separated by approximately 18 Å in the center of the grooves of classical peptide-binding class I MHC molecules such as H-2Dd [24], and class I MHC-related proteins that bind other small molecule ligands also contain open grooves with
separated α1 and α2 domain helices [27–31]. In contrast, the neonatal Fc receptor (FcRn) [32], the hemochromatosis protein HFE [33], and MIC-A [34], MHC homologs that do not bind small molecule ligands, have collapsed grooves with a smaller separation between the α1–α2 domain helices [32–34].

A superposition of the α1–α2 platforms of M10.5, H-2D^d, FcRn, and HFE illustrates the variation in groove size (Figure 1C) and demonstrates that M10.5 has an open groove more similar to the peptide-binding classical class I MHC molecules than the non–peptide-binding homologs. The M10.5 α1–α2 domain helices superimpose well with the analogous H-2D^d helices (RMS [root mean square] deviation of 1.22 Å for 181 α-carbon atoms), with the largest differences located in the region of the α2 domain helix immediately preceding six residues that are disordered in M10.5 (residues 145 to 150). The overall similarity of the M10.5 and H-2D^d α1–α2 platforms, which are contained within class I heavy chains that are paired with human (M10.5) or mouse (H-2D^d) β2m light chains, demonstrates that the observed structure of the M10.5 α1–α2 platform does not result from an artifactual change induced by pairing with human, rather than mouse, β2m. The more open character of the M10.5 groove as compared with the grooves of non–peptide-binding MHC homologs is seen when the calculated accessible surface areas are compared: approximately 730 Å^2 for M10.5 compared with approximately 780 Å^2 for typical class I MHC molecules [30,39], approximately 690 Å^2 for H-2D^d, and approximately 415 Å^2 and approximately 235 Å^2 for HFE and FcRn, respectively [33] (see Materials and Methods). Thus the M10.5 groove can be classified as “open” and capable of binding to a ligand.

The α1–α2 domain groove is occupied by a peptide or mixture of peptides in all class I MHC structures solved to date, and continuous electron-density representing peptide(s) is always seen in the α1–α2 groove [35]. Indeed, attempts to crystallize an empty version of H-2K^b, which was expressed in cells lacking peptide-loading machinery and purified in the absence of a binding peptide, resulted in a crystal structure that revealed a peptide derived from the cell growth media [36]. M10.5 omit electron-density maps in which the α1 and α2 helices were removed from structure factor calculations return clear density for the omitted region (Figure 1B), indicating that the maps are of sufficient quality to detect bound molecules of the size of an 8–10 residue peptide. However, M10.5 electron-density maps show no ordered density corresponding to a peptide: Annealed F_o-F_c maps calculated using NCS constraints or tight restraints do not show continuous electron density within the groove (Figure S1), and maps calculated before or after 5-fold real space averaging also failed to reveal unbroken density in the cleft (data not shown). The possibility that a minor peak near the center of the groove in averaged maps represents a portion of a bound polyethylene glycol molecule is discussed in the caption of Figure 1. We conclude that the M10.5 groove does not contain a single defined peptidic or non-peptidic occupant or a mixture of compounds with a similar conformation.

Comparison of M10.5 and Classical Class I MHC Grooves

Crystallographic studies of classical class I MHC–peptide complexes have defined six pockets (A–F) within the peptide-binding groove that interact with various portions of bound peptides (Figure 2A) [37–39]. The A and F pockets at either end of the peptide-binding groove are largely conserved and interact with the N- and C-terminus, respectively, of the bound peptide [37,39]. The B, C, D, and E pockets contain residues that vary between alleles, resulting in different allele-specific peptide-binding preferences.

Figure 2. Groove Surface Characteristics
(A) Electrostatic surface representations calculated using GRASP [42] of the M10.5 (left) and H-2D^d (right) grooves. Positions of the six pockets are labeled A–F in yellow. In H-2D^d, Arg62 and Glu163 create a bridge over the peptide-binding groove. In M10.5, Glu63 and Arg167 define a similar feature. The H-2D^d groove is constricted between Asn70 and Arg155. In M10.5, Ala70 and Gly155 lead to a wider groove and continuous D and E pockets. (B) Comparison of residues in the A and F pockets of classical class I MHC molecules and M10.5. All alanine-peptide (yellow, atoms color-coded according to atom type) is derived from the H-2D^d structure [24]. Left: In class I MHC proteins, a cluster of four tyrosine residues in the A pocket (Tyr7, Tyr59, Tyr159, and Tyr171) form hydrogen bonds with backbone atoms in the peptide N-terminus (residues are listed in single letter code for M10.5 and H-2D^d, respectively). In M10.5, Tyr7 is replaced by a threonine residue and Tyr171 is replaced with a cysteine. Five of the six M10 proteins have an additional tyrosine at position 33 that could potentially replace one of the two missing tyrosine residues. M10.5 has a serine instead of a glycine at position 26, which could participate in additional hydrogen-binding interactions. Right: The F pocket of class I MHC molecules is blocked on one end by Thr80, Tyr84, and Lys146. The M10.5 pocket is open on this end (see panels A and C) due to the substitution of a glutamate residue for Tyr84 and the disorder of the loop containing residue 146 (Asp in M10.5). Lys142 in M10.5 is missing sidechain density and has been modeled as an alanine, but most likely would not further occlude the M10.5 F pocket. (C) Molecular surface representations of M10.5 and H-2D^d with an all alanine-peptide (yellow) derived from the H-2D^d structure [24] superimposed in the M10.5 groove. M10.5 atoms that came within 2.5 Å of the peptide trace were considered clashes and are colored red. Four additional alanine residues (green) were added to the C-terminus of the H-2D^d-binding peptide to illustrate that peptides binding in the M10.5 groove could extend out of the F pocket side.
DOI: 10.1371/journal.pbio.0030257.g002
The A and F pocket regions of the M10.5 groove contain substitutions that prevent the interactions that anchor peptide termini into the groove of a classical class I MHC molecule. Two of the four conserved tyrosines in the MHC class I A pocket (tyrosines 7, 59, 159, and 171), which hydrogen bond directly or through a water molecule with main chain atoms of the peptide N-terminal residue [37,39], are substituted in M10.5 as Thr7 and Cys171 (Figure 2B). Substitutions in other M10s also eliminate one or two of the four A pocket tyrosines [8,12]. It remains possible, however, that M10.5 and other M10s could bind a peptide N-terminus in a non-classical manner, utilizing residues including Thr7, Ser26, and Tyr33 as hydrogen-bond donors. Residues that typically anchor the C-terminus of a peptide in the class I MHC F pocket are also different in M10.5. In H-2Dd, the ninth and tenth residue of the bound peptide (P9 and P10) form hydrogen-bonding and van der Waals contacts with Val76, Asp77, Tyr84, Thr143, Lys146, and Trp147 (Figure 2B). In M10.5 and most other M10s, these residues are replaced by non-conservative substitutions, and the last two are missing entirely from the M10.5 electron-density map. Thus potential interactions between M10 F pocket groove residues and peptides would have to occur in a non-classical manner.

