Structural Immunology of MHC Class I Proteins, Homologs and Receptor Complexes

Roland K. Strong

Div. of Basic Sciences, Fred Hutchinson Cancer Research Ctr., Seattle, WA 98109, USA

The first structure of a major histocompatibility complex (MHC) class Ia protein, HLA-A2, had a significant impact on the immunological community and answered many of the fundamental questions of how peptide antigens are presented to the immune system (1). Many of the details of the recognition process have since been elucidated, in part, by a plethora of subsequent structures of human and murine class I protein/peptide and class I protein/peptide/αβ T cell receptor (TCR) complexes (2). In structural determinations and sequence analyses, the MHC class I fold has been found in proteins involved in a variety of biological processes, playing peripheral roles in immune responses or even fulfilling functions uninvolved in immunology, interacting with a range of different small-molecule and macromolecular ligands (Table 1). Recently determined structures of the murine MHC class Ia protein H-2D^{d} in complex with its receptor Ly49A (3), the murine MHC class Ib protein H-2T22^{b} (4), the unusual MHC class I homolog zinc-α_2-glycoprotein (ZAG) (5), the human natural killer (NK) cell and T cell target MIC-A (6) and the complex between HFE and transferrin receptor (7) reveal the structural and functional extremes that this fold family encompasses (Fig. 1).

Roland K. Strong is Associate Member of the Div. of Basic Sciences, Fred Hutchinson Cancer Research Center. He worked as a Graduate student in the labs of Stephen C. Harrison, Harvard Univ. and of Gregory A. Petsko, MIT & Brandeis Univ. As Consultant in Payload Inc., Cambridge, he participated in efforts to assess the effect of microgravity on the crystallization of biological macromolecules over the course of three missions to the Russian MIR space station. He reached his PhD in the lab of Pamela J. Bjorkman, Caltech and was Affiliate Assistant Professor of the Dept. of Immunology, Univ. of Washington. **Research interests:** Structural immunology - the analysis of proteins mediating mucosal immune responses through molecular biology, solution biochemistry and x-ray crystallography. **http://fhcr.org/labs/strong/** **rstrong@fred.fhcr.org**

● *MHC class I proteins: fold and function*

The human "classical" or class Ia MHC proteins (HLA-A, B and C) are cell surface, heterodimeric glycoproteins consisting of an integral-membrane heavy chain and a soluble light chain, known as β_2-microglobulin (β_2-m) (8). During infection, proteins from pathogens are processed into peptides and packaged into MHC class Ia proteins for presentation on the cell surface; peptides from self proteins are presented in the absence of an active infection. MHC class I/peptide complexes are recognized by circulating cytotoxic T lymphocytes (CTLs) through direct interactions with antigen-specific αβ TCRs and the co-receptor CD8, resulting in the elimination of infected cells from the body (Fig. 2). The heavy chain of these proteins comprises three extracellular domains (α1, α2 and α3), a transmembrane-spanning domain and a small cytoplasmic domain. The a1 and a2 domains together comprise the peptide- and TCR-binding "platform" domain (Fig. 3). The distinctive topology of the platform consists of two long α-helices overlying an eight-stranded anti-parallel β-sheet. The two helices form the walls of the peptide-binding groove. This groove encompasses a series of pockets for the N- and C-termini of the peptide and the side chains of the "anchor" residues, thus determining the peptide specificity of a particular allele; exposed peptide side-chains and specific peptide main-chain and peptide-induced MHC protein conformations are read out by the TCR (Fig. 2 & 4). β_2-m and the heavy chain α3 domain fold into constant-type immunoglobulin domains. Association with β_2-m and peptide is required for proper folding and cell-surface expression, although either or both of these requirements are

Table 1: Funcional roles of MHC class I proteins and homologs

Molecule	Ligand	Receptor	Function
MHC Class Ia (HLA-A, B & C)	Endogenous Peptides	αβ TCRs / CD8	Immune Surveillance
MHC Class Ib (HLA-E)	Class Ia Leader Peptides	NKG2 / CD94	Check for MHC Class Ia Expression
MHC Class Ib (H-2M3)	N-formylated Peptides	(on CD8^{+} T cells)	Immune Surveillance
CD1a	Bacterial Cell Wall Components	αβ / γδ TCRs	Immune Surveillance
H-2T22^{b}	None?	γδ TCRs	Innate Immunity
MIC-A	None?	NKG2D / ?	Innate Immunity
FcRn	IgG	Not Applicable	trans-Intestinal IgG Transport
HFE	Transferrin Receptor	Not Applicable	Iron Metabolism
ZAG	Hydrophobic and Non-peptidic?	?	Fat Metabolism

