

suitable for this strategy, given its approximate eight-fold symmetry that allows functional groups to be directed toward the bound substrate and reaction intermediates from virtually any direction.

One limitation to commercially viable applications of this strategy will be the development of an *in vitro* screen if an *in vivo* selection is not available, especially for reactions that do not produce a change in optical properties as substrate is transformed to product. Perhaps an even more challenging limitation will be expansion of the repertoire of binding specificities available in existing α/β -barrel enzymes. The most difficult step in *de novo* design of an enzyme is the construction of a binding site. As a corollary, structural biologists now are realizing that deducing functions (substrate specificity and reaction) of unknown proteins from their structures alone is a considerable challenge. Perhaps newly developed techniques in combinatorial protein engineering will allow the specificities of Nature's α/β -barrel enzymes to be altered to accommodate unnatural substrates so that genuinely new enzymes can be created.

Nature's strategies

Enzymes that contain α/β -barrel domains catalyze a diverse range of reactions. Much speculation has focused on whether this fold independently evolved to catalyze each of these reaction types or whether these diverged from a single or more limited

number of progenitors⁵. The facile evolution of *ivePRAI* from IGPS supports the proposal that Nature can modify the structures of 'old' α/β -barrel enzymes to generate 'new' enzymes that catalyze reactions whose mechanisms bear no relationship to the mechanism of the 'old' enzyme. Perhaps each of the enzymes in the tryptophan synthesis pathway studied here was evolved successively by Nature from tryptophan synthase, the last enzyme in the pathway. Additional evidence for the possible importance of substrate/product binding in divergent evolution is provided by phosphoribosylformimino-5-aminoimidazole carboxamide isomerase (HisA) and imidazole glycerolphosphate synthase (HisF), which catalyze successive steps in the histidine biosynthetic pathway⁶.

This evolutionary course is in striking contrast to the observation that the reactions catalyzed by many superfamilies of distantly related enzymes share at least a common partial reaction. For example, members of the enolase superfamily catalyze different overall reactions but share initial divalent metal ion-assisted abstraction of an α -proton from a carboxylate anion substrate to form an enolic intermediate⁷. The members of the enolase superfamily have an α/β -barrel domain, with active site functional groups also located in the loops that connect the β -sheets with the α -helices. However, in this superfamily, the differing substrate specificities are

determined not by the barrel domain but by a second N-terminal domain that closes over the C-terminal end of the barrel domain. Not surprisingly, Nature does not use a single strategy to exploit the functional plasticity of the α/β -barrel fold.

The specificity-based evolution of IGPS into *ivePRAI* and the mechanism-based evolution in the enolase superfamily illustrate that Nature's strategies for divergent evolution can be discovered only by comparing enzymes whose sequences are significantly diverged. These demonstrations of the functional plasticity of the α/β -barrel fold suggest that Nature long ago recognized the utility of this fold in evolving new reactions. Now the challenge is to fully exploit these strategies so that designer catalysts can be generated.

John A. Gerlt is in the Departments of Biochemistry and Chemistry, University of Illinois, 600 South Mathews Avenue, Urbana, Illinois 61801, USA. email: j-gerlt@uiuc.edu

1. Altamirano, M.M., Blackburn, J.M., Aguayo, C. & Fersht, A.R. *Nature*, **403**, 617–622 (2000).
2. Wilmanns, M., Priestle, J.P., Niermann, T. & Jansonius, J.N. *J. Mol. Biol.* **223**, 477–507 (1992).
3. Hyde, C.C., Ahmed, S.A., Padlan, E.A., Miles, E.W. & Davies, D.R. *J. Biol. Chem.* **263**, 17857–17871 (1988).
4. Horowitz, N.H. *Proc. Nat. Acad. Sci. USA* **31**, 153–157 (1945).
5. Reardon, D. & Farber, G.K. *FASEB J.* **9**, 497–503 (1995).
6. Fani, R., Mori, E., Tamburini, E. & Lazcano, A. *Orig. life Evol. Biosph.* **28**, 555–570 (1998).
7. Babbitt, P.C. et al. *Biochemistry* **35**, 16489–16501 (1996).

Class (I) will come to order – *not*

Roland K. Strong

Structures of several structurally and functionally divergent MHC class I homologs and receptor complexes have recently been determined. They reveal the unusual versatility of the underlying MHC class I fold.