An interesting feature of possible functional relevance is the relatively open nature of the F pocket side of the M10.5 groove (Figure 2C), suggesting that potential small molecule or peptide occupants could extend out of this side of the groove. Homology models of other M10 proteins constructed using the M10.5 structure and the Swiss-model Protein modeling Server [40] suggest that this open character is a general feature of M10 proteins (data not shown). In class I MHC molecules, Tyr84 and Lys146 occlude the F pocket side of the groove, but the tyrosine to glutamate substitution at position 84 in M10.5 (Figure 2B) contributes to the more open M10.5 groove. In addition, M10.5 Asp146 is part of the disordered 145–150 region of the α2 domain helix, thus it does not contribute to closing the F pocket side of the groove. The disorder of these residues may be related to a need to maintain flexibility in the F pocket end of the M10.5 groove to allow binding of ligands that extend out of the M10.5 groove.

<table>
<thead>
<tr>
<th>Residue</th>
<th>H-2Dd</th>
<th>H-2Dd Pocket</th>
<th>M10.5</th>
<th>M10.5 Pocket</th>
<th>Peptide Clash</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Leu</td>
<td>Buried</td>
<td>Leu</td>
<td>A</td>
<td>No</td>
</tr>
<tr>
<td>7</td>
<td>Tyr</td>
<td>A,B</td>
<td>Thr</td>
<td>A,B</td>
<td>No</td>
</tr>
<tr>
<td>9</td>
<td>Val</td>
<td>Buried</td>
<td>Arg</td>
<td>C,D</td>
<td>Yes</td>
</tr>
<tr>
<td>24</td>
<td>Glu</td>
<td>Buried</td>
<td>Gin</td>
<td>B</td>
<td>No</td>
</tr>
<tr>
<td>26</td>
<td>Gly</td>
<td>Buried</td>
<td>Ser</td>
<td>A,B</td>
<td>No</td>
</tr>
<tr>
<td>33</td>
<td>Phe</td>
<td>Buried</td>
<td>Tyr</td>
<td>A</td>
<td>No</td>
</tr>
<tr>
<td>34</td>
<td>Val</td>
<td>Buried</td>
<td>Gin</td>
<td>A</td>
<td>No</td>
</tr>
<tr>
<td>52</td>
<td>Ile</td>
<td>Buried</td>
<td>Met</td>
<td>A</td>
<td>No</td>
</tr>
<tr>
<td>55</td>
<td>Glu</td>
<td>Buried</td>
<td>Glu</td>
<td>A</td>
<td>No</td>
</tr>
<tr>
<td>59</td>
<td>Tyr</td>
<td>A</td>
<td>Tyr</td>
<td>A</td>
<td>No</td>
</tr>
<tr>
<td>63</td>
<td>Glu</td>
<td>A,B</td>
<td>Glu</td>
<td>A,B</td>
<td>No</td>
</tr>
<tr>
<td>66</td>
<td>Arg</td>
<td>B</td>
<td>His</td>
<td>B</td>
<td>No</td>
</tr>
<tr>
<td>70</td>
<td>Asn</td>
<td>A,B,C</td>
<td>Ala</td>
<td>B,C</td>
<td>No</td>
</tr>
<tr>
<td>74</td>
<td>Phe</td>
<td>C</td>
<td>Ala</td>
<td>C</td>
<td>No</td>
</tr>
<tr>
<td>80</td>
<td>Thr</td>
<td>F</td>
<td>Thr</td>
<td>F</td>
<td>No</td>
</tr>
<tr>
<td>81</td>
<td>Ala</td>
<td>F</td>
<td>Met</td>
<td>F</td>
<td>No</td>
</tr>
<tr>
<td>95</td>
<td>Leu</td>
<td>F</td>
<td>Leu</td>
<td>C</td>
<td>No</td>
</tr>
<tr>
<td>97</td>
<td>Trp</td>
<td>C</td>
<td>Glu</td>
<td>C</td>
<td>No</td>
</tr>
<tr>
<td>99</td>
<td>Ala</td>
<td>C,D</td>
<td>Phe</td>
<td>B,D</td>
<td>Yes</td>
</tr>
<tr>
<td>114</td>
<td>Trp</td>
<td>D,E</td>
<td>Tyr</td>
<td>E</td>
<td>No</td>
</tr>
<tr>
<td>116</td>
<td>Phe</td>
<td>C,E,F</td>
<td>Leu</td>
<td>C,E,F</td>
<td>No</td>
</tr>
<tr>
<td>122</td>
<td>Asp</td>
<td>Buried</td>
<td>Asp</td>
<td>F</td>
<td>No</td>
</tr>
<tr>
<td>123</td>
<td>Tyr</td>
<td>F</td>
<td>Tyr</td>
<td>Buried</td>
<td>No</td>
</tr>
<tr>
<td>124</td>
<td>Ile</td>
<td>Buried</td>
<td>Ile</td>
<td>E,F</td>
<td>No</td>
</tr>
<tr>
<td>133</td>
<td>Trp</td>
<td>Buried</td>
<td>Trp</td>
<td>E</td>
<td>No</td>
</tr>
<tr>
<td>139</td>
<td>Ala</td>
<td>Buried</td>
<td>Val</td>
<td>F</td>
<td>No</td>
</tr>
<tr>
<td>140</td>
<td>Ala</td>
<td>Buried</td>
<td>Gly</td>
<td>F</td>
<td>No</td>
</tr>
<tr>
<td>143</td>
<td>Thr</td>
<td>F</td>
<td>Phe</td>
<td>F</td>
<td>No</td>
</tr>
<tr>
<td>147</td>
<td>Trp</td>
<td>E,F</td>
<td>Arg</td>
<td>Missing</td>
<td>No</td>
</tr>
<tr>
<td>152</td>
<td>Ala</td>
<td>E</td>
<td>Val</td>
<td>D,E</td>
<td>No</td>
</tr>
<tr>
<td>156</td>
<td>Asp</td>
<td>E</td>
<td>Trp</td>
<td>D</td>
<td>No</td>
</tr>
<tr>
<td>159</td>
<td>Tyr</td>
<td>A,D</td>
<td>Tyr</td>
<td>A,D</td>
<td>No</td>
</tr>
<tr>
<td>167</td>
<td>Trp</td>
<td>A</td>
<td>Arg</td>
<td>A</td>
<td>Yes</td>
</tr>
<tr>
<td>171</td>
<td>Tyr</td>
<td>A</td>
<td>Cys</td>
<td>A</td>
<td>No</td>
</tr>
</tbody>
</table>

Conserved (blue) or highly variable (red) residues are defined as in Figure 4.