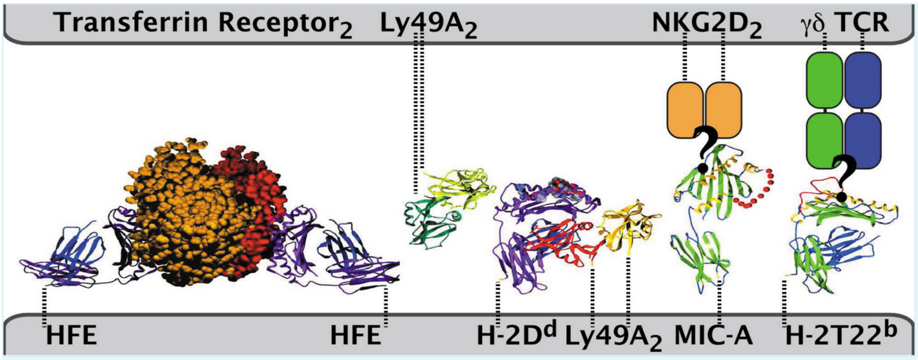

Fig. 1: Representation of several cell-surface interactions mediated by MHC class I homologs. The crystal structures of the complex between dimeric transferrin receptor and HFE, the complex between two Ly49A dimers and H-2Dd, MIC-A and H-2T22b are shown to scale. The MHC class I proteins and Ly49A are shown as backbone ribbons with secondary structure elements indicated; transferrin receptor (red and orange) and the bound peptide (colored by atom type) in H-2Dd are shown in CPK representations. β2-microglobulin is colored blue. Presumed connections to transmembrane anchors are shown as dashed lines. The flexible loops in H-2T22b (red), and the disordered loop in MIC-A (red spheres) are shown. Figures were generated with SwissPDBViewer and rendered with POV-RAY3.

dispensable in some members of the class I family.

Together with the "non-classical" or class Ib MHC proteins (HLA-E, F and G in humans) and their close, but distinct, structural relatives, the MHC class II proteins (DR, DP and DQ in humans), these proteins fill essential functions in the immune system by presenting various types of peptides to various immune system cells in various compartments. In another variation on this theme, the non-MHC encoded CD1 family of class I homologs present not peptides, but mycobacterial lipids to T cells (9,10). While recognizably class I-like, the structure of the binding pocket is altered to accommodate this type of ligand (Fig. 4) (10).

Functional diversity in MHC class I homologs

An early hint of things to come with the description of an MHC class I-like Fc receptor (FcRn) which transports IgG across the intestinal epithelium of nursing neonates (11). The structure of the FcRn/Fc complex showed a non-peptide binding MHC class I homolog using a different site to interact with ligand, a surface distinct from either the TCR or CD8 binding sites (12): αβ TCRs bind to the "top" of a class I molecule, covering much of the bound peptide; CD8 primarily binds the α3 domain but also contacts α2 and β_2-m (Fig. 2). The TCR sits diagonally on the "top" of the platform making extensive contacts with the peptide, and with the Vα domain overlying the α2 domain helix and the Vβ domain contacting the α1:H2 helix (Fig. 3). Fc binds to FcRn in a pH dependent manner near the ends of the α1:H2 and the α2:H1 helices, the edges of the b-strands of the α2 domain, and residues of β_2-m near the N-terminus.

MHC class I proteins and NK cell receptors

H-2D^d is a murine class Ia that is recognized by inhibitory receptors on NK cells, a component of the innate immune system that destroys cells that fail to express class Ia proteins on their surfaces. Some pathogens, such as cytomegalovirus (CMV), have developed clever strategies for down-regulating the expression of class Ia proteins, thus evading αβ T cell-mediated destruction – only to be caught by NK cells. D^d is recognized through the NK cell receptor Ly49A, a member of the growing family of proteins that contain C-type lectin domains (CTLDs) (13). Two Ly49A homodimers bind to one D^d molecule in the crystal (3). The first binding site is not particularly extensive, but overlaps the position of a conserved N-linked sugar, absent from the protein used for crystallization, suggesting that carbohydrate may contribute significantly to Ly49A binding - even though Ly49A lacks several of the features thought to mark carbohydrate-binding CTLDs, particularly two tightly-bound calcium ions. This first site represents yet another binding surface on the MHC class I fold, on the opposite side of the platform from the Fc binding site on FcRn. The second site, which