Few crystal structures have had a greater impact on a field than the first structure of a major histocompatibility complex (MHC) class Ia protein, HLA-A2, had on immunology¹. The structure revealed how peptide antigens are presented to the immune system. Many of the details of the recognition process have been elucidated by a number of subsequent structures of human and murine class I protein–peptide and class I protein–peptide– $\alpha\beta$ T cell receptor (TCR) complexes². Since then, the MHC class I fold has been found

in proteins involved in a variety of biological processes, some having little to do with the immune system, interacting with a range of different ligands. Recently determined structures of the murine MHC class Ia protein H-2D^d in complex with its receptor Ly49A (ref. 3), the murine MHC class Ib protein H-2T22^b (ref. 4), the human natural killer (NK) cell and T cell target MIC-A (ref. 5), and the complex between HFE and transferrin receptor⁶ reveal the structural and functional extremes that this fold family encompasses.

These structures reveal differences in the underlying fold and show how this fold can interact with diverse ligands through binding sites that, together, cover much of the surface of the molecule.

The 'classical' MHC proteins

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; Fig. 1a,b)⁷.

news and views

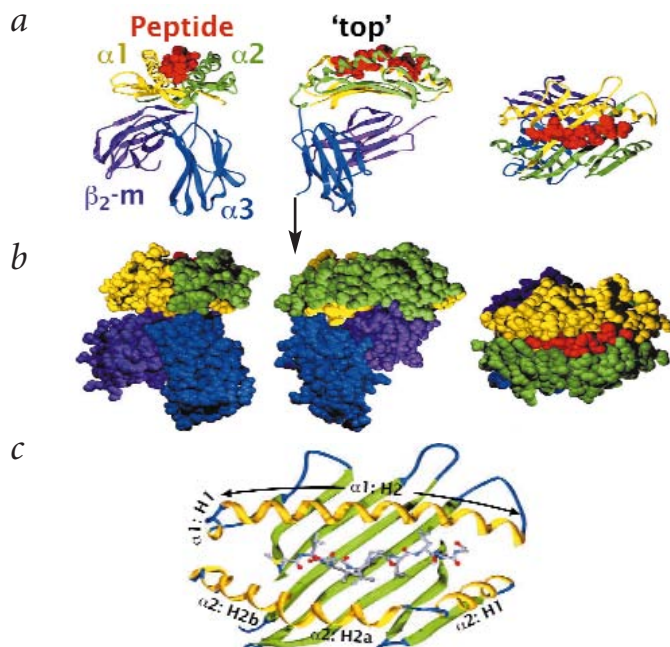


Fig. 1 The structure of a classical MHC class I protein. **a**, Ribbon and **b**, space-filling representations of three views of the structure of HLA-A2, colored to indicate the arrangement of domains ($\alpha 1$: yellow; $\alpha 2$: green; $\alpha 3$: blue; β_2 -m: purple) and the bound peptide (red). The 'top' surface of the platform domain ($\alpha 1\alpha 2$) is indicated, and an arrow shows where the C-terminus of the heavy chain would continue into the transmembrane anchor. **c**, A ribbon representation of the platform domain, colored by secondary structure, is shown. The view is from above, looking down onto the 'top' of the domain (the TCR binding site). The peptide is shown in ball-and-stick style and colored by atom type; the different helical segments are labeled. Figures were generated with SwissPDB Viewer and rendered with POV-RAY3.

During infection, proteins from pathogens are processed into peptides and packaged into MHC class Ia proteins for presentation on the cell surface. MHC class I-peptide complexes are recognized by circulating cytotoxic T lymphocytes (CTLs) through direct interactions with antigen-specific $\alpha\beta$ TCRs and the coreceptor CD8, resulting in the elimination of infected cells from the body.

The heavy chain of these MHC proteins comprises three extracellular domains ($\alpha 1$, $\alpha 2$ and $\alpha 3$), a transmembrane-spanning domain and a small cytoplasmic domain (Fig. 1). The $\alpha 1$ and $\alpha 2$ domains together comprise the peptide- and TCR-binding 'platform' domain (Fig. 1c). The distinctive topology of this domain consists of two long α -helices overlying an eight-stranded antiparallel β -sheet. The two helices form the walls of the peptide-binding groove. β_2 -m and the heavy chain $\alpha 3$ 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 dispensable in some members of the class I family. The $\alpha\beta$ TCRs bind to the 'top' of a class I molecule, covering much of the bound peptide. CD8 primarily binds the $\alpha 3$ domain but also contacts $\alpha 2$ and

β_2 -m (Fig. 2). The TCR sits diagonally on the 'top' of the platform making extensive contacts with the peptide.