*Groove residues have 5.0 Å² or more of accessible surface area to a 1.4 Å radius probe but less than 5.0 Å² to a 5.0 Å radius probe. “Buried” residues have less than 5.0 Å² of accessible surface area to a 1.4 Å radius probe. “Missing” residues are disordered.

*Residues with atoms less than 2.5 Å from a polyalanine version of an H-2Dd-binding peptide superimposed in the M10.5 groove.

*Residues that have accessible surface area in the groove but have more than 5.0 Å² of accessible surface area to a 5.0 Å radius probe.

DOI: 10.1371/journal.pbio.0030257.t002
In order to make additional comparisons between the grooves of M10.5 and peptide-binding class I MHC molecules, we used a previously described algorithm [30,33] to identify residues that comprise the grooves of M10.5 and H-2D\(^d\). Residues that line the M10.5 and H-2D\(^d\) grooves were defined as having more than 5.0 Å\(^2\) accessible surface area to a 1.4 Å probe and less than 5.0 Å\(^2\) to a 5.0 Å probe (Table 2, Figure 3A and 3B), resulting in 33 groove residues in M10.5 and 22 groove residues in H-2D\(^d\). The additional groove residues in M10.5 that are buried in H-2D\(^d\) are divided into two clusters, one in the area surrounding the A and B pockets, and the other around the F pocket (see Figure 2A). The M10.5 A pocket is larger than its H-2D\(^d\) counterpart due to the substitution of class I MHC residues Tyr7 and Tyr171 with less-bulky threonine and cysteine residues, respectively, and a different location of the sidechain of M10.5 Arg167 as compared to H-2D\(^d\) Trp167. Arg167 was postulated to likely block peptide binding in M10.5 [8] by analogy to Arg167 in FcRn, which partially occludes the A-B pocket [32]. However, the sidechain of Arg167 in M10.5 is partially disordered, suggesting mobility that could result in a rearrangement to allow a groove occupant to access the open pocket area below the arginine sidechain. Other differences that create a larger M10.5 A pocket compared to H-2D\(^d\) involve several large aromatic residues lining the H-2D\(^d\) groove that are replaced by smaller non-aromatic residues in M10.5. These include H-2D\(^d\) Tyr7, Phe74, Trp167, Phe116, Trp167, and Tyr171, which are substituted as Thr7, Ala74, Glu97, Leu116, Arg167, and Cys171 in M10.5. Another difference between the M10.5 and H-2D\(^d\) grooves is a considerable enlargement of the M10.5 D pocket, which is continuous with the E pocket. The enlargement is caused by the replacement of H-2D\(^d\) residues Asn70 and Arg155 with alanine and glycine residues, respectively, removing the constriction that sandwiches the peptide residue at position 4 (glycine) in the H-2D\(^d\) structure [24].

The chemical character of the groove of MHC–related protein can sometimes reveal the nature of its ligand. For example, the largely hydrophobic grooves of CD1 [30] and the class Ib MHC molecules Qa-2 [22] and HLA-E [41] allow binding of lipids (CD1) and hydrophobic peptides (the class Ib proteins). By contrast, the M10.5 groove, like the grooves of H-2D\(^d\) and other classical class I MHC molecules, contains a mixture of polar and non-polar residues (Table 2). A surface potential map generated by GRASP [42] shows that the A, C, E, and F pockets of M10.5 are slightly acidic whereas the B and D pockets are uncharged (see Figure 2A). The sidechain of Arg9 (a buried valine in H-2D\(^d\)) points into the center of the M10.5 groove and forms a salt-bridge with the sidechain of Glu97. A second salt-bridge between Glu63 and Arg167 bridges the A and B pockets. The groove in H-2D\(^d\) is primarily acidic, with a similar salt bridge formed between Arg62 and Glu163 over the A–B pocket boundary (Figure 2A). Peptides bound to H-2D\(^d\) are accommodated under the salt bridge [24], thus the salt bridge in the M10.5 groove would not necessarily prevent binding of a peptide or other small molecule.

**Sequence Conservation and Receptor Binding**

Like classical class I MHC molecules, variability within M10 proteins is mainly localized to the \(\alpha_1–\alpha_2\) platform, with the \(\alpha_3\) domain being more constant [8,12]. As previously predicted [8] and now confirmed by the M10.5 structure, amino acids with the highest degree of sequence variability within the M10 and M1 families cluster on the top face of the \(\alpha_1\) and \(\alpha_2\)
Figure 4. Sequence Conservation in M10 and H-2 Loci

(A) Ribbon diagram of the M10.5 α1–α2 platform with positions of residues that are 100% identical in the nine M10 and M1 alleles colored blue; residues that are among the top 10% most variable are colored red. Variability is determined by the number of amino acids at a given position divided by the frequency of the most common allele at that position. Residues 145–150, which are disordered, are designated by a dashed line.

(B) Ribbon diagram of the H-2Dα α1–α2 domain. Conserved (blue) and highly variable (red) residues are derived from an alignment of 19 H-2D alleles [80].

DOI: 10.1371/journal.pbio.0030257.g004

domain helices. Indeed, three of the six disordered residues (145–150) in the M10.5 α2 domain helix belong to the top 10% of variable residues within the M10 α1 and α2 domains. By contrast, M10 residues that point into the groove show less variability (Figure 4A). This pattern of variability is opposite to the pattern in classical MHC class I molecules, in which residues that display the greatest sequence variability point towards the peptide-binding groove, resulting in allelic specificity for binding peptides, whereas residues on the tops of the α1 and α2 domain helices are more conserved [43,44] (Figure 4B).