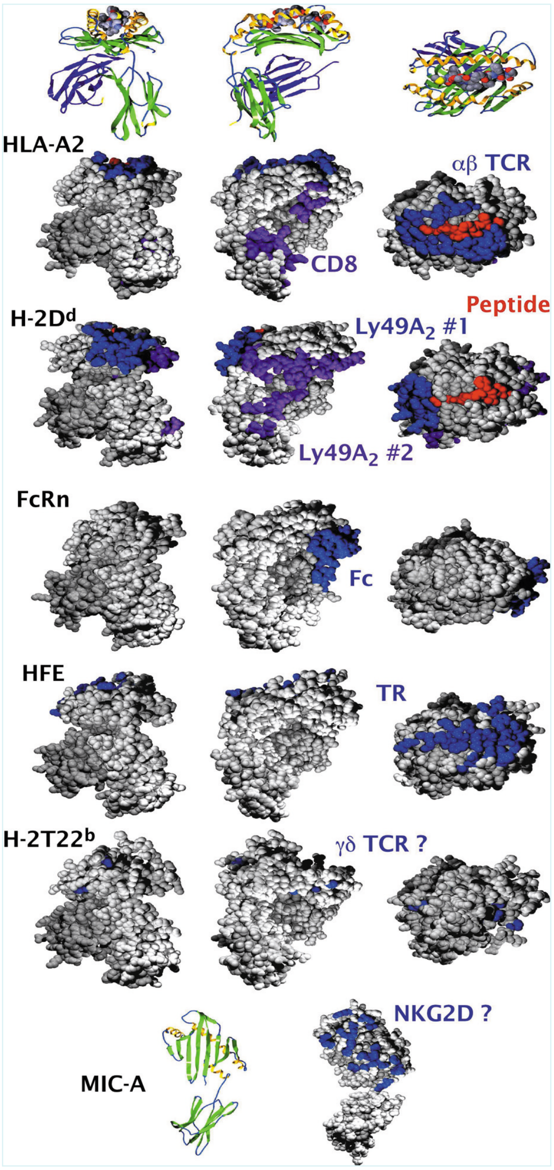

Fig. 2: Footprints of protein ligands on several MHC class I molecules. CPK representations of the structures of HLA-A2, H-2D^d, FcRn, HFE, H-2T22^b and MIC-A are shown. Bound peptides (red) and β2-microglobulin (dark gray) are shown when present. Ribbon representations of the fold of HLA-A2 are shown at the top of the figure to orient the views in each column; the fold of MIC-A is shown at the bottom of the figure in a ribbon representation to orient the view of MIC-A. Residues defining the ligand footprints (blue and purple) are labeled by molecule for HLA-A2, H-2D^d, FcRn and HFE (TfR: transferrin receptor); residues defining putative ligand binding sites in H-2T22^b and MIC-A are similarly indicated.

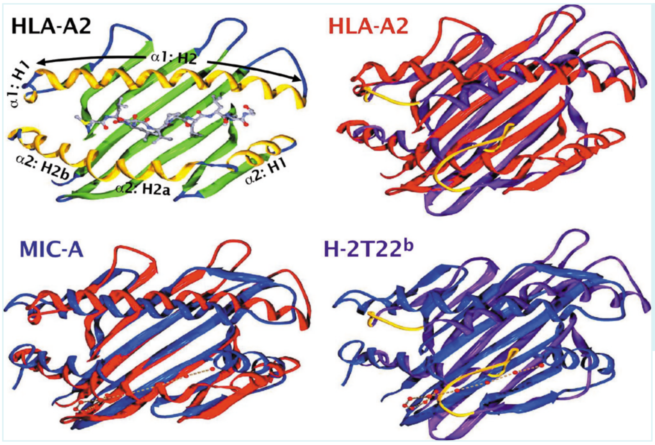

Fig. 3: Superpositions of the platform domains of HLA-A2 (red), MIC-A (blue) and H-2T22^b (purple). A ribbon representation of HLA-A2 alone is shown to orient the views; secondary structure elements are indicated and the bound peptide is shown in ball-and-stick style and colored by atom type. This view looks down onto the top of the platform domain, the surface recognized by a TCR. In this orientation, the α1 domain is to the upper left and the α2 domain is at the lower right in each panel. The flexible loops in H-2T22^b are colored yellow; the disordered loop in MIC-A is represented by spheres (red) connected by dotted lines to indicate the length of this segment.

overlaps the CD8 binding site, while more extensive and intimate, may be a *cis* interaction between D^d and Ly49A on the surface of the NK cell, possibly playing a role in positioning the receptor for target binding. Ly49A homodimers are able to crosslink D^d molecules through the interactions seen in the crystal, and do not occlude the TCR binding site, both mechanisms possibly contributing to signaling. Leading questions raised by this structure include how Ly49A and other members of the Ly49 family distinguish between different murine MHC class Ia alleles and the details of the carbohydrate/ Ly49A interaction.