Class I-like MHC proteins

An early hint that this family of MHC class I proteins could interact with different ligands in different ways came with the description of an MHC class I-like Fc

receptor (FcRn) which transports IgG across the intestinal epithelium of nursing neonates⁸. 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 (Fig. 2)⁹. Fc binds to FcRn in a pH dependent manner near the ends of the $\alpha 1$:H2 and the $\alpha 2$:H1 helices, the edges of the β -strands of the $\alpha 2$ domain, and residues of β_2 -m near the N-terminus.

Murine class Ia and NK receptors

H-2D^d is a murine class Ia protein that is recognized by inhibitory receptors on NK

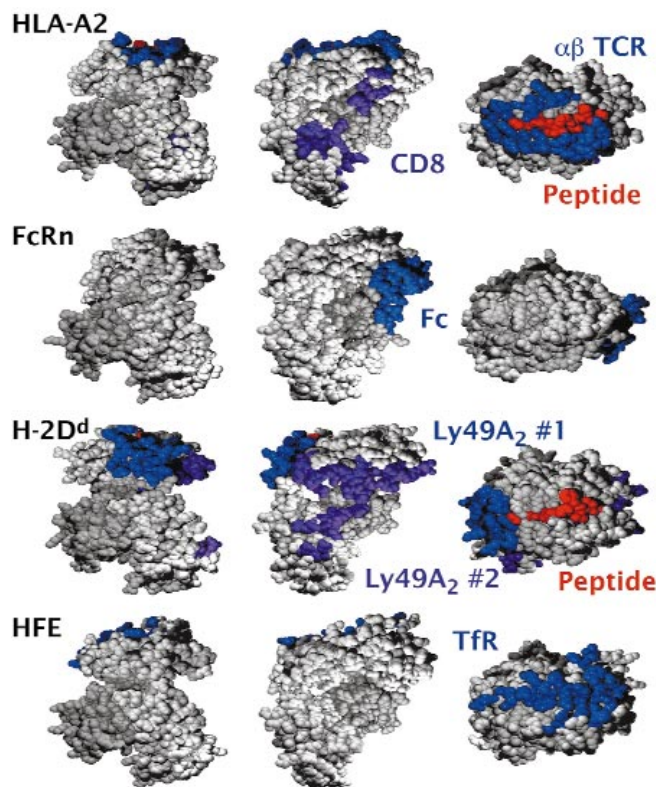


Fig. 2 Footprints of protein ligands on several MHC class I molecules. Space-filling representations of the structures of HLA-A2, FcRn, H-2D^d, and HFE are shown, oriented as in Fig. 1a,b. Bound peptides (red) are shown when present. Residues defining the binding sites (blue and purple) are labeled by ligand.

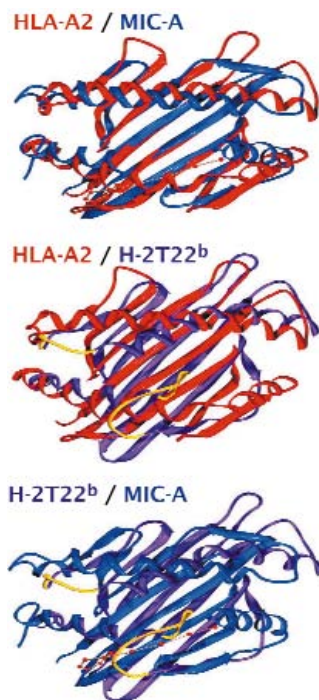


Fig. 3 Superpositions of the platform domains of HLA-A2 (red), MIC-A (blue) and H-2T22^b (purple) in the same orientation as in Fig. 1c. The flexible loops in H-2T22^b are yellow; the disordered loop in MIC-A is represented by spheres (red) connected by dotted lines to indicate the length of this segment.

cells, a component of the innate immune system that destroys cells that fail to express class Ia proteins on their surfaces. 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)¹⁰. Two Ly49A homodimers bind to one D^d molecule in the crystal (Fig. 2)³. 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 overlaps the CD8 binding site, while more extensive and intimate, may be an interaction between D^d and Ly49A on the surface of the same NK cell, possibly playing a role in positioning the receptor for binding to D^d on a target cell. The Ly49A–D^d interaction can crosslink multiple Ly49A receptors and does not occlude TCR binding.

Fig. 4 Space-filling representations of the structures of MIC-A (top) and H-2T22^b (bottom). Residues defining putative ligand binding sites in H-2T22^b and MIC-A are blue, and are labeled by ligand. H-2T22^b is oriented as in Figs. 1 and 2. MIC-A is oriented to highlight the exposed 'underside' of the platform domain.

