Crystal Structures of Mouse CD1d-iGb3 Complex and its Cognate Vα14 T Cell Receptor Suggest a Model for Dual Recognition of Foreign and Self Glycolipids

Dirk M. Zajonc1, Paul B. Savage2, Albert Bendelac3, Ian A. Wilson1,4,*, and Luc Teyton5,*

1Department of Molecular Biology, The Scripps Research Institute, 10550 North Torrey Pines Rd., La Jolla, CA 92037, USA
2Department of Chemistry, Brigham Young University, Provo, UT 84602, USA
3Committee on Immunology, University of Chicago and Howard Hughes Medical Institute, Chicago, IL 60637, USA
4Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Rd., La Jolla, CA 92037, USA
5Department of Immunology The Scripps Research Institute, 10550 North Torrey Pines Rd., La Jolla, CA 92037, USA

Abstract

The semi-invariant Vα14Jα18 T cell receptor (TCR) is expressed by regulatory NKT cells and has the unique ability to recognize chemically diverse ligands presented by CD1d. The crystal structure of CD1d complexed to a natural, endogenous ligand, isoglobotrihexosylceramide (iGb3), illustrates the extent of this diversity when compared to the binding of potent, exogenous ligands, such as α-galactosylceramide (α-GalCer). A single mode of recognition for these two classes of ligands would then appear problematic for a single T cell receptor. However, the Vα14 TCR adopts two different conformations in the crystal where, in one configuration, the presence of a larger cavity between the two CDR3 regions could accommodate iGb3 and, in the other, a smaller cavity fits α-GalCer more snugly. Alternatively, the extended iGb3 headgroup could be “squashed” upon docking of the TCR and accommodated between the CD1 and TCR surfaces. Thus, the same TCR may adopt alternative modes of recognition for these foreign and self-ligands for NKT cell activation.

Keywords

iNKT cells; CD1; TCR; glycolipids; immune system

Introduction

Invariant NKT (iNKT) cells express a semi-invariant T cell receptor (TCR) composed of an invariant α chain (Vα14Jα18 in mouse, Vα24Jα18 in human) paired with a limited set of Vβ chains (Vβ8, Vβ7 and Vβ2 in mouse, Vβ11 in human). This highly restricted αβ pairing recognizes glycolipids bound to CD1d. iNKT cells have at least two functions: regulation of immune responses mediated by the autoreactive recognition of endogenous ligands; and anti-
bacterial activity through recognition of exogenous ligands. On the basis of extensive studies from several laboratories, isoglobotrihexosylceramide (iGb3) is the only glycosphingolipid that has been found to be stimulatory for both mouse and human iNKT cells. Although its synthetic route and its distribution remain uncertain, natural and synthetic forms of iGb3 are broadly stimulatory to iNKT cells, whereas its related isomer Gb3 has no stimulatory activity. Further, on the basis of its preferential usage of Vβ7>Vβ8>Vβ2 chains, iGb3 mimics closely the stimulatory properties of endogenous ligands, whereas the prototypic exogenous ligand of iNKT cells, α-galactosylceramide (α-GalCer) does not have such a bias towards Vβ7. Alternate microbial ligands of iNKT cells include α-galacturonosylceramide and α-galactosyl diacylglycerols, while mycobacterial phosphatidylinositol-tetramannosides, ganglioside GD3, and phosphatidylinositol or phosphatidylethanolamine have been suggested to stimulate minor subsets of iNKT cells, but that awaits experimental confirmation. The capacity of a single TCR to recognize multiple ligands is usually limited by very strict structural constraints that are generally concerned with recognition of highly conserved major histocompatibility complex (MHC) presenting different peptide sequences. However, some alloreactive, as well as xenogenic, pairs of ligands have illustrated the most extreme situations of TCR recognition of multiple peptide-MHC (pMHC) partners. The geometry of TCR to pMHC binding is often altered to some extent, but more substantial shifts have been seen when the divergence between selecting and exogenous ligands is more extreme, as in the case of TCR 2C bound to H-2Kb-deV8 and H-2Ld-QL9. The crystal structure of CD1d-iGb3 that we present here reveals the structural impact of such chemical differences on the presentation between this endogenous compound and the highly stimulatory α-GalCer exogenous ligand. While the headgroup of α-GalCer is a monosaccharide that is α-linked to the ceramide backbone and oriented flush with and parallel with the surface of CD1, iGb3 is a β-linked trihexosyl ceramide that requires its more complex headgroup protrudes out of the binding groove, perpendicular to the CD1 surface. A single mode of recognition of these two distinct surfaces is difficult to envisage for a single TCR. Nevertheless, the presence of a copious cavity between the two equivalent complementarity determining regions 3 (CDR3) of the α and β chains of the TCR, in one of the two forms observed in the crystal, and structural insights into α-GalCer recognition by the human Va24 TCR suggests an alternative model of iGb3 recognition, in which the more complex carbohydrate headgroup of iGb3 can either be accommodated inside this cavity or fold flat and not perturb the CD1d-TCR interactions. Both models are supported by the substantially different affinities of the two ligand classes; it is high for exogenous and low for the endogenous ligands.

Results

Structure of the CD1d-iGb3 complex

The crystal structure of the mouse CD1d-iGb3 complex was determined to 1.95 Å resolution (Table 1). The overall architecture of CD1d and its binding to a variety of different glycolipids has been described elsewhere. The lipid backbone of each antigen is generally buried deep inside the hydrophobic binding groove, which is composed of two large pockets. The A' pocket binds one of the alkyl chains, while the second alkyl chain is inserted into the F' pocket. This mode of glycolipid binding results in exposure of the polar, glycolipid headgroup to the exterior via the portal from the buried hydrophobic CD1 groove for interaction and recognition by the incoming TCR (Figure 1a).