MIC-A is recognized by NK cells and T cells through NKG2D, which, unlike the Ly49A/D^d signal, stimulates, rather than inhibits, cytolysis. MIC-A binds neither peptides, peptide surrogates (such as lipids) nor β_2-m. NKG2D is the most distantly related member of the human NKG2 family of CTLD-containing NK receptors, which have low sequence similarity to the Ly49 family. The MIC-A platform differs dramatically from an MHC class Ia platform, particularly in the loops at the edges of the α1 domain and the apparent disordering of the α2:H2α helix into a flexible loop that is not visible in the crystal structure (6). The platform and α3 domains are also flexibly linked. MIC-A and H-2D^d are structurally dissimilar in the regions of the Ly49A binding sites, so it is very unlikely that NKG2D binds MIC-A in a manner analogous to the D^d/Ly49A interaction. Even though a receptor complex is not yet available, mapping of allelic and species sequence conservations onto the MIC-A structure suggests that NKG2D may bind to the underside of the platform domain (Fig 2.). Approximately half of this surface corresponds to the β_2-m binding site in class I proteins that associate with β_2-m. A receptor complex structure will be necessary to sort out the details of this important interaction, and determine if and how carbohydrate is involved.

MHC class I proteins and γδ T cell receptors

The structure of the non-peptide binding, murine MHC class Ib protein, H-2T22^b, which is recognized by a subset of γδ TCRs, reveals the most distorted example of an MHC class I platform fold yet seen, even though this molecule interacts with β_2-m in the usual manner (4). The distortions are partly the result of deletions in the sequence of H-2 T22^b. Two loops, corresponding to the a1:H1 and part of the α1:H2 helices, and the α2:H1 helix, are conformationally flexible, dramatically demonstrated by comparisons of the four molecules in the asymmetric unit of the crystal. The remainder of the α1:H2 helix curves into what would be the peptide-binding groove in a previously unobserved manner. Differences between binding and non-binding alleles of T22, and the closely related T10 molecules, have been used to delineate possible receptor binding sites, one of which lies in the most distorted part of the T22 platform in the α2 domain (Fig. 2). With only part of the structure of a γδ TCR currently available (14), the structure of a γδ/H-2T22^b complex would be an important advance in studying the function of these receptors, since the radical changes seen in the structure of the H-2T22^b platform, and differences in the complementarity determining regions (CDRs) of the different classes of TCRs (15), almost guarantee that αβ and γδ TCRs recognize their ligands in distinct manners.

MHC Class I Proteins and Metabolism

Mutations in the HFE gene are responsible for the iron absorption disorder hereditary hemochromatosis

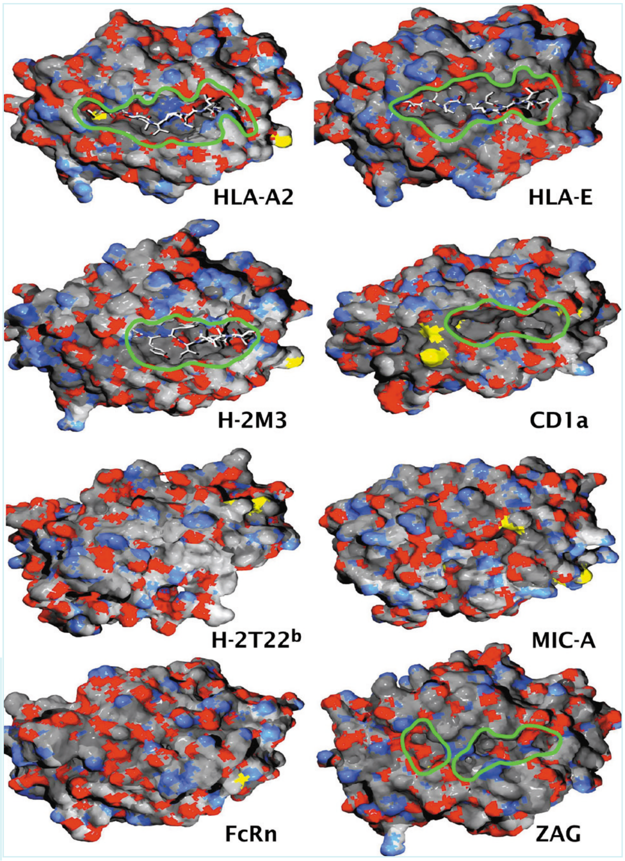

Fig. 4: Molecular surfaces of MHC class I proteins and homologs, colored by atom type (carbon: gray; oxygen: red; nitrogen: blue; sulfur: yellow). When present, bound peptides are shown in stick representation colored by atom type. The outline of the windows into well defined binding pockets are shown in green when present. The orientation of the molecules is the same as in Fig. 3.