HFE and iron metabolism

Mutations in the HFE gene are responsible for the iron absorption disorder hereditary hemochromatosis (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 presence⁶ or absence¹¹ 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 $\alpha\beta$ 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 α 1 domain, an interaction very unlike the binding of class Ia proteins to $\alpha\beta$ TCRs. 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 (His 41) is not involved in the TfR interaction, suggesting that other mechanisms, possibly interactions with other molecules, may be affected in some forms of HH.

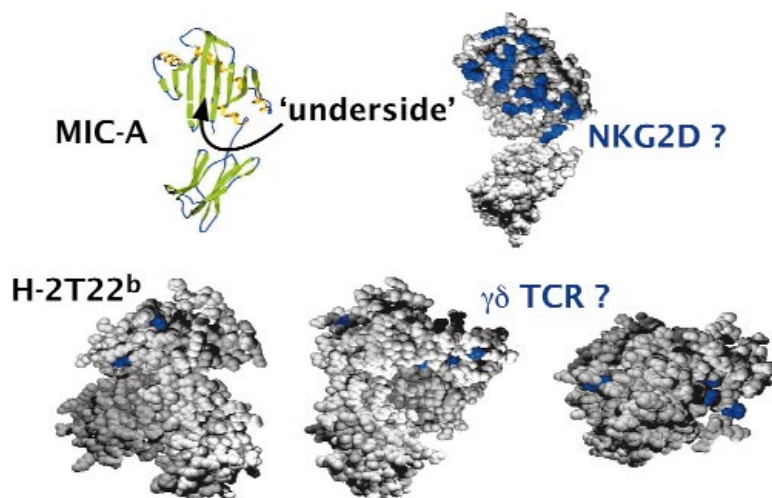
Human class Ib and NK receptors

MIC-A is recognized by NK cells and T cells through NKG2D, which, unlike the Ly49A–D^d signal, stimulates, rather than inhibits, cytotoxicity. MIC-A binds neither peptides, peptide surrogates (such as lipids) nor β_2 -m. The MIC-A platform differs dramatically from an MHC class Ia platform, particularly in the loops at the

edges of the α 1 domain and the apparent disordering of the α 2:H2 α helix into a flexible loop that is not visible in the crystal structure (Fig. 3)⁵. The platform and α 3 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 conservation onto the MIC-A structure suggests that NKG2D may bind to the 'underside' of the platform domain (Fig. 4). Approximately half of this surface corresponds to the β_2 -m binding site in class I proteins that associate with β_2 -m.

Murine class Ia and $\gamma\delta$ TCRs

The newest structure — of the non-peptide binding, murine MHC class Ib protein, H-2T22^b, which is recognized by a subset of $\gamma\delta$ TCRs—reveals the most distorted example of an MHC class I platform fold observed to date, even though this molecule interacts with β_2 -m in the usual manner (Fig. 3)⁴. The distortions are partly the result of deletions in the sequence of H-2T22^b. Two loops, corresponding to the α 1:H1 and part of the α 1:H2 helices, and the α 2:H1 helix, are conformationally flexible, dramatically demonstrated by comparisons of the four molecules in the asymmetric unit of the crystal. The remainder of the α 1: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 plat-



news and views

form in the $\alpha 2$ domain (Fig. 4). With only part of the structure of a $\gamma\delta$ TCR currently available¹², the structure of a $\gamma\delta$ -H-2T22^b complex would be an important advance in studying the function of these receptors. In fact, radical changes seen in the structure of the H-2T22^b platform, and differences in the sequences of the different classes of TCRs¹³, almost guarantee that $\alpha\beta$ and $\gamma\delta$ TCRs recognize their ligands in distinct manners.

The future

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 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 UL18¹⁴, or functionally divergent interactions, such as the interaction between class I proteins and insulin receptor¹⁵, remain to be seen. The structures of receptor-MIC-A or -H-2T22^b complexes will almost certainly add to the diversity of class I-ligand interactions. Analysis of the structures of the FcRn-Fc, H-2D^d-Ly49A and MHC protein-TCR complexes has not only detailed the intermolecular interactions, but has also provided clues for understanding the functional consequences of complex formation: ligand transport and ligand-mediated signaling events. Further structures will expand the focus of these studies from elucidating biologically important recognition events to explaining the broader functional context that these recognition events mediate.