Two independent copies of the CD1d-iGb3 complex (A and B) are present in the asymmetric unit of the crystal. However, the iGb3 ligand in complex A is less well ordered and, therefore, we refer to complex B throughout this paper for all structural aspects of iGb3 binding. The crystal structure of the CD1d-iGb3 complex (Figure 1a) revealed the high level of exposure of the linear trisaccharide headgroup (Galα1-3-Galβ1-4-Glc) of iGb3. The distal galactose (Figure
1b, in red) was disordered in the crystal structure and was, consequently, omitted from the final structure. The electron density for the ligand revealed that only the proximal glucose was well ordered, whereas the second galactose already showed a high degree of flexibility, as it does not interact extensively with CD1d (Figure 2a and b). The proximal glucose of iGb3 forms hydrogen bonds with CD1d residues of the \( \alpha_2 \)-helix, in particular with Asp153 and Thr156, and with Gly155 through a water-mediated H bond. Residues from the \( \alpha_1 \)-helix (Asp80 and Met69) primarily interact with and stabilize the lipid backbone, while Arg79, a key interacting residue with exogenous ligands, does not participate in ligand binding (Figure 2b). As the iGb3 used for crystallization was a short-chain version with a \( C_8 \) fatty acid, as with the previously crystallized short-chain \( \alpha \)-GalCer (PBS-25; Figure 1b), a spacer lipid (most likely palmitate) was sequestered deep inside the bottom of the \( A' \) pocket (Figure 2c, yellow).

Interestingly, differences in the positioning of the palmitate could be observed when compared to the spacer lipid found in the CD1d-\( \alpha \)-galacturonosyl ceramide structure.\(^{18}\) The longer \( C_{14} \) fatty acid of \( \alpha \)-galacturonosyl ceramide pushes the spacer lipid (cyan) slightly deeper into the pocket, while the shorter \( C_8 \) of iGb3 allows the spacer lipid to be situated slightly closer to the entrance of the \( A' \) pocket, with the carboxyl moiety forming hydrogen bonds with Gln14 and, via two water molecules, with Arg74 and Ser28. In that regard, such flexible placing of the same spacer lipid (i.e., the abundant cellular palmitate) could compensate for differential insertion of the variable length alkyl chains into the \( A' \) pocket (\( C_8 \) versus \( C_{14} \)) thereby stabilizing the \( A' \) pocket. Alkyl chains of \( C_{16} \) in length seem to be sufficient to stabilize the \( A' \) pocket by themselves without a residual pocket for a spacer lipid.\(^{17}\)

**Comparison of CD1d-glycolipid complexes**

We compared the binding of iGb3 to that of sulfatide, a less complex \( \beta \)-linked glycolipid and \( \alpha \)-GalCer (an \( \alpha \)-anomeric glycolipid) (Figure 2d–f). As iGb3 and sulfatide are very similar in chemical structure up to the proximal sugar, the binding was indeed very similar as well, with only a slight lateral shift of the carbohydrate and a slight rotation of the 3’-OH group of the sphingosine chain. The 3’ sulfate is absent from iGb3, but the second sugar of the 4-position occupies a similar space. As expected, \( \alpha \)-GalCer binding is very different from that of iGb3, due to the different glycosidic linkage (for a more detailed comparison between \( \alpha \) and \( \beta \)-linked glycolipids, see Ref. 19). However, this raises the interesting question of how the same TCR can recognize \( \alpha \)-GalCer and the structurally different iGb3, but not sulfatide or non- sulfated \( \beta \)-GalCer, which are both similar in structure to iGb3.