M10.5 Is Thermally Unstable

Classical class I MHC molecules and UL18 are thermally unstable in the absence of bound peptide [21,45,46]. In contrast, non-peptide binding class I MHC homologs, such as FcRn and the hemochromatosis protein HFE, do not show reduced thermal stability, presumably due to structural rearrangements that close the counterparts of their peptide-binding grooves [32,33,47]. The M10.5 groove is open, but apparently unoccupied, in the crystal structure, raising the question of whether M10.5 is stable in the absence of a groove occupant.

To determine the thermal stability of M10.5, we monitored heat-induced unfolding by recording the circular dichroism (CD) signal at 223 nm as a function of increasing temperature. Two unfolding transitions, an upward-sloping transition with a Tm of 43 °C and a downward-sloping transition with a Tm of 64 °C, are evident in the melting curve of insect cell-derived M10.5 (Figure 5A). The M10.5 melting curve and derived Tm values are similar to previously reported results for M10.5 Is Thermally Unstable

Crystal Structure of the MHC Ib Molecule M10.5

Thermal denaturation of M10.5 (A) compared with empty and peptide-filled H-2Kd (B) (modified from Figure 3A in Fahnestock et al. [21]). The CD signal at 223 nm was monitored as a function of temperature for insect cell-derived M10.5 and empty and peptide-filled versions of soluble H-2Kd produced in CHO cells. Tm values for the melting of the M10.5 heavy chain and β2m light chain (marked with arrows) were derived by estimating the half-point of the ellipticity change between the beginning and end of each transition. The M10.5 denaturation profile is similar to the profile obtained from empty H-2Kd, suggesting that the M10.5 groove is not occupied. The possibility that polyethylene glycol, a component in the crystallization solutions, might bind to and stabilize M10.5 was investigated by repeating the unfolding experiment in the presence of 20% polyethylene glycol 1000, with no significant changes to the thermal stability profile (data not shown).

DOI: 10.1371/journal.pbio.0030257.g005

to the low thermal stability of M10.5, the peptide-filled form of the H-2Kd heavy chain melts with a Tm of approximately 57 °C [21] (Figure 5B) and FcRn melts with a Tm of approximately 62 °C at pH 6 [47]. The low thermal stability of purified M10.5 suggests that a ligand or binding partner not present in the purified preparation is necessary to stabilize the native conformation at 37 °C in vivo.

M10.5 Does Not Associate with Endogenous Peptides

The empty groove in the M10.5 structure could result from expression in invertebrate cells, which do not possess the machinery to load peptides into class I MHC molecules [48,49]. When classical class I MHC molecules are expressed in vertebrate cells or purified from native sources, mixtures of short peptides derived from cytoplasmic proteins can be eluted [50]. Sequence analyses of eluted peptides revealed that any given class I allele can bind a wide variety of short (8- to 10-mer) peptides that conform to a particular allelic-specific peptide-binding motif involving preferred residues at the peptide C-terminus and an internal position (usually position 2 or 5/6) [50]. To determine whether M10.5 binds endogenous or exogenous peptides with these or other characteristics, we expressed the M10.5 ectodomain in CHO cells, mammalian cells that support peptide loading, and examined acid eluates from purified M10.5 using methods previously used to identify peptides eluted from class I MHC molecules [51].

Purified M10.5-β2m heterodimers expressed in CHO and in insect cells were treated with acetic acid to dissociate potential peptide material. Insect cell-derived M10.5 was analyzed before and after incubation with a mixture of short (8- to 9-mer) peptides. As it would be impossible to test all possible 8- to 9-mer peptides, we prepared a mixture of peptides including those reported to activate V2R-expressing neurons [19] in order to see if any of these class I MHC-
binding peptides bind to M10.5. We used soluble versions of two other β2m-binding class I MHC–like proteins expressed in CHO cells as controls: UL18, a viral class I MHC homolog that binds endogenous peptides resembling class I MHC–binding peptides [52], and FcRn, a class I MHC homolog that does not bind peptides or contain a non-peptidic groove occupant [32,47].

Low molecular weight acid eluates derived from M10.5 and the control proteins were sequenced by Edman degradation and analyzed by MALDI-TOF (matrix-assisted laser desorption ionization time-of-flight) mass spectrometry. Acid eluates from CHO–derived M10.5 and both peptide-treated and untreated insect cell–derived M10.5 resemble eluates from FcRn, the non–peptide-binding class I homolog, rather than eluates from UL18, a peptide-binding protein (Table 3). With the exception of cycle 1, which typically shows a high background for FcRn and other proteins that do not bind peptides [52], the total yield of the amino acids from each cycle of pool sequencing of the M10.5 acid eluates remains nearly constant and not significantly above background. By contrast, the UL18 acid eluate shows the presence of a mixture of peptides with characteristics similar to those previously reported [52]. MALDI-TOF analysis of the acid eluates indicated molecules with masses consistent with peptides in the UL18 sample, but not in the negative control or M10.5 samples (data not shown). Thus M10.5 does not appear to bind any form of peptide, including N-terminally blocked peptides, or other small molecule ligand when expressed in CHO or insect cells. In addition, purified M10.5 does not bind any of a collection of class I MHC–binding peptides, as would have been expected if M10 proteins were the molecules responsible for binding the class I MHC–binding peptides reported to activate VNO neurons [19].

### Computer Modeling of Potential M10.5 Groove Occupants

Given the open character of the M10.5 groove and the thermal instability of M10.5–β2m heterodimers, we believe that the groove is likely to be a binding site for a peptide, protein, or a non-peptidic small molecule ligand. Here we use computer modeling to evaluate which potential ligands can fit into the M10.5 groove.