(HH). HFE is a non-peptide binding MHC class I homolog that binds β_2-m and affects iron metabolism by binding transferrin receptor (TfR) in a pH dependent manner reminiscent of FcRn. Comparisons of the structure of TfR in the presence (7) or absence (16) of HFE reveal conformational changes that provide a possible explanation for altering the affinity for transferrin *versus* HFE, which likely bind to overlapping sites on TfR. A TfR homodimer binds to the top of two HFE platforms, one on each chain of the TfR dimer, overlapping the αβ TCR-binding surface in class Ia proteins (Fig. 2). However, the HFE-TfR interaction surface consists of a three-helix bundle, two helices contributed from TfR and one helix from the HFE α1 domain, an interaction surface very unlike the CDR loops in αβ TCRs. TfR is also positioned over the "peak" formed by the bend between the α2:H1 and α2:H2a helices, a feature that αβ TCRs avoid by their diagonal interaction with class Ia proteins. The resulting packing of helices at the interface is unlike coiled-coil interfaces seen other proteins. The complex structure also suggests that pH-mediated changes in complex formation are the result of conformational changes in TfR sensed by HFE. Interestingly, one HH mutation (His41) is not involved in the TfR interaction, suggesting that other mechanisms, possibly interactions with other molecules, may be affected in some forms of HH.

ZAG is a soluble, secreted protein present in most body fluids and has been found to accumulate in breast cysts and some breast carcinomas. ZAG has been associated with the catabolism of lipids and cachexia, though through an, as yet, uncharacterized mechanism. ZAG binds neither β_2-m nor endogenously-expressed peptides. The crystal structure of ZAG revealed a more extensive interaction between the platform and α3 domains, which has been forwarded as the likeliest explanation accounting for the stable expression and secretion of this molecule in the absence of β_2-m association (5). ZAG also retains an open binding groove corresponding to the peptide binding groove of classical class I proteins, although with a significantly altered shape. Electron density for a hydrophobic, non-peptidic small-molecule ligand was visible in the ZAG groove. It remains to be seen if either further structural or biochemical analyses can pin down the identity of this compound.

Future directions

Structures of MHC class I homologs illustrate that most of the surface of this fold can be suborned into a binding site for either a broad range of proteins, smaller molecules, or both. Rather than recapitulating a single structural theme, these recent structures demonstrate that many of the key elements of the basic class I fold are mutable to yield molecules with dramatically different functions. What surprises might be revealed by structural studies of other divergent MHC class I homologs, such as the CMV protein UL18 (17), or functionally divergent interactions, such as the interaction between class I proteins and insulin receptor (18), remain to be seen.

REFERENCES

1. Bjorkman PJ et al. Nature, 329, 506, 1987
2. Maenaka K et al. Curr Opin Struct Biol, 9, 745, 1999
3. Tormo J et al. Nature, 402, 623, 1999
4. Wingren C et al. Science, 287, 310, 2000
5. Sanchez LM et al. Science, 283, 1914, 1999
6. Li P et al. Immunity, 10, 577, 1999
7. Bennett MJ et al. Nature, 403, 46, 2000
8. Bjorkman PJ et al. Ann Rev Biochem, 90, 253 1990
9. Porcelli SA et al. Immunol Today, 19, 362, 1998
10. Zeng, Z.-H., et al., Science, 277, 339, 1997
11. Simister NE et al. Eur J Immunol, 15, 733, 1985
12. Burmeister WP et al. Nature, 372, 379, 1994
13. Drickamer K, Curr Opin Struct Biol, 9, 585, 1999
14. Li H et al. Nature, 391, 502, 1998
15. Davis MM et al. Immunol Today, 16, 316, 1995
16. Lawrence CM et al. Science, 286, 779, 1999
17. Fahnestock ML et al. Immunity, 3, 583, 1995
18. Verland S et al. J Immunol, 143, 945, 1989