**Expression of mouse \( V\alpha 14 \) TCR for crystallization**

We have expressed multiple \( V\alpha 14 \) recombinant TCRs in a fly expression system and measured their affinities in multiple CD1-lipid complexes.\(^{16,25}\) Although we could achieve high levels of homogeneity and purity by chromatography, low yields remained a stumbling block on the way to crystallographic studies. Because of the wealth of expression success, as well as structural information, on TCR 2C, its \( V\beta 8 \) chain, which is in the set of \( V\beta \) families that pair with \( V\alpha 14J\alpha 18 \) for iGb3 and other glycolipid recognition, was used for design of recombinant proteins. The various constructs that were engineered and tested for expression are given in Table 2. The main lesson from those multiple partial successes and failures is that pairing of \( V\alpha 14 \) \( \alpha \) chains with \( \beta \) chains is very inefficient and that, in the absence of a proper transmembrane segment, \( \alpha\beta \) association is of low affinity. Expression was initially achieved by modifying our prototypic soluble TCR construct\(^ {14}\) with the addition of leucine zippers to both truncated chains in order to increase heterodimerization.\(^ {25}\) This single modification resulted in secretion of soluble protein with low yield (50–100 \( \mu \)g/l). Removal of the FG loop of the \( \beta \) chain was sufficient to abrogate expression altogether, recapitulating for recombinant molecules what had been observed for \( \Delta FG \) transgenic mice in which NKT cells were absent.\(^ {26}\) Further efforts were focused on re-engineering the \( \alpha/\beta \) interface. Addition of a single
disulfide to the center region of the Ca/Cβ interface, as proposed by Jakobsen, did not improve the level of expression. Similarly, re-engineering of the Va/Vβ interface to make it more 2C Va/Vβ-like (Figure 3b) resulted in a low level of expression. However, the combination of a stronger Ca/Cβ association combined with a re-designed Va/Vβ interface that recapitulates several interactions specific for the 2C TCR, produced a more robust expression (>1 mg/l). These modifications applied to the Va14 chain did not perturb the display of the CDRs or the anatomy of the interacting part of the receptor (Figure 3a). To verify the functionality of the new recombinant receptor, affinities for CD1/α-GalCer complexes were measured by surface plasmon resonance (Figure 4). The short chain α-GalCer, PBS-25, was loaded onto CD1d and compared with binding of TCR to “empty” CD1d molecules. After subtraction of the sensorgrams corresponding to empty CD1, association and dissociation were fitted using the BIAl evaluation package and measured as $4 \times 10^4$ M$^{-1}$ s$^{-1}$ for $K_{ass}$ and $8.3 \times 10^{-3}$ s$^{-1}$ for $K_{diss}$, almost identical with the values measured with the original TCR construct. Measurement of the affinity TCR/CD1d-iGb3 revealed a very transient interaction with a fast off-rate (0.5 – 1 s$^{-1}$) (Figure 4). The resulting modest affinity (50–100 µM) is a likely explanation for our inability to stain iNKT cells with CD1d-iGb3 tetramers. The large difference in TCR affinity (250-fold) towards CD1d-iGb3 and CD1d-α-GalCer may suggest a model in which the two CD1d complexes are seen differently (see below), although affinities and physical contacts (nature and number) have been shown in many antibody/antigen systems not to correlate well. However, in the case of iGb3, it is likely that the extended and exposed sugar moieties are essential for binding. However, protein–carbohydrate affinities are notoriously low, as has been well documented for antibodies and lectins, except when multivalency enhances the interaction. Therefore, the low affinity of the Va14 TCR for iGb3 could be due to its lectin-like properties. This concept was tested directly by investigating the binding of recombinant Va14 TCRs to individual sugars on a glycan microarray (Consortium for Functional Glycomics), but no clear binding for any sugars on the array were identified among a fairly high background (data not shown).

Crystal structure of mouse Va14 TCR

To address the structurally puzzling question of how different lipids can be recognized by the same TCR, we determined the crystal structure of the cognate Va14 TCR to 1.85 Å resolution (Table 1), which can then be used to model CD1-glycolipid-TCR interactions for further studies. The overall structure of the Va14 TCR is illustrated in Figure 5a, with the CDR3s color-coded to match the sequence highlighted in Figure 5b. The crystal lattice contains two TCR molecules in the asymmetric unit, which are overall very similar in structure (root-mean-square deviation (r.m.s.d.) of 1.04 Å for all 422 Ca atoms of the αβ-heterodimeric TCR), except for the CDR loops, especially CDR3α and CDR3β (Figure 5c), which give rise to two differently configured antigen binding pockets (outlined in more detail in the following section). Whereas the two independent variable α and variable β chains superimpose quite well (r.m.s.d. of ~0.6–0.7 Å for 49 Va and 57 Vβ central β-sheet residues, respectively), with their human Va24Ja18 counterparts (PDB codes 2EYR and 2CDE31,32). The juxtaposition of the Va14Vβ8.2 heterodimer in mouse is different from that of human Va24β11 and, hence, the assembled αβ heterodimers do not superimpose well (r.m.s.d. of 1.7 Å for 106 Cα atoms overall) (see Figure 5d, superimposition of the α chain). In addition, CDR2α, CDR3α and CDR3β adopt different conformations in the human and mouse TCR, whereas the equivalent CDR1 loops superimpose quite well. For pMHC-restricted TCRs, the CDR3s primarily recognize the ligand and, if it were also true for Va14 TCRs, their flexibility in the unbound TCR state may reflect their capacity to adapt to different conformations that permit recognition of diverse glycolipid epitopes presented by CD1.

†www.functionalglycomics.org
CD1d-glycolipid-Vα14 TCR model

Analysis of the Vα14 TCR revealed two differently shaped pre-formed binding pockets in the two TCR molecules of the crystal. Similar pockets have been reported for the human Vα24 TCR binding pockets, but not to the extent seen here for the mouse TCR.31,32 The larger of the two Vα14 binding pockets is wider and much deeper than the corresponding human Vα24 pocket (Figure 6). The structure of the human CD1d-α-GalCer-Vα24 TCR complex has been determined recently.33 It revealed a surprising “parallel” TCR docking mode in contrast to the more general “diagonal” docking of pMHC TCR’s. Furthermore, CDR1α, CDR3α and CDR2β of the TCR form the most contacts with the CD1d-α-GalCer complex. Five hydrogen bonds are formed between CDR1α and CDR3α and the glycolipid ligand, while CDR2β interacts with Glu83 of the α1-helix of hCD1d. We tried to generate a mouse CD1d-TCR model, where both TCR and CD1d are in identical positions with regards to the human Va24-TCR complex (Figure 7a). Not surprisingly, however, this model resulted in serious steric clashes between the exposed iGb3 and several TCR residues, as the iGb3 glycolipid ligand is very different in size and structure from α-GalCer (versus 1 sugar). Next, we prepared two alternative CD1d-iGb3-TCR models, in which either the TCR or the iGb3 ligand required repositioning to form adequate interaction with each other (Figure 7b and c). In the “cavity” model, the mouse Vα14 TCR binding pocket is a positively charged, large and deep cavity that can accommodate the distal two carbohydrates of iGb3. The distal galactose that was not ordered in the crystal structure but was included in both TCR models (Figure 7b and c). The CDR3s of both the α chain and the β chain are responsible for shaping the groove. It should be noted, however, that due to the deletion of four residues of the J-region of CDR3α (to match the length of CDR3α of the 2C TCR), the native TCR binding pocket may be slightly different in size and shape and even deeper, as the CDR3α loop may extend further toward the CD1d surface and make better contact with the α1-helix of CD1d (Figure 7c). The “squashed” model requires no re-orientation and allows for a conserved TCR docking onto CD1d (Figure 7b). Essentially, rotation of two bonds between the lipid backbone and the iGb3 headgroup was enough to accommodate the three sugars between the CD1d and TCR surfaces without any steric clash. Recent mutational data on the mouse Vα14 TCR clearly favor the squashed model, as it conserves the CD1d-TCR specific interactions (e.g., Y48 and Y50 of CDR2β with E83 of mCD1d). These interactions change upon repositioning the TCR to form the cavity model.34 However, both models have their limitations, especially as the process of breaking existing and forming new hydrogen bonds between CD1d and iGb3 to obtain the squashed model is energetically unfavorable. On the other hand, different CDR3β usage would change the size and charge of the cavity, and could influence ligand recognition and discrimination. However, it is commonly accepted that the CDR3β region of the semi-invariant Vα14 TCR has, in general, very little influence on ligand recognition. Moreover, the human CD1-αGalCer-Vα24 TCR structure demonstrates clearly that, at least for α-GalCer recognition, CDR3β contacts neither CD1d nor the glycolipid.

