Mouse Clr-g, a Ligand for NK Cell Activation Receptor NKR-P1F: Crystal Structure and Biophysical Properties

Tereza Skálová, Kristýna Kotýnková, Jarmila Dusková, Jindrich Hasek, Tomáš Kovač, Petr Kolenko, Petr Novák, Petr Man, Pavel Hanc, Ondrej Vanek, Karel Bezouska and Jan Dohnálek

*J Immunol* 2012; 189:4881-4889; Prepublished online 15 October 2012;
doi: 10.4049/jimmunol.1200880
http://www.jimmunol.org/content/189/10/4881
Mouse Clr-g, a Ligand for NK Cell Activation Receptor NKR-P1F: Crystal Structure and Biophysical Properties

Tereza Skálova,* Kristýna Kotýnková,‡+ Jarmila Dušková,* Jindřich Hašek,* Tomáš Kovařík,† Petr Kolenko,*, Petr Novák,‡+ Petr Man,‡+ Pavel Hanč,‡+ Ondřej Vanček,‡+ Karel Bezouška,‡ and Jan Dohnálek*†

Interactions between C-type lectin-like NK cell receptors and their protein ligands form one of the key recognition mechanisms of the innate immune system that is involved in the elimination of cells that have been malignantly transformed, virally infected, or stressed by chemotherapy or other factors. We determined an x-ray structure for the extracellular domain of mouse C-type lectin related (Clr) protein g, a ligand for the activation receptor NKR-P1F. Clr-g forms dimers in the crystal structure resembling those of human CD69. This newly reported structure, together with the previously determined structure of mouse receptor NKR-P1A, allowed the modeling and calculations of electrostatic profiles for other closely related receptors and ligands. Despite the high similarity among Clr-g, Clr-b, and human CD69, these molecules have fundamentally different electrostatics, with distinct polarization of Clr-g. The electrostatic profile of NKR-P1F is complementary to that of Clr-g, which suggests a plausible interaction mechanism based on contacts between surface sites of opposite potential. The Journal of Immunology, 2012, 189: 4881–4889.
The mouse activating NK cell receptor NKR-P1F was shown to bind Clr-g in the original study (8). However, more recently, a much broader pattern of reactivities has been described, allowing the separation of the NKR-P1 receptor family into two subfamilies on the basis of chromosomal location, multiple sequence alignment of their amino acid sequences, and three-dimensional structures predicted by molecular modeling (10, 17–19). Based on the chromosomal gene arrangement, the proximal (centromeric) cluster encoding NKR-P1A, NKR-P1B, NKR-P1C, and NKR-P1D is distinguished from the distal (telomeric) cluster represented by NKR-P1F and NKR-P1G. The rat and mouse NKR-P1F and NKR-P1G receptors demonstrate a striking, cross-species conservation of specificity for a wide range of Clr ligands and, thus, were proposed to serve as promiscuous receptors for the rapidly evolving Clr family (18). Nevertheless, the activation potential of these receptors on NK cells has been subject to question, because cross-linking of mouse NKR-P1F (mNKR-P1F) using the recently obtained specific mAb failed to induce any enhancement of cytotoxicity or production of IFN-γ (20), and NKR-P1G has been identified as an inhibitory receptor (18). In contrast, NKR-P1F expressed on the surface of dendritic cells was shown to engage Clr-g on T cells and costimulate T cell proliferation and IL-2 production, thus maintaining the T cell response and memory in vivo (21). This role for the NKR-P1F/Clr-g interaction in the immune system is further supported by results showing that silencing of Clr-g expression leads to defects in T cell response to antigenic stimulation (22).

Previously, the evidence for specific interactions between NKR-P1 receptors and their Clr ligands was obtained with interaction assays based on the use of cellular transfectants, artificial receptor oligomers, and functional immunological tests. These determinations have a particular advantage of being very sensitive and reasonably specific, provided that a large number of appropriate controls has been included in the assays. However, such analyses cannot provide much data about the molecular details of these interactions, which can be obtained from high-resolution structures of the corresponding receptors and their ligands, as well as detailed structural and physicochemical investigation of individual receptor–ligand complexes performed with structurally well defined proteins corresponding to their natural arrangement within the plasma membrane. Recently, amino acid residues critical for the interaction between the human receptor NKR-P1 and its ligand LLT1, an ortholog of rodent Clr, were identified using mutagenesis and surface plasmon resonance to monitor the complex formation (23). For this receptor–ligand pair, the key residues are localized on the surfaces of the molecules near the surface loop of the C-type lectin-like (CTL) fold distal to the stem region.

Recently, the three-dimensional structure of the extracellular domain of mouse activation receptor NKR-P1A has been solved by x-ray diffraction (24). This is the only known structure of an NK receptor with CTL fold and also interacting with a CTL ligand. The core of the CTL domain of this receptor is homologous to the one revealed in other structures belonging to this protein family while about one quarter of the protein is in the form of extended loops interacting tightly with a neighboring loop in the crystal. A functional dimeric form of the receptor was suggested, with a potential role of the evolutionarily conserved loops in the interaction with ligands.

To our knowledge, we report the first three-dimensional structure of a Clr ligand for NKR-P1, mouse Clr-g (mClr-g). The structure closely follows the fold of the CTL family immune receptors, and the receptor dimer mimics that of human CD69 (hCD69) (25, 26). Thus, this study establishes the framework for subsequent detailed structural investigations of NKR-P1–Clr protein complexes and reveals the importance of electrostatics in the specificity of interaction of NK cell receptors with their ligands. To the best of our knowledge, this aspect has not been studied for this group of receptors and their ligands.