We first consider whether the M10.5 groove can accommodate a class I MHC–binding peptide by superimposing a polyanalanine version of an H-2Dd–binding peptide [24] or a Qa-2–binding peptide [22] on the M10.5 structure (see Figure 2C). The H-2Dd–binding peptide fits into most regions of the M10.5 groove, but clashes with three residues in the A and B pockets (see Table 2). The Qa-2 peptide adopts a sharper upward bend near residue P2 (data not shown), and consequently, two of these clashes (Arg9 and Phe99) do not occur. The third clash involves Arg167 with the N-terminus of the peptide, but arginine sidechains are often flexible and might be able to move out of the way of the N-terminus of a bound peptide. Several residues in the vicinity of Arg167 have smaller sidechains in M10.5 than in H-2Dd, which could open up additional space for alternative conformations of a peptide terminus. These include Thr7 (Tyr7 in H-2Dd) and Cys171 (Tyr171 in H-2Dd), as well as Arg167 (Trp167 in H-2Dd) if its sidechain can adopt an alternate conformation. We conclude that the M10.5 groove can accommodate a peptide that adopts a class I MHC–binding conformation, but that differences between the A and F pocket regions of M10.5 and classical class I MHC molecules (see Figure 2B) would require a peptide bound in the M10.5 groove to be anchored differently than a class I MHC–binding peptide.

We next considered whether longer peptides or an extended region of a protein could fit into the M10.5 groove by extending the polyanalanine version of the H-2Dd–binding peptide by four alanine residues at its C-terminus. The resulting peptide was fit into the M10.5 groove with the extra residues extending out of the F pocket side of the groove (Figure 2C). The relatively open F pocket end of the M10.5 groove is able to accommodate the extra residues, suggesting that M10 proteins could bind extended regions of other proteins. Candidate M10–binding proteins include the V2Rs and/or the major urinary proteins (MUPs), which are thought to deliver small-molecule compounds to chemosensory receptor neurons [1,12]. Although the nature of the interaction is unknown, biochemical data demonstrate an association between M10 and V2R proteins [8], thus the open M10.5

### Table 3. Amino Acid Yields from Acid Eluates

<table>
<thead>
<tr>
<th>Cycle Number</th>
<th>UL18</th>
<th>FcRn</th>
<th>M10.5 (CHO)</th>
<th>M10.5 (Insect)</th>
<th>M10.5 – Peptide (Insect)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>15.22</td>
<td>14.69</td>
<td>1.75</td>
<td>4.93</td>
<td>8.10</td>
</tr>
<tr>
<td>2</td>
<td>11.74</td>
<td>3.23</td>
<td>0.60</td>
<td>1.07</td>
<td>2.65</td>
</tr>
<tr>
<td>3</td>
<td>8.89</td>
<td>3.19</td>
<td>0.39</td>
<td>1.15</td>
<td>0.41</td>
</tr>
<tr>
<td>4</td>
<td>4.99</td>
<td>4.68</td>
<td>0.35</td>
<td>0.23</td>
<td>0.74</td>
</tr>
<tr>
<td>5</td>
<td>4.55</td>
<td>2.01</td>
<td>0.66</td>
<td>1.49</td>
<td>1.14</td>
</tr>
<tr>
<td>6</td>
<td>4.52</td>
<td>2.30</td>
<td>0.54</td>
<td>0.51</td>
<td>0.31</td>
</tr>
<tr>
<td>7</td>
<td>2.46</td>
<td>1.85</td>
<td>0.50</td>
<td>0.79</td>
<td>0.95</td>
</tr>
<tr>
<td>8</td>
<td>3.35</td>
<td>0.92</td>
<td>0.40</td>
<td>0.23</td>
<td>0.55</td>
</tr>
<tr>
<td>9</td>
<td>1.95</td>
<td>1.31</td>
<td>0.38</td>
<td>0.65</td>
<td>0.55</td>
</tr>
<tr>
<td>10</td>
<td>1.20</td>
<td>0.00</td>
<td>0.33</td>
<td>0.55</td>
<td>0.84</td>
</tr>
<tr>
<td>11</td>
<td>1.21</td>
<td>1.57</td>
<td>0.49</td>
<td>0.38</td>
<td>1.01</td>
</tr>
</tbody>
</table>

Total yield (pmols) of amino acids for each Edman degradation sequencing cycle for acid eluates. Only cycles that exhibited an increase in signal from the previous cycle were included in the total yield and the amino acids threonine and aspartate were removed due to a large signal from sample contaminants. Acid eluates were derived from equivalent amounts of UL18, FcRn, CHO cell–derived M10.5, insect cell–derived M10.5, and insect cell–derived M10.5 to which a 20-fold molar excess of the H-2Kd–binding peptide SYIPSAEKI [73,74] was added. Peptide–incubated M10.5 was separated from unbound peptide by gel-filtration chromatography prior to the acid elution procedure. The peptide incubation procedure was repeated using mixtures of six or 13 MHC–binding peptides (see Materials and Methods) with similar results (i.e., no detectable peptides), as assayed by N-terminal sequencing (for the experiment involving six peptides) or mass spectrometry (for the experiment involving 13 peptides) (data not shown). MALDI-TOF analysis did not indicate the presence of N-terminally blocked peptides or other small molecules in the M10.5 or FcRn samples (data not shown).

DOI: 10.1371/journal.pbio.0020257.t003
groove may form a binding site for a piece of a V2R. MUPs, however, are not known to associate with M10 proteins, and a surface plasmon resonance-based binding experiment did not reveal an interaction between immobilized M10.5 and MUPs in fractionated mouse urine (data not shown; see Materials and Methods).

It is also possible that pheromones bind within the M10 groove and either are presented directly to V2R receptors or initiate a conformational change within the M10 molecule that results in the activation of the receptor. Because there are no known V2R–activating ligands, we examined pheromones known to activate V1R receptors including 2-heptanone, α- or β-farnesene, and 2,5-dimethylpyrazine [53]. Although these molecules fit easily within the M10.5 groove (data not shown), the hydrophobic nature of most pheromones is not complementary to the charged character of the groove (see Figure 2A), which is much larger than a single pheromone molecule. Smaller pheromones might associate with residues within the relatively hydrophobic M10.5 D pocket (Figure 2A), however, the rest of the groove would remain unoccupied and presumably still unstable, and it is not obvious how such an interaction would lead to activation of a V2R if it occurs through a conformational change in an M10 groove.

**Discussion**

The M10 family of murine class Ib MHC molecules are expressed exclusively in the VNO and appear to act as chaperones to facilitate cell surface expression of the V2R class of pheromone receptors [1,8]. Although the role of M10 molecules in V2R signaling is unclear, direct homologs of proteins typically associated with immunogenic identity such as MHC proteins provide attractive candidates to mediate MHC–disassortative mating preferences [17] and pregnancy-block phenomena [18]. Here we present a biochemical and structural analysis of M10.5, a representative M10 molecule, aimed at providing new insights into the function of M10 proteins and their association with V2Rs. As M10 proteins are related by greater than 60% sequence identity [8,12], above the approximately 30% sequence identity threshold suggesting similar three-dimensional structures [54], the results obtained for M10.5 are relevant to other M10 proteins.