Roland K. Strong is in the Division of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, USA. email: rstrong@fhcrc.org

1. Bjorkman, P.J. et al. *Nature* **329**, 506–512 (1987).
2. Maenaka, K. & Jones, E.Y. *Curr. Opin. Struct. Biol.* **9**, 745–750 (1999).
3. Tormo, J., Natarajan, K., Margulies, D.H. & Mariuzza, R.A. *Nature* **402**, 623–631 (1999).
4. Wingren, C., Crowley, M.P., Degano, M., Chien, Y.-H. & Wilson, I.A. *Science* **287**, 310–314 (2000).
5. Li, P. et al. *Immunity* **10**, 577–584 (1999).
6. Bennett, M.J., Lebrón, J.A. & Bjorkman, P.J. *Nature* **403**, 46–53 (2000).
7. Bjorkman, P.J. & Parham, P. *Annu. Rev. Biochem.* **90**, 253–288 (1990).
8. Simister, N.E. & Rees, A.R. *Eur. J. Immunology* **15**, 733–738 (1985).
9. Burmeister, W.P., Huber, A.H. & Bjorkman, P.J. *Nature* **372**, 379–383 (1994).
10. Drickamer, K. *Curr. Opin. Struct. Biol.* **9**, 585–590 (1999).
11. Lawrence, C.M. et al. *Science* **286**, 779–782 (1999).
12. Li, H. et al. *Nature* **391**, 502–506 (1998).
13. Davis, M.M. & Chien, Y. *Immunology Today* **16**, 316–318 (1995).
14. Fahnestock, M.L. et al. *Immunity* **3**, 583–590 (1995).
15. Verland, S. et al. *J. Immunology* **143**, 945–951 (1989).

Passing the baton in base excision repair

Samuel H. Wilson and Thomas A. Kunkel

Apurinic/apyrimidinic endonuclease 1 (APE1) plays a central role in DNA repair by cleaving the DNA backbone 5' of AP sites that result from removal of damaged bases. New structural findings on APE1-DNA cocrystals provide insights into how this enzyme binds and cleaves its substrate and how, like one member in an efficient relay team, it coordinates potentially dangerous steps in the base excision repair pathway.

Many different substances produce DNA base damage. Such an onslaught would be ultimately lethal were it not for the ability of mammalian cells to remove much of this damage through base excision repair (BER)^{1–5}. The importance of this system is illustrated by the fact that a BER deficiency leads to cellular hypersensitivity to certain toxins and to genomic instability and mutations^{6–8}. The main BER pathway in mammalian cells involves removal of a single damaged nucleotide and its replacement with an undamaged base through template directed synthesis to fill the single-nucleotide gap. BER can be initiated (Fig. 1) by any of several damage-specific DNA glycosylases^{9,16} that remove damaged bases to produce apurinic/apyrimidinic (AP) sites. AP sites, which can also arise due to spontaneous depurination, are a very common lesion in mammalian cells². AP sites are also potentially dangerous lesions because they are unstable, can

strongly impede DNA synthesis by many DNA polymerases, and lack base coding potential when replicated. AP sites are the substrates for AP endonuclease 1 (APE1), which cleaves the sugar-phosphate backbone 5' of the AP site^{10,11}. This generates the nicked intermediate for the next protein in the BER pathway, DNA polymerase β . Pol β incorporates the required undamaged nucleotide and then removes the abasic sugar-phosphate (dRP) group, generating a nicked duplex DNA product for subsequent ligation by either DNA ligase I or the DNA ligase III-XRCCI complex⁴.

Recent studies, including one by Tainer and coworkers¹² in a recent issue of *Nature*, suggest that, except for the DNA glycosylases, the steps in BER may involve recognition of a product-enzyme complex by the next enzyme in the pathway, rather than binding to an intermediate that is free in solution. Thus, these

enzymes likely coordinate with one another to receive the damaged DNA substrate and efficiently pass the resulting DNA product along to the next enzyme, just as a baton is passed from one runner to the next in a relay.

Clues to the mechanisms of product-complex recognition and coordination in BER have come from biochemical studies of protein-protein interactions and from ternary complexes between two enzymes and a DNA intermediate (Table 1). Most of the proteins involved in single-nucleotide BER are relatively small, single polypeptide molecules that are active as monomers, and this has facilitated extensive structural and biochemical characterization of these proteins⁴. Key insights into possible mechanisms that coordinate product-substrate hand-off in the BER relay are also being derived from structural studies of BER enzymes and their substrate and product complexes. The