Discussion

Beyond their capacity to regulate early adaptive immune responses, iNKT cells have emerged as a fascinating system to investigate due to their cell surface display of a semi-invariant TCR that recognizes multiple lipids and glycolipids.1 IGb3 and α-GalCer represent two extremes of what the iNKT receptor can recognize. The former is a β-linked ceramide and displays three sugars, whereas the second one is α-linked and has only one sugar. These fundamental differences coincide with the origin of these ceramides: α-linked ceramides are all exogenous, as mammalian genomes do not encode the proper enzymes to produce this linkage, whereas all endogenous ceramides that are potential ligands of iNKT cells are β-linked, like iGb3. A comparison of the current iGb3 structure with structures of CD1d-α-GalCer16,20 clearly shows the physical impact of an α-versus β-linkage. Whereas, the galactose sugar ring of α-
GalCer lies almost parallel with the surface of CD1, the glucose of iGb3 projects perpendicularly upwards compared to α-GalCer. The rigidity of the linkage and the positioning of the first sugar is re-enforced by a network of hydrogen bonds that immobilizes the neck and first sugar of the ceramide. This scenario seems to argue against the possibility that, upon recognition by TCR, the sugar might adopt a different conformation to fit the interface, as in the case of some peptide side chains in TCR/pMHC complexes. However, we could still argue that CDR3α, the proposed main player in these glycolipid ligand interactions, could be the flexible domain in the resulting interface. However, the fact that the on-rates for these diverse ligands are very similar does not favor this model. Similarly, the second sugar of iGb3 also points straight up, with the third sugar attached to it, and would form a very extended structure that would be difficult to accommodate in a normal TCR/MHC-like interface.

These issues require us to speculate on the potential role and shape of the two cavities in the crystal TCR structure that are found between the two CDR3 regions. These cavities are found in the mouse receptor, and in its human ortholog Vα24/Vβ11, which also can recognize CD1-iGb3, whilst none of the other TCR structures reported to date have such major cavities.

Several members of the lectin superfamily, as well as alginate binding proteins from Gram-negative bacteria, reveal a central cleft, which is often formed by two globular domains of the proteins. These clefts are the binding sites for various molecules, including oligosaccharides. We were intrigued by this mode of carbohydrate binding and were able to dock the carbohydrate headgroup of iGb3 into the central cavity of the Vα14 TCR. Modeling three sugars bound in the cavity gives a near-perfect fit, suggesting that, if it is indeed the mode of binding for endogenous ligands, most of them could be tri-saccharide ceramides. We know that a diglycosylceramide with the same first two sugars as iGb3 (lactosylceramide, LacCer) is non-stimulatory and that one additional sugar, as found in iGb4 (GalNAc β1-3Gal α1-3Gal β1-4Glc β1-1Cer) is also non-stimulatory. We have established that the last sugar of iGb3 was critical for TCR recognition since Gb3 (α1-4 instead of α1-3 linkage of the terminal galactose), mannosyl and glucosyl-iGb3 (α1-3man-LacCer and α1-3gluc-LacCer) were also non-stimulatory (P.B.S., A.B. and L.T., unpublished results). The importance of length for β-linked ceramides, the nature of the linkage for the terminal sugar (α versus β) and fine chemistry of the third sugar argue for very strict physical requirements for recognition, suggestive of an intimate interaction between all of the sugars, especially the terminal galactose with the end of the cavity. In this way, the cavity could act as a ruler (length) and, at the same time, would have specificity (chemistry of the terminal sugar). On the other hand, a comprehensive mutagenesis approach in which every CDR residue of mouse Vα14Vβ8.2 TCR was mutated to alanine suggests that recognition of the glycolipids α-GalCer, the structurally related Sphingomonas antigen GalA-Gsl and iGb3 are mediated by the same residues of the TCR. Hence, this would imply that the antigens are recognized by the TCR in a similar way and, therefore, docking of the TCR onto CD1 glycolipid should be highly similar, if not identical.