Proteins with an MHC fold generally have open, occupied grooves, as in classical MHC proteins, other class Ib proteins, CD1 [30], Zn–α2-glycoprotein [28], and the protein C receptor [55], or closed, unoccupied grooves, as in FcRn [32], HFE [33], MIC-A [34], Rae-1β [56], and T22 [57]. Thus it is surprising that the M10.5 groove is open, but apparently unoccupied, in the crystal structure. Consistent with the lack of defined extra electron density for a groove occupant, analyses of acid eluates derived from M10.5 expressed in insect cells and in CHO cells failed to reveal peptidic or non-peptidic material (Table 3). In addition, M10.5 did not bind MHC–binding peptides, including peptides reported to activate VNO neurons expressing M10 family members [19]. These results do not rule out a peptide-binding role for M10 proteins because the peptides used for these experiments may not have been optimal for binding to M10.5. However, the fact that M10.5 does not associate with any endogenous peptides when expressed in CHO cells suggests that it either has more stringent criteria for peptide binding than conventional class I proteins, which bind endogenous peptides when expressed in CHO cells [21], or that it does not bind peptides at all. If M10.5 does bind peptides, changes in the M10 A and F pockets from their counterparts in class I MHC grooves (see Figure 2B) suggest that M10.5 and other M10 molecules cannot bind the same sort of peptides that are bound by class I MHC molecules. Thus it is unlikely that the class I MHC–binding peptides reported to activate VNO neurons [19] exert their effects by binding to M10 family members, consistent with the observation that purified M10.5 showed no detectable binding to peptides used in that study.

Although the ectodomains of classical class I MHC molecules can fold in the absence of peptide [21,58], and full-length empty class I proteins can reach the cell surface, they are thermally unstable and are rapidly degraded unless an appropriate binding peptide is added exogenously or the cells are grown at 26 °C [45]. Heat-induced unfolding of recombinant M10.5 reveals a thermal stability similar to empty, rather than peptide-filled, class I MHC molecules (Figure 5A and 5B). It has been assumed that the low thermal stability of empty class I MHC molecules contributes to the apparent inability of empty class I molecules to crystallize (unpublished data). Indeed, the results of molecular dynamics simulations have been used to predict that empty α1–α2 platforms do not adopt a single defined conformation [59]. However, in the case of the empty M10.5 α1–α2 platform, we are able to generate moderately well-ordered crystals, and we see a single conformation for the five copies of M10.5 in the crystallographic asymmetric unit (Figure S1). The empty platforms may be stabilized somewhat by crystal contacts, but different crystal contacts for the five M10.5 molecules do not produce different conformations of the α1–α2 platform, as evidenced by relatively low temperature factors for α1–α2 residues (see Materials and Methods).

The low thermal stability of empty recombinant M10.5 suggests that the groove is normally occupied to stabilize the protein at 37 °C. Although our results cannot be used to identify a physiologically relevant groove occupant, the structural and biochemical results can be used to determine which types of ligands are unlikely. Our analysis of the M10.5 structure and the fact that endogenous peptides were not found associated with M10.5 expressed in CHO cells (Table 3) do not support a model in which M10 proteins are stabilized by class I MHC–binding peptides from either endogenous or exogenous sources. In addition, the M10.5 groove does not have the largely hydrophobic character that would be expected if it were a binding site for hydrophobic pheromones serving as chemical cues in urine, and we observed no detectable binding signals when fractionated mouse urine was injected over purified M10.5 in a surface plasmon resonance-based binding assay (data not shown, see Materials and Methods).

A clue as to the possible nature of an M10 groove occupant comes from the different patterns of variability within the grooves of M10 and classical class I MHC proteins (Figure 4), which likely reflect the different functional roles of vomeronasal versus classical MHC molecules. Variability within the grooves of classical class I MHC molecules, which mainly involves inward-pointing groove residues [43], creates different peptide-binding preferences such that different alleles present different types of peptides that conform to allelic-specific peptide-binding motifs [35]. If the grooves of M10
proteins are occupied, the relative conservation of residues that point toward the M10 groove suggests that different M10 proteins bind a more limited set of ligands than the ligands of different class I MHC alleles. The greater variability in the upward-pointing residues on the M10 helices (Figure 4A) suggests allele-specific interactions with other proteins, consistent with the suggestion of specificity in binding between different V2R and M10 proteins [8]. The variability analysis combined with the M10.5 structure suggests a potential interaction site: The disordered six-residue loop within the M10.5 α2 domain (residues 145–150) contains three highly variable residues that could be involved in an interaction with a V2R. As a precedent for a disordered region of an MHC-like structure being at a receptor binding site, a portion of the α2 domain helix that is disordered in the structure of MIC-A (residues 152 to 161, corresponding to approximately the same M10.5 residues) [34] becomes ordered in a structure of MIC-A bound to the NKGD2 receptor [60].

The disordered region of the M10.5 α2 domain helix may also be related to a need for flexibility within the F pocket region of the M10 groove, which could allow binding of larger ligands than the 8- to 10-mer peptides bound in the grooves of classical class I MHC molecules. Although the M10.5 groove is about the same size as the grooves of classical class I MHC molecules (approximately 730 Å² versus approximately 760 Å²), and the A pocket side of both types of grooves is closed, the F pocket side of the M10.5 groove is more open than the counterpart region of a class I molecule (see Figure 2). The closed ends of class I MHC grooves result in the preference for binding short (8- to 10-mer) peptides that do not extend out of either end of the groove [35]. The open F pocket end of the M10.5 groove may allow it to bind larger ligands, perhaps even an extended region of an intact protein, such as a V2R.

As a precedent for this type of interaction, the open grooves of class II MHC molecules bind to an extended region with the invariant chain protein during transit through the ER and Golgi (reviewed in [61]). In this case, the invariant chain not only occludes the class II MHC groove to prevent association with ER peptides, it also serves as a chaperone to direct class II proteins to acidic compartments where they acquire peptides derived from exogenous proteins. A related situation could occur for M10 proteins, such that an extended region within the ectodomain of a V2R would bind within the groove of an M10 protein during transit through the ER and Golgi. An interaction in which an extended loop from a V2R protein binds to an M10 groove would explain why peptides or other small molecule ligands were not found in recombinant M10.5 molecules expressed in the absence of V2R proteins and why the empty recombinant molecules are thermally unstable. In this hypothesized scenario, newly synthesized M10 and V2R proteins would be stabilized through mutual interactions with a V2R loop in the M10 groove, enabling the M10 to escort the V2R to the cell surface, rationalizing the observation that M10 proteins are required for cell surface expression of V2Rs [8].