Materials and Methods

Bacterial expression, refolding, and purification of mClr-g

The extracellular portion of mClr-g (Asn86–Asn206) (8) was amplified using RT-PCR from total RNA isolated from splenocytes of C57BL/6 mice and subcloned into expression plasmid pET-30a (24). Mutation of Cys148 to Ser was performed using the QuikChange II Site-Directed Mutagenesis kit, and the final product was verified by DNA sequencing. Protein was expressed in Escherichia coli BL-21 Gold cells with initiation Met35 and inclusion bodies were isolated using a standard protocol (24). Inclusion bodies were dissolved in a guanidine HCl buffer under reducing conditions, and the solubilized protein was refolded by drop-wise addition of denaturing solution into cold (4°C) refolding buffer (20 mM HEPES [pH 7.5] with 1 M L-arginine, 3 mM cystamine, 9 mM cysteamine, 0.3 M 3-(benzylidemethylammonio) propane sulfonate, 1 mM NaN3, and 1 mM PMSF, 1 μM leupeptin, and 1 μM pepstatin). The resulting mixture was dialyzed twice, and the protein was purified by a two-step chromatography on S-Sepharose FF and Superdex 75. A single monodisperse protein peak eluted from the Superdex 75 column in 10 mM HEPES (pH 7.5) with 100 mM NaCl and 1 mM NaN3 was concentrated to 2.3 mg/ml and stored in this buffer at 4°C.

Characterization of mClr-g

The identity, purity, and native size of the produced protein were characterized using SDS-PAGE and mass spectrometry under reducing and nonreducing conditions, peptide mapping, gel filtration, and analytical ultracentrifugation (27). Disulide bond arrangement was analyzed using a previously described procedure (28). Mass spectrometric analyses were performed using direct infusion or by liquid chromatography-mass spectrometry (MS) mode analysis on a Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometer (APEX-ULTRA Qe FT-ICR; Bruker Daltonics) (26).

Dynamic light scattering

The dynamic light-scattering experiments were performed at 18°C with a Zetasizer Nano (Malvern Instruments) and a 45-μl quartz cuvette with ∼40 μl mClr-g solution diluted to 1.5 mg/ml in a buffer solution containing 10 mM HEPES (pH 7.5), 100 mM NaCl, and 1 mM NaN3.

Crystallization and data collection

The purified mClr-g, at a concentration of 2.3 mg/ml, was stored in 10 mM HEPES (pH 7.5), 100 mM NaCl, and 1 mM NaN3. The Hampton Research Crystal screen I, II, and Index were used for crystallization screening. The trials were performed in sitting drops of volume 0.4 + 0.4 μl (protein plus reservoir solution) and using 70 μl reservoir solution. Rod-shaped crystals, with size up to 500 μm, grown in the Crystal screen II no. 17 (35% [v/v] tert-butanol, 0.1 M sodium citrate tribasic dihydrate [pH 5.6]), gave no diffraction and were not reproducible, even when the seeding, sitting and hanging drop, and microbatch techniques were tested. Finally, spontaneously grown rod-shaped crystals with a length ∼200 μm were found in the protein stock. The crystals were flash-cooled after cryoprotection with polyethylene glycol 400, ethanol, and glycerol, as well as without any cryoprotectant. Polyethylene glycol 400 and ethanol damaged crystals visibly, whereas glycerol did not. However, a crystal without any cryoprotection provided the best diffraction and was finally used for data collection.

Single-crystal diffraction data were collected at the synchrotron radiation source Bessy II, beamline BL 14.1 (29), with the crystal mounted in a cryo-loop, at 100 K using a MAR Mosaic 225 CCD detector.

Data processing and refinement

The data were indexed, integrated, and scaled in HKL2000 (30) in the primitive orthorhombic Bravais lattice. The data collection and processing parameters are shown in Table I. Salt and ice diffraction were apparent in the diffraction images. To exclude reflections at the ice diffraction rings, the “Slope” parameter in SCALEPACK was set to 30 to decrease the maximal allowed slope of background of an included reflection. The data were then imported into the CCP4 package (31) and converted to structure factor amplitudes using TRUNCATE (32).

The phase problem was solved by molecular replacement in the online version of BALBES (33) in space group P21 using a chain of an extracellular

Downloaded from http://www.jimmunol.org/ by guest on April 30, 2017
domain of human receptor CD69 with Protein Data Bank (PDB) code 3HUP (34) and sequence identity 36% (ClustalW) (35). Four chains forming two dimers were placed in the asymmetric unit, the mClr-g sequence was assigned, and the structure was refined in REFMAC5.5 (36), with the following final BALBES parameters: \(Q = 0.587, R = 0.444,\) and \(R_{	ext{free}} = 0.479,\) where \(Q\) is a quality factor (37).

After several initial manual building and refinement cycles in REFMAC and COOT (38), the model of mClr-g was transformed to space group R2\(_1\)/2\(_1\)2\(_1\)2, which is the correct space group. No molecular replacement solution could be found directly in P2\(_1\),2\(_1\)2\(_1\), possibly because of tight crystal contacts between not-so-large mClr-g molecules (e.g., extended N terminus of chain B inserted into a neighbor molecule). In P2\(_1\),2\(_1\)2\(_1\) there is one dimer of mClr-g in the asymmetric unit.

The structure was refined using REFMAC, and manual modifications were performed in COOT. 5% of reflections were used as test reflections for the \(R_{	ext{free}}\) statistic. The last cycle of the refinement was performed using all reflections. The quality of geometry, contacts, and consistency with electron density maps were assessed by the validation tools available in COOT and by the ADIT–PDB validation server (39).

Figs. 2–4 were prepared in PYMOL (DeLano Scientific, San Carlos, CA; http://www.pymol.org).

### Accession numbers

The coordinates and structure factor files were deposited in the PDB under accession code 3RS1 (http://www.rcsb.org/pdb/explore/explore.do?structureId=3rs1).

### Homology modeling

A homology model of mNKR-P1F was prepared using the structure of mouse NKR-P1A (mNKR-P1A) (PDB code 3MQ2) (24). Only the part of the mNKR-P1F sequence overlapping with the structurally determined part of mNKR-P1A was taken into account. These parts are 50% identical, matching 139 residues of mNKR-P1A to 129 residues of mNKR-P1F (ClustalW) (35).

A homology model of mClr-b was prepared using the structure of mClr-g in an analogous procedure. Only 122 residues of the mClr-b sequence aligned with the structurally determined part of mClr-g (122 residues) were modeled, with sequence identity of 76% (ClustalW) (35). The levels of identity in both cases enable homology modeling in the safe homology-modeling zone (40).