In addition, it has been demonstrated that certain pMHC-restricted TCRs have the capacity to flatten bulged peptides, similar to our squashed CD1-iGb3-TCR model. This could also result in intimate contact between iGb3 and the TCR α chain, possibly providing similar specificity for the terminal sugar, as the shape complementarity between the sugars and the surface of the α chain could then become increasingly important for ligand discrimination. However, if the iGb3 headgroup is accommodated between the CD1 and TCR surfaces, we would expect more interaction of the TCR with the much larger tri-hexosyl headgroup of iGb3 compared to the simple galactose sugar of α-GalCer. These additional interactions, in turn, should result in differences in the recognition of iGb3 versus α-GalCer, and it is currently not clear why no difference has been detected in mutational studies.

Our future efforts are focused on the crystallization of the Vα14TCR/CD1d-iGb3 complex, even though it is likely to be challenging because of its low affinity. In parallel, we will attempt to probe the function of the cavity by closing it using the 2Cαβ TCR structure as a rational
template for the design of the required mutations. A successful closure should produce a receptor still able to respond to CD1-iGalCer, but unable to recognize CD1-iGb3. Mutagenesis of the residues lining the bottom of the cavity will also address specificity with respect to the various β-linked trihexosyl ceramides that we have at our disposal. These experiments should confirm and aid in understanding of the unusual dual mode of recognition of NKT cell receptors that is strongly suggested by the CD1 and TCR structures presented here. As we have already seen for antibody and TCR diversity systems for antigen recognition, there are often multiple ways of recognizing the same ligand, and similar ways of recognizing multiple ligands.39,40

Material and Methods

Protein expression, purification and crystallization

Expression and purification of CD1d has been reported elsewhere.16 The synthesis of the short-chain iGb3, PBS-74, was carried out by PBS (unpublished results). PBS-74 was solubilized in phosphate-buffered saline, 0.05% (v/v) Tween20 and loaded into recombinant CD1d overnight at room temperature at a tenfold molar excess in the presence of Saposin B. Loaded complexes were re-purified by gel filtration and concentrated to 10 mg/ml in 100 mM Hepes (pH 7.5), 100 mM NaCl. For TCR expression, the original cDNAs used for the expression of zippered Vα14/2Cβ heterodimers were modified by PCR to produce the various mutants discussed in the text and figures. All transfections in S2 cells were carried out by calcium-phosphate precipitation, and puromycin-resistant lines were derived for expression. The double mutant TCR for structural studies (additional Cα/Cβ disulfide bridge, modified Vα/Vβ interface) was purified by a succession of Ni-NTA-agarose chromatography, ion-exchange chromatography and gel filtration. For crystallization studies, zippers were removed by thrombin cleavage (50 units/10 mg protein) and treatment with carboxypeptidase B (1:200 w/w) for 3 h at room temperature. Digested TCR was re-purified by hydrophobic interaction chromatography and gel filtration before being concentrated to 10 mg/ml in 100 mM Hepes (pH 7.5), 100 mM NaCl. Crystallization was tested using the sitting-drop technique in 96-well plates with commercial (Hampton Research, Aliso Viejo, CA, and Emerald Biosystems, Bainbridge Island, WA) and in-house matrix screens. Conditions were optimized in 1 μl drops using in-house reagents. Final crystallization conditions were 18% (w/v) PEG 4000, 0.2 M ammonium citrate and 2% (v/v) butanol for CD1d-PBS-74 at 10 mg/ml, and 16% (w/v) PEG 4000, 0.2 M magnesium acetate, 0.1 M Tris (pH 8.5) for the TCR at 10 mg/ml.

Structure determination

Crystals were flash-cooled at 100 K in mother liquor containing 25% (v/v) glycerol. Diffraction data from a single CD1-iGb3 crystal were collected at beamline 8.2.1 of the Advanced Light Source (ALS) and processed to 1.9 Å resolution with the Denzo-Scalepack suite41 in monoclinic spacegroup P2_1 (unit cell dimensions a=58.94 Å, b=96.2 Å, c=77.93 Å, β=106.23°). Two CD1-lipid complexes occupy the asymmetric unit with an estimated solvent content of 44.2% (v/v) based on a Matthews' coefficient (V_M) of 2.2 Å³/Da. Molecular replacement in P2_1 was carried in CCP442, using the program PHASER43 with the CD1d-sulfatide structure (PDB code 2AKR) as the search model minus the ligand, and resulted in a Z-score of 63. Subsequent rigid-body refinement in REFMAC 5.2 produced an R_cry of 36.3%. Initial refinement included several rounds of restrained refinement against the maximum likelihood target in REFMAC 5.2. At a later stage of refinement, carbohydrates were built at all three N-linked glycosylation sites in CD1d. Refinement progress was judged by monitoring the R_free for cross-validation.44 The model was rebuilt into σA-weighted 2F_0−F_c and F_0−F_c difference electron density maps using the program O.45 Water molecules were assigned during refinement in REFMAC using the water ARP module for >3σ peaks in a F_0−F_c map and retained if they satisfied hydrogen-bonding criteria and returned 2F_0−F_c density >1σ after
refinement. Starting coordinates for the iGb3 ligand were obtained using the molecular modeling system INSIGHT II (Accelrys, Inc.). The iGb3 library for REFMAC was created using the Dundee PRODRG2 server. Final refinement steps were performed using the TLS procedure in REFMAC with three anisotropic domains per CD1-iGb3 complex (α1-α2 domain including carbohydrates and glycolipid ligand, α3 domain and β2M) and resulted in improved electron density maps for the glycolipid ligand and a further drop in R_free. The CD1d-iGb3 structure has a final R_cryst = 20.1% and R_free = 24.6% and the quality of the model (Table 1) was excellent as assessed with the program Molprobity.