Materials and Methods

M10.5 expression and purification. A construct encoding soluble M10.5 (corresponding to the ectodomain with the preceding hydrophobic leader sequence and an upstream insect Kozak sequence (CTTATAAT) plus a C-terminal Factor Xa site and a 6xHis tag) was subcloned into the BamHI site of the dicistronic baculovirus transfer vector pAcUW31 (BD Biosciences Clontech, Mountain View, California, United States). The ectodomain was truncated after residue Gly299 (Gly275 using numbering corresponding to mature class I MHC heavy chains). cDNA encoding human β2m plus its hydrophobic leader sequence was inserted into the BglII site of the transfer vector. Recombinant baculovirus was generated by infection of the transfer vector with linearized viral DNA (Baculogold; BD Biosciences Clontech) and supernatants were harvested from baculovirus-infected Tr5 (High Five) insect cells. Cell supernatants were bulk-exchanged into 20 mM Tris (pH 7.4)/150 mM NaCl, and concentrated to 1 liter using an Amicon R20000 tangential-flow concentrator (Amicon, Beverly, Massachusetts, United States) with a 10 kDa cutoff membrane (Pall Corporation, East Hills, New York, United States). The resulting solution was adjusted to 50 mM Tris (pH 7.4)/300 mM NaCl/10 mM imidazole/10% glycerol, and loaded onto an Ni-NTA column (Qiagen, Valencia, California, United States) equilibrated in 50 mM Tris (pH 7.4)/300 mM NaCl/10% glycerol. The column was washed with a similar buffer containing 40 mM imidazole and the bound M10.5 eluted in 250 mM imidazole. M10.5-containing fractions were pooled, concentrated to 2 ml, and loaded onto a 16/60 Superdex 75 column (Amersham Biosciences Corp., Piscataway, New Jersey, United States) equilibrated in 20 mM Tris (pH 7.4)/150 mM NaCl/1 mM EDTA/1 mM β-mercaptoethanol. SDS-PAGE analysis of the major peak on the gel filtration column revealed two bands, one at apparent molecular masses of 32.6 kDa (protein) and 36.2 kDa (protein plus carbohydrate), corresponding to the M10.5 ectodomain, and 12 kDa (expected mass 11.7 kDa), corresponding to β2m. The M10.5-β2m peak eluted at a position corresponding to a protein of approximately 47 kDa, suggesting that the protein is monomeric (i.e., a single heterodimer). Fractions containing M10.5 were pooled resulting in a yield of approximately 1.5 mg per liter of insect cell supernatant.

N-terminal sequencing of the M10.5 heavy chain was accomplished by first blotting gel-run protein to a PVDF membrane (polyvinylidene fluoride, Millipore, Billerica, Massachusetts, United States), followed by sequencing using a 492 ALC Procise protein micro-sequencer (Applied Biosystems, Foster City, California, United States). The sequence obtained, SHWLKT, corresponds to the processed mature form of M10.5. In our numbering system, the first amino acid of the M10.5 heavy chain is denoted as residue 2 to correspond with the numbering of the processed forms of classical class I MHC molecules.

A vector for expression of M10.5 in CHO cells was constructed by subcloning the analogous region of M10.5 into a derivative of pB5-GS, which carries the glutamine synthetase gene as a selectable marker and means of gene amplification in the presence of methionine sulfonamide [62]. The M10.5 expression vector was transfected into CHO cells together with a human β2m expression vector as described [33]. Selection, amplification, maintenance of methionine sulfonamide-resistant cells, and identification of M10.5-expressing cells were described [46]. CHO cells were transfected with the M10.5 expression vector, and the bound M10.5 eluted in 250 mM imidazole. M10.5-containing fractions were pooled, concentrated to 2 ml, and loaded onto a 16/60 Superdex 75 column (Amersham Biosciences Corp., Piscataway, New Jersey, United States) equilibrated in 50 mM Tris (pH 7.4)/300 mM NaCl/10% glycerol. The bound M10.5 eluted in 250 mM imidazole. M10.5-containing fractions were pooled, concentrated to 2 ml, and loaded onto a 16/60 Superdex 75 column (Amersham Biosciences Corp., Piscataway, New Jersey, United States) equilibrated in 50 mM Tris (pH 7.4)/300 mM NaCl/10% glycerol. The bound M10.5 eluted in 250 mM imidazole. M10.5-containing fractions were pooled and concentrated to 1 liter using an Amicon R20000 tangential-flow concentrator (Amicon, Beverly, Massachusetts, United States) with a 10 kDa cutoff membrane (Pall Corporation, East Hills, New York, United States). The resulting solution was adjusted to 50 mM Tris (pH 7.4)/300 mM NaCl/10 mM imidazole/10% glycerol, and loaded onto an Ni-NTA column (Qiagen, Valencia, California, United States) equilibrated in 50 mM Tris (pH 7.4)/300 mM NaCl/10% glycerol. The column was washed with a similar buffer containing 40 mM imidazole and the bound M10.5 eluted in 250 mM imidazole. M10.5-containing fractions were pooled, concentrated to 2 ml, and loaded onto a 16/60 Superdex 75 column (Amersham Biosciences Corp., Piscataway, New Jersey, United States) equilibrated in 20 mM Tris (pH 7.4)/150 mM NaCl/1 mM EDTA/1 mM β-mercaptoethanol. SDS-PAGE analysis of the major peak on the gel filtration column revealed two bands, one at apparent molecular masses of 32.6 kDa (protein) and 36.2 kDa (protein plus carbohydrate), corresponding to the M10.5 ectodomain, and 12 kDa (expected mass 11.7 kDa), corresponding to β2m. The M10.5-β2m peak eluted at a position corresponding to a protein of approximately 47 kDa, suggesting that the protein is monomeric (i.e., a single heterodimer). Fractions containing M10.5 were pooled resulting in a yield of approximately 1.5 mg per liter of insect cell supernatant.