The monomers of mNKR-P1F and mClr-b were modeled using the Swiss-model server (41). A dimer of each model was then built, based on secondary structure matching (SSM) of the individual chains to the dimer of the template. For mNKR-P1F, we selected the dimer type B of the receptor originally observed in the crystal structure of mNKR-P1A (24).

### Electrostatic potential

Electrostatic potentials for selected dimers with the CTL fold were calculated by solving the Poisson–Boltzmann equation in APBS (42). Partial charges were assigned in PDB2PQR (43), based on potential Amber. Hydrogens were added in PROPKA (44), including optimization of hydrogen bonds for the proteins at pH 7.5. The Poisson–Boltzmann equation was solved for dielectric constants \(\varepsilon_{\text{protein}} = 78.54\) and \(\varepsilon_{\text{solvent}} = 2\), assuming a 0.225-M solution of ions in solution with charge +\(\varepsilon\), the radius 2 Å.

### Macromolecular docking

The macromolecular docking was performed using the RosettaDock protein–protein docking server (45) in three preset orientations of the observed mClr-g dimer and the model of the mNKR-P1F dimer.

### Results

#### Production and biochemical characterization of mClr-g

The studied mouse protein Clr-g (UniProtKB/Swiss-Prot entry Q9WV9P.1) is a type II transmembrane glycoprotein of 217 aa. The N-terminal cytoplasmic sequence encompasses amino acids Met\(^{\text{1}}\) to Lys\(^{\text{53}}\) and includes splice variants responsible for individual isoforms of this receptor. The transmembrane region spanning Leu\(^{\text{24}}\) to Ala\(^{\text{28}}\) represents a potential helical signal-anchor for type II membrane protein. The extracellular part (Leu\(^{\text{2}}\)–Val\(^{\text{117}}\)) includes the predicted CTL domain (Cys\(^{\text{92}}\)–Lys\(^{\text{104}}\)), which is stabilized by two canonical predicted disulfide bonds (Cys\(^{\text{92}}\)–Cys\(^{\text{103}}\) and Cys\(^{\text{285}}\)–Cys\(^{\text{292}}\)), and contains a single potential N-glycosylation site at Asn\(^{\text{112}}\). For structural studies, an extracellular soluble portion of the protein was expressed corresponding to the above-defined CTL domain with a few extra amino acids (Asn\(^{\text{112}}\)–Asn\(^{\text{206}}\) and with the translation initiation Met\(^{\text{53}}\)) introduced artificially at the beginning of the polypeptide.

The above-described portion of mClr-g was amplified by RT-PCR using total RNA isolated from spleens of C57BL/6 mice and specific oligonucleotide primers. Because of the initial problems with protein in vitro refolding caused by unpaired cysteine Cys\(^{\text{148}}\), this residue was mutated to serine using the standard site-directed mutagenesis protocol. The final product was verified by DNA sequencing.

For large-scale protein production, bacterial cultures were grown at 37°C and induced with 0.1 mM isopropyl-\(\beta\)-D-thiogalactopyranoside, and cells were harvested 3 h after induction. The identity of the produced proteins was confirmed by automated Edman degradation, providing the expected N-terminal sequence NKTYAAAXSKN, where X indicates the position of the cysteine residue not easily identifiable by this method.

The in vitro refolding of the protein from inclusion bodies was optimized. In addition to the standard compounds, including 1 M L-arginine used as a low molecular mass chaotrope and a redox refolding buffer composed of 3 mM cysteamine and 9 mM cysteine that had proven useful previously (24, 27), two other conditions turned out to be particularly critical during mClr-g refolding. First, much higher yields of the soluble refolded monomeric protein were obtained in HEPES buffer (pH 7.5) compared with PIPES buffer (pH 6.8) (data not shown), suggesting a pH dependence of the proper folding. Second, the addition of nondetergent sulfobetaine 3-(benzyl(dimethylammonio)propanesulfonate (suggested by the commercial iFOLD2 kit) was beneficial for the yield of the correctly refolded protein, as observed by size-exclusion chromatography (SEC). This effect proved to be concentration dependent up to 0.3 M, at which point it reached a plateau (Fig. 1A).

Thus, this concentration of the sulfobetaine was used for large-scale protein preparations. The final purification of mClr-g was performed by SEC, providing protein of high purity when analyzed by SDS-PAGE (Fig. 1B) and FT-ICR MS. Mass spectrometric analysis of the entire protein confirmed correctness of the mutation, showed that all cysteines are in the oxidized form (two disulfide bridges), and provided detailed information on the disulfide arrangement. The first cysteine in the sequence [C0 according to the notation of cysteines usual for the long carbohydrate-recognition domain of the C-type lectin family (46)] was shown to be paired to the second cysteine (C0’) (cystic peptide TYAAC92SK-C103FYFSGYPR, M\(_{\text{th}}\) 1879.8146, M\(_{\text{exp}}\) 1879.8127). Similarly, the third cysteine (C1) was shown to be paired with the fourth cysteine (C4) based on the identification of cystic peptide YNFTVAFQIC120MAEQAILAR-MW1C202SK (M\(_{\text{th}}\) 2850.2988, M\(_{\text{exp}}\) 2850.2942).

### Oligomerization of mClr-g in solution

The molecular size of the extracellular domain of mClr-g in solution was estimated using several techniques. SEC revealed a single symmetrical peak of a size corresponding to a compactly folded monomer. Analytical ultracentrifugation also indicated the prevalence of a monomeric protein, although a minor fraction of dimeric and trimeric species could also be detected (Fig. 1C).

Dynamic light scattering provided a radius of mClr-g of 18 Å, which corresponds to the Stokes translational radius of 20 Å calculated for a monomer using HYDROPRO (47) and structure coordinates of mClr-g. However, under native conditions, mClr-g is anchored in the cell membrane, and its oligomeric state may differ significantly from that in solution. The dimer
The protein used for crystallization represents an almost complete overall structure of mClr-g because of its similarity to other biological dimers of the C-terminus. The measured crystallographic data have reliable statistics and the refined structure fulfills standard validation criteria (Table I).

mClr-g occurs in the crystal structure as a dimer (Fig. 2), which is also the content of the asymmetric unit of the crystal. All of the residues 85–206 of the extracellular part of the protein used in crystallization are localized in electron density. At the level of the secondary structure, mClr-g forms two α-helices, two anti-parallel β-sheets (each consisting of three β-strands), and two disulfide bonds (Cys92–Cys103 and Cys120–Cys202; Fig. 2). The fold closely follows that of hCD69, with the typical loop-in-loop arrangement of the CTL domains.