Diffraction data from a single TCR crystal were collected at Beamline 11.1 of the Stanford Synchrotron Radiation Laboratory and processed to 1.85 Å resolution with the Denzo-Scalepack suite in spacegroup P2_1 (unit cell dimensions a=65.29 Å; b=150.52 Å; c=65.81 Å, β=117.04°). Two TCR molecules occupy the asymmetric unit with an estimated solvent content of 53.0% (v/v) based on a Matthews’ coefficient (V_M) of 2.6 Å^3/Da. Molecular replacement in P2_1 was carried out using the program MOLREP as part of the CCP4 program suite, and the 2C TCR (PDB code 1TCR) as the search model, with the ligand removed, and resulted in an R_cryst of 49.4% and a correlation coefficient (CC) of 0.34. Subsequent rigid-body refinement in REFMAC 5.2 using tight NCS restraints produced an R_cryst of 38.1%. Refinement and model building was performed as reported above for the CD1-iGb3 structure, but without maintaining any NCS restraints. TLS refinement was carried out as described above, but with eight anisotropic domains corresponding approximately to the Cα, Vα, Cβ and Vβ domains of molecules A and B of the asymmetric unit, including N-linked carbohydrates. The Vα14 TCR was refined to a final R_cryst of 19.3% and R_free of 22.8% and the quality of the model was assessed in Molprobity.

CD1-TCR complex modeling

To model the cavity CD1d-iGB3-Vα14 ternary complex, the coordinates of CD1d and the TCR were superimposed onto the corresponding regions of the previously prepared model of the CD1α-lipopeptide-TCR complex. The position of the TCR was manually adjusted to optimally position the large binding cavity of the TCR above the exposed iGb3 headgroup at the CD1 surface. The terminal, α1-3 linked galactose of iGB3, which was not ordered in the crystal structure, was then modeled into the remaining space of the TCR binding cavity. The final orientation showed a maximal fit of the binding cavity of the TCR and the headgroup of iGB3. The α chain of the TCR needed to be tilted slightly to interact better with the α1 helix of mouse CD1. The squashed model was prepared by analogy to the human CD1d-α-GalCer-Vα24 TCR structure by superimposing mCD1d onto hCD1d and Vα14 TCR onto the α chain of the Vα24 TCR. Clearly, other models are possible, but without further structural data on TCR-CD1 complexes, the model here is at least consistent with several independent pieces of data.

Affinity measurements

All surface plasmon resonance experiments were carried out using a BIAcore 2000 machine (BIAcoreAB, Uppsala, Sweden). For analysis of CD1d/αGC interaction, PBS-25 was loaded onto recombinant CD1d as reported, and re-purified after loading by gel filtration. CD1d-PBS-74 was prepared as reported above for crystallization trials. Loading was evaluated by native isoelectrofocusing as reported, and was estimated to be >90%. TCR was immobilized on CM-5 research grade sensor chip by amine coupling. Filtered and degassed PBS was used as the running buffer. Empty CD1d was used as negative control and sensorgrams for empty CD1d were subtracted from sensorgrams for CD1d/PBS-25 or CD1d/PBS-74 to evaluate real binding. Experiments were carried out at 25 °C. Analysis was performed with the BIAevaluation software package using global fitting. Measurements were repeated twice.
Structure presentation

The program PyMol‡ was used to prepare Figures 1, 2, 3 and 5, 6, 7 and 1, 2, 3 and 5, 6, 7. The PDB2PQR server,52 and the program APBS,53, were used to calculate the electrostatic surface potentials of Figures 2, 6 and 7. The Chemdraw program was used to prepare all chemical structures.

Protein Data Bank accession numbers

Coordinates and structure factors have been deposited in the PDB with accession codes 2Q7Y and 2Q86 for CD1d-iGb3 and TCR, respectively.

Acknowledgements

We thank Randy Stefanko for excellent technical support. A.B., P.B.S. and L.T. are supported by NIH grant AI053725. A.B. is an HHMI investigator. I.A.W. is supported by NIH grants CA58896 and GM62116.

Abbreviations used

TCR, T cell receptor; iGb3, isoglobotrihexosylceramide; α-GalCer, α-galactosylceramide; MHC, major histocompatibility complex; pMHC, peptide-major histocompatibility complex; CDR, complementarity determining region; r.m.s.d., root-mean-square deviation.