Crystallization and data collection. Crystals (space group P2₁2₁2₁; a = 124.11 Å, b = 124.71 Å, c = 149.37 Å; five molecules per asymmetric unit) were grown at room temperature utilizing the hanging drop method by combining 1 μl of protein solution (approximately 8 mg/ml of insect cell-derived M10.5) with 1 μl of precipitant containing 0.1 M imidazole (pH 8.0)/20% PEG 1000/0.2 M calcium acetate. Crystals were cryopreserved in liquid nitrogen after soaking in mother liquor containing 10% glycerol and 10% PEG 1000. Data were collected at 100 K at beamline 12.3.1 at the Advanced Light Source (Berkeley, California, United States) and beamline 9–2 at the Stanford Synchrotron Radiation Laboratory (Stanford, California, United States). The data were indexed and scaled using the HKL suite of programs (HKL, Research, Charlottesville, Virginia, United States) (see Table 1). The statistics reported in Table 1 refer to a single merged native dataset obtained by including frames from both native datasets in the scaling procedure.

Structure solution and refinement. A protein–protein BLAST search indicated that Qa-2 and H-2Dd are the closest-related proteins in sequence to M10.5 for which crystal structures are available. Molecular replacement was carried out using the CCP4 program AMORE [63,64] with an all atom version of Qa-2 (not including the bound peptide) as a search model. Molecular replacement was not successful using each dataset individually, but a solution was obtained when the data were merged into a single native dataset (see Table 1). Five molecules were located in the asymmetric unit giving an R-factor of 47.9% after rigid-body refinement.
The model was refined using all reflections to 3.0 Å, but as the data are only approximately 60% complete between 3.1 Å and 3.0 Å, the effective resolution of the structure is approximately 3.2 Å. Test set reflections (5% of total) were picked using the thin shell method in DATAMAN [65] to reduce the influence of NCS correlations on R-factor calculations. To reduce the possibility of a model bias, initial averaging of the model was performed using Qa-2 (two 60°-rotated, annealed, and truncated electron density maps in which 5% of the molecule was removed at a time. Simulated annealing and grouped B-factor refinement with 5-fold NCS constraints was carried out using CNS [67]. Once the Rfree value stopped improving with successive cycles of refinement, the NCS constraints were relaxed to NCS restraints using a weight of 300 kcal/mol.

Annealing and grouped B-factor refinement with 5-fold NCS maps in which 5% of the molecule was omitted at a time. Simulated annealing and grouped B-factor refinement with 5-fold NCS constraints was carried out using CNS [67]. Once the Rfree value stopped improving with successive cycles of refinement, the NCS constraints were relaxed to NCS restraints using a weight of 300 kcal/mol.

The NCS model stereochemistry was checked after each round of refinement using PROCHECK [68] and WHAT_CHECK [69]. The model (R_{free} = 26.6%; R_{free} = 30.5%) includes 264 out of 274 residues in the M10.5 heavy chain and all 99 residues of β2m (Table 1). Residues 145–150 and 195–198 are not seen in the electron-density map, and 36 sidechains in the molecule 1 M10.5 heavy chain and 13 sidechains in β2m are disordered and were modeled as alanines. The average main chain temperature factor per domain for molecule 1 are 48.9 Å² (a2–26), 68.2 Å² (β2), and 44.1 Å² (β2m). The temperature factors are the highest in the distal half of the β2 domain, correlating with this region having less continuous electron density than other regions of the maps.

Greater surface area calculations were performed as previously described [30,54,55,58,68,71,72,128,161,182,202,203]. Alignments of the M10.5 a1–2 domains with other a1–2 domains were carried out using the CCP4 program Lsoqab [63] using platform b-sheet residues 3–13, 21–28, 34–37, 46–47, 93–103, 111–118, 122–126, and 133–135. Figures were produced using Molscript [70], Raster3D [71], and PyMOL [72].

Thermal stability analyses. An AVIV 62A DS CD spectrometer with a thermoelectric cell holder and a cuvette with a 1-mm path length was used to monitor heat-induced unfolding of insect cell-derived M10.5 using samples containing 15 μM protein in 50 mM phosphate buffer. In one experiment, the protein solution also contained 20% polyethylene glycol 1000. The CD signal was monitored at 223 nm in one experiment, the protein solution also contained 20% polyethylene glycol 1000. The CD signal was monitored at 223 nm in one experiment, the protein solution also contained 20% polyethylene glycol 1000. The CD signal was monitored at 223 nm.
Crystal Structure of the MHC Ib Molecule M10.5

NK2D receptor (P26718), Qa-2 (P79567), Rae-1β (O08693), T22 (Q9BCZ1), TAP1 (P62427), TAP2 (P36371), UL18 (P08560), and Zn-21.

The Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank accession number (http://www.rcsb.org/pdb/) for the M10.5 structure is 1Z88. The accession numbers for the other structures discussed in this paper are as follows: FcRn (3FRU), H-2Dβ (1BII), and Qa-2 (1K8D).

Acknowledgments

Protein sequencing analyses carried out by the Protein/Peptide Microanalytical Lab were supported by the Beckman Institute at the California Institute of Technology (Caltech). We thank Kyle Lassila and Heidi Privett for help with CD measurements, Tony Giannetti and Anthony West for help with Biacore experiments, Peter Snow, Inderjit Nangiana, and Cynthia Jones at the Caltech Protein Expression Center for assisting with insect cell expression, Fabio Papes for providing fractionated urinary samples, the Ralph M.Parsons Foundation for computational support, and members of the Bjorkman laboratory for critical reading of the manuscript. This work was supported by a Rosalind Alcott Post-doctoral Fellowship administered by Caltech (R.O.), the Howard Hughes Medical Institute (C.D. and P.J.B.), and the NIH/NIHDCD (R01 DC003903 (C.D.)).

The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, Materials Sciences Division, of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098 at Lawrence Berkeley National Laboratory. Portions of this research were carried out at the Stanford Synchrotron Radiation Laboratory, a national user facility operated by Stanford University on behalf of the U.S. Department of Energy, Office of Basic Energy Sciences.

Competing interests. The authors have declared that no competing interests exist.

Author contributions. RAO and PB conceived and designed the experiments. RAO and KEHT performed the experiments. RAO, KEHT, CD, and PBJ analyzed the data. CD and PBJ contributed reagents/materials/analysis tools. RAO, CD, and PBJ wrote the paper.

References