The dimer has 2-fold noncrystallographic symmetry, except for the N-termini (residues 85–90; chain B penetrating into a neighbor dimer, chain A in contact with surfaces of two neighboring dimers), the dimer interface (residues 141–145 and 106–108), and Tyr195 on the surface of the protein. The side chain of Gln126B is chemically modified at Oε1. There is a clear maximum present in electron density, representing an additional atom behind Oε1 (1.7 Å) and forming a hydrogen bond to water molecule 17B (3.0 Å). Hydroxylation of the Gln side chain is the most likely explanation for this feature, but the peak was left uninterpreted, because no additional evidence was available. No modification of the residue was apparent in MS analysis before crystallization. Three residues are modeled in alternative conformations: Met121A, Met121B, and His153B.

![Image](http://www.jimmunol.org/)

Table I. Data collection statistics and structure refinement parameters

<table>
<thead>
<tr>
<th>Data Collection Statistics</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wavelength (Å)</td>
<td>0.91841</td>
</tr>
<tr>
<td>No. of images</td>
<td>580</td>
</tr>
<tr>
<td>Crystal-detector distance (mm)</td>
<td>197.7</td>
</tr>
<tr>
<td>Exposure time/image (s)</td>
<td>25</td>
</tr>
<tr>
<td>Oscillation width (°)</td>
<td>0.25</td>
</tr>
<tr>
<td>Space group</td>
<td>P2₁₂₁₂</td>
</tr>
<tr>
<td>Unit-cell parameters a, b, c (Å)</td>
<td>118.3, 61.4, 32.4</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>50.0–1.95 (1.98–1.95)</td>
</tr>
<tr>
<td>No. of observations</td>
<td>756630</td>
</tr>
<tr>
<td>No. of unique reflections</td>
<td>17372 (802)</td>
</tr>
<tr>
<td>Data completeness (%)</td>
<td>96 (90)</td>
</tr>
<tr>
<td>Average redundancy</td>
<td>5.6 (4.6)</td>
</tr>
<tr>
<td>Mosaicity (°)</td>
<td>0.7</td>
</tr>
<tr>
<td>Average I0(I)</td>
<td>10.4 (4.5)</td>
</tr>
<tr>
<td>Solvent content (%)</td>
<td>39.7</td>
</tr>
<tr>
<td>Matthews coefficient (Å³/Da)</td>
<td>2.0</td>
</tr>
<tr>
<td>Rmerge&lt;sup&gt;+&lt;/sup&gt;</td>
<td>0.081 (0.306)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Structure Refinement Parameters</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>R&lt;sub&gt;work&lt;/sub&gt;&lt;sup&gt;b&lt;/sup&gt;</td>
<td>18.4</td>
</tr>
<tr>
<td>R&lt;sub&gt;free&lt;/sub&gt;</td>
<td>26.8</td>
</tr>
<tr>
<td>R&lt;sub&gt;all&lt;/sub&gt;</td>
<td>18.6</td>
</tr>
<tr>
<td>Average B-factor (Å&lt;sup&gt;2&lt;/sup&gt;)</td>
<td>14.3</td>
</tr>
<tr>
<td>RMSD bond lengths from ideal (Å)</td>
<td>0.014</td>
</tr>
<tr>
<td>RMSD bond angles from ideal (°)</td>
<td>1.34</td>
</tr>
<tr>
<td>No. of nonhydrogen atoms</td>
<td>2311</td>
</tr>
<tr>
<td>No. of monomers/asymmetric unit</td>
<td>2</td>
</tr>
<tr>
<td>No. of water molecules</td>
<td>256</td>
</tr>
<tr>
<td>Other localized moieties</td>
<td>1 × C1⁻</td>
</tr>
</tbody>
</table>

Values in parentheses refer to the highest resolution shell.

\[ R_{merge} = \left( \sum_{k,l} \sum_{hkl} \left| F_o(hkl) - F_c(hkl) \right|^2 / \sum_{k,l} \sum_{hkl} \left| F_o(hkl) \right|^2 \right)^{1/2} \]

are the observed and calculated structure-factor amplitudes for reflection with indices hkl, respectively, \( F_o(hkl) \) and \( F_c(hkl) \) are the observed and calculated structure-factor amplitudes for reflection with indices hkl, for working set of reflections. \( R_{free} \) is the same as \( R_{work} \) but for 5% of the data that were omitted from refinement. \( R_{merge} \) is the sum over all reflections.

---

observed in the crystal structure is the expected biological unit because of its similarity to other biological dimers of the C-type lectin family.

**Overall structure of mClr-g**

The protein used for crystallization represents an almost complete extracellular domain of mClr-g (14.4 kDa), covering residues 85–206 (UniProtKB/Swiss-Prot, Q9WVF9.1) with mutations I85M (initiation methionine) and C148S (mutation of an unpaired cysteine).

![Image](http://www.jimmunol.org/)
ion in the asymmetric unit (i.e., in the solvation shell of a dimer). Ninety-five percent of the residues lie in the preferred regions of the Ramachandran plot, 4% lie in the allowed regions, and there are two outliers: Lys$^87B$ and Ser$^{186B}$. All atoms of Lys$^87B$ are clearly localized in electron density, and the residue adjusts its conformation to enable binding of the N terminus of chain B into the neighboring dimer. Ser$^{186B}$ is involved in a tight turn and is well supported by electron density. Mutated C148S is clearly localized in electron density and located in the core of the protein.