References


‡http://pymol.sourceforge.net/


Fig. 1.
A representation of the mCD1d-iGb3 complex (a) and chemical structures of CD1d ligands (b). a, The self-antigen iGb3 (yellow) is bound in the hydrophobic binding groove between the \( \alpha_1 \) and \( \alpha_2 \) helices of the CD1d heavy chain (grey) that associates non-covalently with \( \beta_2\)-microglobulin (\( \beta_2\M, blue-grey) to form a biological heterodimer. Three of the four N-linked glycosylation sites (Asn20 (N20), Asn42 and Asn165) carry well-ordered carbohydrates (grey sticks). The spacer lipid (C\(_{16}\), orange) present in the binding groove complements the short C\(_8\)-alkyl chain of the synthetic ligand iGb3. b, The chemical structure of short-chain iGb3 is different from that of cis-tetracosenoyl sulfatide (sulfatide C\(_{24:1}\)), which it resembles in the core structure, and the short-chain \( \alpha\)-GalCer, which is dissimilar due to the different anomeric conformation of the galactose (\( \alpha\)-versus \( \beta\)-glycosidic linkage). The terminal \( \alpha_1\)-3 linked galactose (red) is not ordered in the crystal structure and, therefore, not shown in Figures 1 and 2).
Fig. 2.
Overview of the CD1d-iGb3 structure (top row), and ligand binding comparison (bottom row).
(a) The 2F_o–F_c electron density map is contoured at 1σ and is shown as a blue mesh around
the iGb3 ligand. The proximal glucose is well ordered, while the distal (α1-3) galactose is
not ordered in the crystal structure due to lack of interaction with CD1 and, therefore, is not
shown. (b) Hydrogen bond interactions between CD1d residues (gray) and the polar moieties
of iGb3 (yellow). Note that CD1d residues of the α2-helix interact exclusively with the
carbohydrate headgroup of iGb3, while α1-helix residues participate in orientating the lipid
backbone. (c) A palmitic acid spacer lipid (yellow) is apparent in the A′ pocket, where it is
orientated by hydrogen bond interactions (blue broken lines). The depth and exact location of
the spacer lipids differ between the CD1d-iGb3 complex and the α-galacturonosyl complex
(spacer lipid in cyan, PDB code 2FIK) due to the different lengths of glycolipid alkyl chains
that are inserted into the A′ pocket. (d) Presentation of the iGb3 headgroup by CD1d looking
along the binding groove. Surface representation with electrostatic potential (red,
electronegative and blue, electropositive contoured from −30 to +30 kT/e). (e) Binding of the
iGb3 ligand is very similar to that of cis-tetracosenoyl sulfatide (green, PDB code 2AKR). (f)
Compared to α-GalCer, the iGb3 headgroup is far more exposed, as it projects away from the
CD1d surface. Several residues that are involved in the shaping of the CD1 binding grooves
(a and d) or in polar interactions with iGb3 (b and d) and the spacer lipid (c) are depicted in
the one-letter code with the residue number.
Fig. 3.
A cartoon representation and sequence of the engineered Vα14 TCR. (a) Mutated residues on the α chain (cyan) are highlighted in yellow. Several interacting partners of the β chain (gray) are depicted. Residues in red indicate the position in the α chain where the four residues were removed by mutagenesis. An interchain disulfide bond (green) between C158α and C167β was incorporated to increase the expression and stability of the TCR. (b) Sequence alignment between the parental Vα14 sequence and the mutated Vα14 sequence (Vα14m). Colored residues are depicted in a.
Fig. 4.
Low-affinity binding of CD1d-PBS-74 to recombinant TCR Vα14/2Cβ. (a) Successive
dilutions of empty CD1d or CD1d-PBS-74 were injected over immobilized TCR. Subtraction
(CD1-d-PBS-74 sensorgrams minus empty CD1 sensorgrams) and a 1:1 Langmuir fit of CD1d-
PBS-74 binding are presented. (b) Magnified view of the dissociation phase comparing empty
CD1d and CD1d-PBS-74. (c) Association constants of CD1d-PBS-25 and CD1d-PBS-74 for
Vα14/2Cβ TCR. Measurements were reproduced in two separate experiments.
Fig. 5. Vα14 TCR structure overview. (a) A representation of the TCR Vα14/2Cβ constant (C) and variable (V) domains of the α (orange) and β (gray) chains. The CDR3 loops are color-coded to match the sequence in (b). (c) Superimposition of the two independent copies of the TCR molecules in the asymmetric unit of the crystal. Top view into the TCR binding site (molecule A, gray with colored CDRs, molecule B orange). Note that the changes in the conformation of the CDR3 loops give rise to the different “pre-formed” binding pockets of the TCR. d, Superimposition of mouse Vα14 TCR (molecule A) onto the α chain of human Vα24 TCR (from Ref. 31; PDB code 2EYR).
Fig. 6.
Comparison of mouse and human iNKT cell TCR “binding pockets”. The variable domains of the mouse Vα14 TCR (left-hand column) and human Vα24 TCR (PDB code 2EYR, right-hand column) are shown in a surface representation with calculated electrostatic potentials (upper row) or zoomed in onto the putative binding pocket (lower row). The TCR α chain is cyan, and the β chain is gray. TCR residues that form the binding pocket are depicted. Note that Arg95α is disordered in the Vα14 structure and built only as an alanine. However, the remaining electron density in the vicinity of this residue suggests that the arginine side chain points away from the groove, rather than towards it, which would close the groove. In addition, Arg95α of the second Vα14 TCR molecule in the crystal projects straight up towards the viewer and is not involved in closing the preformed binding pocket. Rather, movement of the protein backbone in CDR3α and CDR3β is responsible for the closure of the pocket in one of the two observed TCR structures of the crystal.
Fig. 7.
A model of CD1d-iGb3-Vα14 TCR binding site. (a) A model of the mouse complex based on the human CD1d-aGC-Va24 structure (PDB code 2PO6) demonstrates steric clashes between the iGb3 ligand and the TCR surface. (b) The “squashed” iGb3 model was created to preserve the CD1d-TCR interaction while re-positioning the iGb3 headgroup. (c) Slight re-orientation of the TCR above the iGb3 ligand leads to the “cavity” TCR model, with the iGb3 headgroup accommodated within a deep TCR pocket. Note the positively charged nature of the binding pocket (assignment of the electrostatic surface potential of c is identical with Figure 2). Several residues involved in forming the binding pocket in c are illustrated.
Table 1