Dimer interface

The surface area of the dimer is $12830 \text{Å}^2$, and the total buried surface area of the dimer is $1740 \text{Å}^2$. The estimated change in solvent-free energy due to dimerization is $-14.0 \text{kcal/mol}$. The dimer observed in the asymmetric unit is the expected form of quaternary structure (PISA 1.18) (48).

The contact between monomers involves 10 hydrogen bonds (PROTORP) (49). The core of the dimer interface is formed by hydrophobic residues - phenyl rings of Phe$^{138A}$, Phe$^{142B}$, and Phe$^{106A}$, and by Arg$^{141B}$, always from both chains. Interestingly, as a consequence of the change in the conformer of Phe$^{106B}$ between chains A and B, the peptide bond Ser$^{107B}$–Gly$^{108B}$ is flipped to remove the carbonyl oxygen from a position that would lead to a steric clash with the Phe$^{106A}$ side chain.

Intermolecular contact in the crystal: N terminus bound to the central pocket

The N terminus of chain B (residues 85–87) is inserted into a central pocket located at the dimeric interface between the chains of a neighboring dimer (Fig. 3). Hydrogen bonds between the N terminus and the pocket are summarized in Table II, and the binding mode is shown in Fig. 3A–C. Sulfur of the first residue Met$^{85B}$ has the closest contacts with Phe$^{142B} \text{C}^\beta$ (3.8 Å), Phe$^{138B} \text{C}^\alpha$ (3.7 Å), and Phe$^{138B} \text{C}^\beta$ (3.8 Å). The “tail pocket” interaction relies on hydrophobic contacts mainly of Met$^{85B}$ and on several main chain–main chain, side chain–side chain, and mixed hydrogen bonds. It should be noted that the arrangement of the first residues of the crystallized protein is not native; the protein corresponds only to the extracellular domain and, therefore, it is truncated artificially at its N terminus, and the terminal Met$^8$ is not present in the native sequence (initiation mutation I85M). This N terminus influences the assembly of two neighboring dimers and is a possible cause of the spontaneous crystallization of the protein.

Structure comparison, molecular modeling, and electrostatics calculations

Structural comparisons were performed with proteins summarized in Table III using the secondary structure matching algorithm.
Structures of mClr-b and mNKR-P1F were modeled, and electrostatic potentials of selected dimers of receptors and ligands were calculated (Fig. 4).

Discussion

**mClr-g and hCD69: structure comparison**

mClr-g and hCD69 show 36% sequence identity (ClustalW) (35). A root-mean-square deviation (RMSD) between the two monomers (chain B) is only 1.0 Å (113 residues, SSF in PDBeFold) (50). Superposition of 118 Cα atoms of chains A and B of mClr-g results in an RMSD of 0.6 Å (SSM in COOT) (38) and, similarly, in an RMSD of 0.9 Å for 118 Cα atoms of hCD69. Thus, the structural variance between the single chains of mClr-g and hCD69 is approximately on the same level as the structural difference between chains A and B in one dimer of mClr-g or hCD69. RMSD of Cα atoms of the complete dimers is much higher (2.4 Å, superposition of 229 residues “dimer to dimer,” hCD69, PDB code 3HUP) (34). This indicates that the global fold of a monomer main chain is preserved more rigorously than is the assembly of the dimer.

The N and C termini of both chains lie close to one another for both proteins. In both cases, continuation of the protein chains prior to the N termini of the extracellular part would enable formation of stalks directed away from the described extracellular domains toward a cell membrane where the proteins would be anchored by residues that were not included in the current protein-expression constructs. Therefore, our newly described structure of mClr-g fully fits with the current model of molecular architecture of the CTL NK cell receptors and ligands and confirms the particular dimerization pattern previously observed for hCD69 (26).

**Architecture of the dimers.** The difference in organization of the mClr-g and CD69 dimers (PDB 3HUP) can be described by the rotation required to superimpose one chain of a dimer onto the other. Chains A and B of the hCD69 dimer are related by a rotation of 179.97°, whereas this rotation angle equals 179.1° in mClr-g.

The details become clear when only halves of the dimers are superimposed (chains A of hCD69 and mClr-g, Fig. 3D). Then chain B of one protein must be rotated by ~25° to overlap with the other, and the direction of the axis of this rotation is very close to the direction perpendicular to the “plane” of the dimer interface (if a roughly planar interface surface is assumed).

A detailed analysis does not reveal a simple relationship between this structural difference and sequence differences. Moreover, slight differences in the orientation of monomers in such a dimer are not found between only mClr-g and hCD69. Various structures of hCD69 differ from each other: CD69 in a tetragonal lattice (PDB 1E8I) and in a trigonal lattice (PDB 1E87) are more similar to mClr-g than is CD69 in a hexagonal lattice (PDB 3HUP), and the Cα trace of chain B of 1E87 and 1E81 is roughly in the middle between 3HUP and mClr-g when chains A of the structures are superimposed. A different orientation of subunits within dimers of proteins with a CTL fold was also described when comparing the heterodimer CD94-NKG2A with Ly49A (51). Thus, these observations suggest that this feature corresponds to a real dimer flexibility that is also observed for other CTL proteins.

**Dimer interface.** The dimer interface features of mClr-g and hCD69 are strikingly similar. Both molecules have a central hydrophobic core with similar features and level of asymmetry. The positions of the central residues Phe138 and Phe142 correspond to residues Phe131 and Tyr135 in hCD69. Similarly to Tyr135 in CD69, residue Phe142 in mClr-g takes different rotamers in chains A and B. Phe106, corresponding to Ile29 in hCD69, also adopts different rotamers. Smaller differences between chains A and B are also observed for other residues of the interface.

**N terminus binding to the central pocket.** The N terminus binding to the central pocket of the dimer was also observed for CD69 (25) (PDB code 1FM5). A tripeptide bound to the central pocket is interpreted as N-terminal triad Ser78, Asp79, and Ser80, whereas the next residues in the chain are not localized, and the chain continues, starting with residue Cys85. Two structures raise the question about a possible functional peptide binding to the central pocket of Clr-g and CD69. A different type of interaction between hCD69 and an immunogenic hsp65-derived peptide was observed experimentally. This interaction is charge based and is due to glutamic acid residues of the peptide ligand interacting with hCD69 lysines (52).

**mClr-g and mNKR-P1A: structure comparison**

The structure of the extracellular domain of mouse activation receptor NKR-P1A was determined in our previous study (24). Sequence comparison of mClr-g and mNKR-P1A is shown in Supplemental Fig. 1. A total of 94 of 124 aa residues of one chain of mNKR-P1A can be superimposed, by the SSM algorithm (implemented in COOT) (38), with chain A of mClr-g with an RMSD of Cα atoms of 1.4 Å. The region of 30 residues missing from this set of superimposable residues is the extended loop of mNKR-P1A (161–187). mClr-g follows the same fold as in the case of hCD69 in the compact part of the CTL domain (residues 90–153 and 181–206). Most of the large surface loop of mClr-g (the extended loop in mNKR-P1A) also folds in the conformation of hCD69, with the exception of a small difference in residues 163–168 and a larger one for the “tip” of loop 155–160. The former region differs primarily because Val122A of hCD69 is replaced by a larger Arg129A in mClr-g in a neighboring chain.
The latter difference is a consequence of Pro\textsuperscript{149}A of hCD69 being replaced by a more flexible Ser\textsuperscript{156}A in mClr-g and an insertion with respect to hCD69: Ser\textsuperscript{157} in mClr-g.

Electrostatics of mClr-g and other CTL proteins

In addition to the already-discussed hCD69 and mNKR-P1A, the following proteins with known three-dimensional (3D) structure, and similar to mClr-g, are discussed (references in Table III): NK receptors human CD94 (hCD94); human NGK2D (hNGK2D); human NKG2A; KLRG1; mouse LY49L4 (mLY49L4); mouse LY49A (mLY49A); CD207 (langerin), a type II transmembrane cell surface receptor produced by Langerhans cells; and hLOX-1, a human receptor encoded by the OLR1 gene, which binds and degrades oxidized low-density lipoprotein. Table III is sorted by the level of structural agreement of dimers of these proteins with mClr-g.

Additionally, two structures were modeled using homology modeling: mClr-b, a mouse ligand for inhibitory receptors, and mNKR-P1F, the binding partner of mClr-g (for sequence alignments see Supplemental Fig. 1).

To help explain the structure and function of mClr-g and its similarity or complementarity to other CTL receptors and ligands, we investigated their electrostatics. Fig. 4 shows the electrostatic equipotentials (based on solution of the Poisson–Boltzmann equation in APBS) (42) of these CTL proteins (Table III) with the exception of human CD207 and human KLRG1, and with the addition of the models of mNKR-P1F and mClr-b. Electrostatic equipotentials (surfaces with a constant value of electrostatic potential) are shown instead of the more common values of electrostatic field on the surface of the protein. This representation suitably shows dominant electrostatic features of a molecule, as they appear from the point of view of an approaching binding partner.

It is apparent that, despite similar folds, the molecules studied have highly diverse distributions of electrostatic potential. mClr-g has a predominantly positive potential in its surroundings, with the exception of the central pocket, which is negative. The electrostatics of CD69, the most similar protein, has an opposite character: it is negative and has positive potential in the central pocket. This may imply different roles for mClr-g and hCD69, despite their high similarity in sequence and structure. mClr-b, a ligand for inhibitory receptors, has a featureless electrostatic potential, contrary to the polar potential of the activation receptor ligand mClr-g. mNKR-P1F has negative potential in the middle of the receptor. Thus, the electrostatic potentials of mNKR-P1F and mClr-g, its binding partner, appear complementary. This would suggest the basis of their interaction and the contact driving force.

Table III. Summary of CTL proteins (with known 3D structure) most similar to mClr-g

<table>
<thead>
<tr>
<th>Protein</th>
<th>Reference</th>
<th>PDB Code</th>
<th>Sequence Identity (No. of Overlapping Residues)</th>
<th>RMSD Monomer (Å; No. of Overlapping Residues)</th>
<th>RMSD Dimer (Å; No. of Overlapping Residues)</th>
</tr>
</thead>
<tbody>
<tr>
<td>hCD69</td>
<td>(26)</td>
<td>1E8I</td>
<td>38% (118)</td>
<td>1.13 (116)</td>
<td>1.75 (228)</td>
</tr>
<tr>
<td>hCD69</td>
<td>(34)</td>
<td>3HUP</td>
<td>36% (130)</td>
<td>1.22 (115)</td>
<td>2.42 (229)</td>
</tr>
<tr>
<td>Human NKG2A-CD94, heterodimer bound to HLA-E</td>
<td>(51)</td>
<td>3BDW</td>
<td>NKG2A 32% (80)</td>
<td>2.03 (108)</td>
<td>2.41 (209)</td>
</tr>
<tr>
<td>hCD94</td>
<td>(55)</td>
<td>1B6E</td>
<td>16% (128)</td>
<td>1.79 (104)</td>
<td>2.77 (200)</td>
</tr>
<tr>
<td>hLOX-1</td>
<td>(56)</td>
<td>1YPO</td>
<td>23% (132)</td>
<td>1.32 (116)</td>
<td>2.84 (217)</td>
</tr>
<tr>
<td>hNGK2D</td>
<td>(57)</td>
<td>1MPU</td>
<td>20% (138)</td>
<td>1.73 (108)</td>
<td>3.05 (208)</td>
</tr>
<tr>
<td>mL494A</td>
<td>(58)</td>
<td>1OQ3</td>
<td>4% (146)</td>
<td>1.88 (108)</td>
<td>3.40 (189)</td>
</tr>
<tr>
<td>mL49L4</td>
<td>(59)</td>
<td>3G8K</td>
<td>28% (130)</td>
<td>2.00 (108)</td>
<td>3.41 (197)</td>
</tr>
<tr>
<td>Human CD207 (langerin)</td>
<td>(60)</td>
<td>3OQG</td>
<td>21% (182)</td>
<td>1.39 (111)</td>
<td>Different configuration</td>
</tr>
<tr>
<td>Human KLRG1 bound to E-cadherin</td>
<td>(61)</td>
<td>3FF7</td>
<td>6% (100)</td>
<td>1.63 (107)</td>
<td>Different configuration</td>
</tr>
<tr>
<td>mNKR-P1A</td>
<td>(24)</td>
<td>3M9Z</td>
<td>27% (139)</td>
<td>1.44 (94)</td>
<td>Different configuration</td>
</tr>
</tbody>
</table>

Overview of their similarity to mClr-g, from the point of view of a sequence, monomer structure, and dimer structure. Sequence identity with mClr-g was computed in ClustalW (pair-wise alignment), and structure superposition with mClr-g was performed using EBI-SSM. The structures are ordered according to dimer RMSD with respect to mClr-g.