Data collection and refinement statistics

<table>
<thead>
<tr>
<th></th>
<th>CD1d-iGb3</th>
<th>Va14 TCR</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A. Data collection</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolution range Å</td>
<td>40.0–1.95 (2.02–1.95)</td>
<td>50.0–1.85 (1.90–1.85)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>97.9 (97.8)</td>
<td>93.0 (82.5)</td>
</tr>
<tr>
<td>No. unique reflections</td>
<td>59,816</td>
<td>90,127</td>
</tr>
<tr>
<td>Redundancy</td>
<td>2.7</td>
<td>2.0</td>
</tr>
<tr>
<td>R&lt;sub&gt;sym&lt;/sub&gt; (%)</td>
<td>6.0 (44.9)</td>
<td>4.0 (32.7)</td>
</tr>
<tr>
<td>I/σ</td>
<td>17.9 (2.3)</td>
<td>21.6 (2.0)</td>
</tr>
<tr>
<td><strong>B. Refinement statistics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of reflections (F&gt;0)</td>
<td>58,076</td>
<td>88,133</td>
</tr>
<tr>
<td>Maximum resolution Å</td>
<td>1.95</td>
<td>1.85</td>
</tr>
<tr>
<td>R&lt;sub&gt;cryst&lt;/sub&gt; (%)</td>
<td>19.9 (26.7)</td>
<td>19.2 (27.4)</td>
</tr>
<tr>
<td>R&lt;sub&gt;free&lt;/sub&gt; (%)</td>
<td>24.5 (35.2)</td>
<td>22.7 (32.5)</td>
</tr>
<tr>
<td>No. atoms</td>
<td>6592</td>
<td>7269</td>
</tr>
<tr>
<td>Protein</td>
<td>5952</td>
<td>6589</td>
</tr>
<tr>
<td>Ligands (iGb3/palmitate)</td>
<td>93/36</td>
<td></td>
</tr>
<tr>
<td>N-linked carbohydrate</td>
<td>159</td>
<td>134</td>
</tr>
<tr>
<td>Water molecules</td>
<td>352</td>
<td>546</td>
</tr>
<tr>
<td>Ramachandran statistics (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Favored</td>
<td>98.6</td>
<td>96.0</td>
</tr>
<tr>
<td>Outliers</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>r.m.s.d. from ideal geometry</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond length (Å)</td>
<td>0.015</td>
<td>0.016</td>
</tr>
<tr>
<td>Bond angles (deg.)</td>
<td>1.71</td>
<td>1.64</td>
</tr>
<tr>
<td>Average B-values&lt;sup&gt;e&lt;/sup&gt; (Å&lt;sup&gt;2&lt;/sup&gt;)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>31.7</td>
<td>35.3</td>
</tr>
<tr>
<td>Ligands (iGb3/palmitate)</td>
<td>55.0/46.3</td>
<td>None</td>
</tr>
<tr>
<td>Water molecules</td>
<td>36.7</td>
<td>37.1</td>
</tr>
<tr>
<td>Carbohydrates</td>
<td>49.1</td>
<td>49.0</td>
</tr>
</tbody>
</table>

<sup>a</sup>Number in parentheses refer to the highest resolution shell.

<sup>b</sup>R<sub>sym</sub> = (∑<sub>h</sub>Σ<sub|i</sub>|I<sub>i(h)</sub>−<i<l(h)></i|)/|Σ<sub>h</sub>Σ<sub>i</sub>I<sub>i(h)</sub>|×100, where <i<l(h)></i> is the average intensity of i symmetry-related observations for reflections with Bragg index h.

<sup>c</sup>R<sub>cryst</sub> = (∑<sub>hkl</sub>|Fo − Fc|/Σ<sub>hkl</sub>|Fo|)×100, where Fo and Fc are the observed and calculated structure factors, respectively, for all data.

<sup>d</sup>R<sub>free</sub> was calculated as for R<sub>cryst</sub>, but on 2% (CD1d) and 1.5% (TCR) of data excluded before refinement.

<sup>e</sup>B-values were calculated with the CCP4 program TLSANL<sup>48</sup>.
<table>
<thead>
<tr>
<th>Modifications</th>
<th>Expression</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Removal of cytoplasmic tails and transmembrane domains</td>
<td>−</td>
<td>25,54</td>
</tr>
<tr>
<td>• Removal of cytoplasmic tails and transmembrane domains</td>
<td>+</td>
<td>25</td>
</tr>
<tr>
<td>• Addition of leucine zippers</td>
<td>−</td>
<td>L.T., unpublished results</td>
</tr>
<tr>
<td>• Removal of the β chain FG loop</td>
<td>+</td>
<td>This study</td>
</tr>
<tr>
<td>• Addition of additional disulfide bridge Cα/Cβ</td>
<td>+</td>
<td>This study</td>
</tr>
<tr>
<td>• Removal of cytoplasmic tails and transmembrane domains</td>
<td>+</td>
<td>This study</td>
</tr>
<tr>
<td>• Addition of leucine zippers</td>
<td>+</td>
<td>This study</td>
</tr>
<tr>
<td>• Re-engineering of Vα/Vβ interface</td>
<td>+++</td>
<td>This study</td>
</tr>
<tr>
<td>• Addition of leucine zippers</td>
<td>+</td>
<td>This study</td>
</tr>
<tr>
<td>• Re-engineering of Vα/Vβ interface</td>
<td>+</td>
<td>This study</td>
</tr>
<tr>
<td>• Addition of additional disulfide bridge Cα/Cβ</td>
<td>+</td>
<td>This study</td>
</tr>
</tbody>
</table>

Symbols: −, expression below 50 μg/l, +, expression between 50 and 100 μg/l, ++, expression between 100 and 500 μg/l, +++ expression above 1 mg/l.