FIGURE 4. Electrostatic potential (equipotential surfaces at levels 3kT [blue] and $\pm$3kT [red]) for mClr-g and seven dimeric proteins with known x-ray structure ordered according to the level of structural similarity (dimer RMSD) to the mClr-g dimer: hCD69 (PDB code 3HUP), hCD94-NKG2A, hCD94, hLOX-1, hNGK2D, mL49L4, and mL49A. For PDB codes and level of similarity see Table III. Additionally, mClr-g was used as a template for homology modeling of mClr-b, a mouse ligand for inhibitory receptors, and mNKR-P1F, the binding partner of mClr-g (for sequence alignments see Supplemental Fig. 1).

To help explain the structure and function of mClr-g and its similarity or complementarity to other CTL receptors and ligands, we investigated their electrostatics. Fig. 4 shows the electrostatic equipotentials (based on solution of the Poisson–Boltzmann equation in APBS) (42) of these CTL proteins (Table III) with the exception of human CD207 and human KLRG1, and with the addition of the models of mNKR-P1F and mClr-b. Electrostatic equipotentials (surfaces with a constant value of electrostatic potential) are shown instead of the more common values of electrostatic field on the surface of the protein. This representation suitably shows dominant electrostatic features of a molecule, as they appear from the point of view of an approaching binding partner.

It is apparent that, despite similar folds, the molecules studied have highly diverse distributions of electrostatic potential. mClr-g has a predominantly positive potential in its surroundings, with the exception of the central pocket, which is negative. The electrostatics of CD69, the most similar protein, has an opposite character: it is negative and has positive potential in the central pocket. This may imply different roles for mClr-g and hCD69. Despite their high similarity in sequence and structure, mClr-b, a ligand for inhibitory receptors, has a featureless electrostatic potential, contrary to the polar potential of the activation receptor ligand mClr-g. mNKR-P1F has negative potential in the middle of the receptor. Thus, the electrostatic potentials of mNKR-P1F and mClr-g, its binding partner, appear complementary. This would suggest the basis of their interaction and the contact driving force.
In the calculated distributions of electrostatic potential, negative potential near the surface of NKR-P1A (Fig. 4) is mostly due to two charged residues the surface of Glu\textsuperscript{162} and Asp\textsuperscript{124} and their symmetry mates from the other chain. Interestingly, these negative residues are perfectly conserved throughout the family of NKR-P1 receptors [Fig. A2 of the supplementary material of Kolenko et al. (24)] whereas hCD69 and mClr-b proteins are excluded from this pattern (Ser\textsuperscript{192} and Asn\textsuperscript{197} in mClr-g).

**mClr-g-mNKR-P1F complex formation**

An interaction between mClr-g and mNKR-P1F was suggested based on studies using transfected cells (21). Any data on the atomic or molecular level are expected to provide better insight into this matter. Structural complementarity between the Clr and NKR-P1 proteins cannot be established based on the current structural knowledge. An unexplained mechanism must govern the formation of these complexes and consequent signaling events. The electrostatic profile provides one of the possible ways to explain this mechanism. The negative region in the central area of mNKR-P1F (Fig. 4) would complement the positive parts of the interacting partner mClr-g. Based purely on the distribution of the potentials, the interaction between mNKR-P1F and Clr-g would likely prefer the positive tip of one chain of the mClr-g dimer binding to the central part of dimeric mNKR-P1F.

We modeled the complex of mClr-g with mNKR-P1F using the RosettaDock protein–protein docking server (45). Three models of interaction were tested: interaction of the extended loop of mNKR-P1F placed into the pocket of mClr-g, mimicking binding of extended mClr-g N terminus to the neighboring molecule, as observed in the crystal structure of mClr-g; positioning of mClr-g between the extended loops of mNKR-P1F so that the loops embrace a monomer of the mClr-g dimer; and the loops of mNKR-P1F embracing the mClr-g dimer in the middle. The second model gives the most stable results and stands as the most likely mode (Supplemental Fig. 2).

It should be mentioned that the model of mNKR-P1F is based on the structure of mNKR-P1A with extended loops 161–187. There were observations of a refolded monomer of mNKR-P1A in solution with a compactly folded loop (53); however, our model is based on the only 3D structure available. Nevertheless, a different orientation of the loops does not change the basic properties of electrostatics of NKR-P1F (results based on modeling; data not shown). No experimental information on the conformation of the loops of mNKR-P1F in the complex with mClr-g is available.

The model is consistent with the electrostatic expectation of binding one chain of the mClr-g dimer to the central part of an mNKR-P1F dimer and confirms the spatial feasibility of such an interaction. This hypothetical binding mode would enable interaction of the second chain of the mClr-g dimer with another receptor dimer. Such interactions would be in agreement with the idea that NK receptors bind to their Clr ligands in a chain-like manner (zipper-type complexes).

Recent work on the human NKR-P1A/LLT1 interaction pair (23) identified the importance of residues that would also play a role in the interaction mode discussed in this article. The positive potential areas of Clr-g are determined by the presence of five arginine residues (Arg\textsuperscript{154} corresponds to Arg\textsuperscript{124} in LLT1, Arg\textsuperscript{180} corresponds to Glu\textsuperscript{162} in LLT1, Arg\textsuperscript{198} to Lys\textsuperscript{180}, Arg\textsuperscript{193} to Arg\textsuperscript{175}, and Arg\textsuperscript{198} to Arg\textsuperscript{180}). Point mutations of three of these basic residues in LLT1 (Lys\textsuperscript{180}, Arg\textsuperscript{175}, and Arg\textsuperscript{180}) always led to loss of detectable interaction with hNKR-P1A, and our results indicate the key role of the corresponding residues in the interaction pair NKR-P1F/Citr-g as well. A multiple-sequence alignment of the individual mouse Clr isoforms clearly indicates that three of five Arg residues mentioned above are absolutely conserved within this mClr subfamily (Supplemental Fig. 3). These would be expected to interact with mNKR-P1F and mNKR-P1G, because recent data strongly indicate a broader reactivity of mNKR-P1F than originally reported (8), possibly involving not only mClr-g but also mClr-d/x (18) or even mClr-c and mClr-f (16).

Two approaches led to conclusions about the type of interactions in two different CTL receptor/ligand interaction pairs: one based on mutagenesis and surface plasmon resonance measurements (23) and the currently presented one based on x-ray structure and modeling. The results of both are generally in agreement; the membrane distal surface of the receptor or ligand is involved in the interaction, with some key residues reoccurring and also conserved in Clr proteins. Currently, a somewhat different orientation of the interacting molecules is expected for NKR-P1F/Citr-g, which was discussed in detail in this article. Structural information on complex formation between Clr molecules and NKR-P1 receptors observed experimentally would provide the ultimate answer to these questions.

**Acknowledgments**

We thank Uwe Müller for support at the beam line BL14.1 of Bessy II, Helmholtz-Zentrum Berlin, and the Institute of Biotechnology, Academy of Sciences of the Czech Republic, for access to the x-ray diffractiuon suite.

**Disclosures**

The authors have no financial conflicts of interest.

**References**


Mouse Clr-g, a ligand for NK cell activation receptor NKR-P1F: crystal structure and biophysical properties

Tereza Skálová, Kristýna Kotýnková, Jarmila Dušková, Jindřich Hašek, Tomáš Kovařík, Petr Kolenko, Petr Novák, Petr Man, Pavel Hanč, Ondřej Vaněk, Karel Bezouška, and Jan Dohnálek
SUPPLEMENTAL FIGURE 1. (A) Structure-based sequence alignment of the extracellular part of mClr-g, hCD69 and mNKR-P1A. Secondary structure of mClr-g is denoted. The upper numbering corresponds to mClr-g and the lower one to mNKR-P1A. The secondary-structure matching parts are in cyan, the extended loop of mNKR-P1A is in green. Disulfide bonds are denoted by stars with color marking bonded pairs (green and blue in all three proteins, red in hCD69 and mNKR-P1A only). Secondary structure was assigned with DSSP (Kabsch and Sander, 1983, *Biopolymers* 22(12): 2577-2637). Secondary structure matching was performed by EBI-SSM. (B) Sequence alignment of the extracellular part of mClr-g with mClr-b: a basis for homology modelling. Prepared using CLUSTALW (35). (c) Sequence alignment of the extracellular part of mNKR-P1A with mNKR-P1F: a basis for homology modelling. Prepared using CLUSTALW. The figure was prepared using ESPRIPT (Gouet, P., Courcelle, E., Stuart, D. I., and F. Metoz. 1999. ESPript: multiple sequence alignments in PostScript. *Bioinformatics* 15: 305-308.).
SUPPLEMENTAL FIGURE 2. A model of the mouse NKR-P1F:Clr-g complex, respecting the electrostatic complementarity of both molecules, in molecular surface representation. The model was created by the RosettaDock protein-protein docking server using input model no. 2 as explained in the manuscript (position of mClr-g is expected between the extended loops of mNKR-P1F so that the loops embrace a monomer of the mClr-g dimer). Dimers of the extracellular domains of the receptor and ligand are used in the model. The individual protein chains within the dimers are differentiated by color hue.
SUPPLEMENTAL FIGURE 3. (A) Bioinformatic overview of mClr-g using UniProtKB/Swiss-Prot file Q9WVF9.1. This particular splice variant of mClr-g is 217 amino acids long, and is composed of intracellular N-terminal fragment (Met\textsuperscript{1} – Lys\textsuperscript{53}, \textit{in italics}), the transmembrane domain (Leu\textsuperscript{54} – Ala\textsuperscript{74}, \textcolor{purple}{purple}), and the large C-terminal extracellular portion (Leu\textsuperscript{75} – Val\textsuperscript{217}) that includes the CTL domain (Cys\textsuperscript{92} – Lys\textsuperscript{204}, \textcolor{blue}{underlined}) stabilized by two predicted canonical disulfide bonds formed by Cys\textsuperscript{92} and Cys\textsuperscript{103}, and Cys\textsuperscript{120} and Cys\textsuperscript{202} (\textcolor{yellow}{in yellow}). The unpaired Cys\textsuperscript{148} within the CTL domain is shown in \textcolor{green}{green}, and the C-terminal Cys\textsuperscript{212} predicted to be responsible for covalent dimerization of the receptor in \textcolor{red}{red}. The expression construct used in this study is marked in \textcolor{lightblue}{light blue}. Putative N-glycosylation sites at Asn\textsuperscript{95} and Asn\textsuperscript{112} are marked in \textcolor{darkblue}{dark blue}.

(B) Multiple sequence alignment of mNKR-P1F reactive mClr isoforms mClr-g (accession Q9WVF9.1), mClr-c (accession AEP22558.1), mClr-d/x (accession NP_081838.1), and mClr-f (accession Q8C1T8.1). Cysteine residues forming the canonical disulfide bonds are shown in \textcolor{yellow}{yellow}, those supposed to be involved in covalent receptor dimerization in \textcolor{purple}{purple}, and additional cysteines are in \textcolor{darkblue}{dark blue} (including Cys\textsuperscript{148} mutated to Ser in the expressed protein, the extent of which is indicated in \textcolor{lightblue}{light blue} and includes the initiation Met\textsuperscript{85} replacing Ile shown in \textcolor{red}{red}). Arginines Arg\textsuperscript{154}, 180, 188, 193, and 198 predicted to be involved in mNKR-P1F interaction are shown in \textcolor{green}{green}, and additional predicted interaction residues are marked with asterisks.