Chemokine receptors (CKRs) function in the inflammatory response and in vertebrate homeostasis. Decay and viral receptors are two types of CKR homologs with modified functions from those of the typical CKRs. The decoy receptors are able to bind ligands without signaling. On the other hand, the viral receptors show constitutive signaling without ligands. We examined the sites related to the functional difference. At first, the decoy and viral receptors were each classified into five groups, based on the molecular phylogenetic analysis. A multiple amino acid sequence alignment between each group and the CKRs was then constructed. The difference in the amino acid composition between the group and the CKRs was evaluated as the Kullback–Leibler (KL) information value at each alignment site. The KL information value is considered to reflect the difference in the functional constraints at the site. The sites with the top 5% of KL information values were selected and mapped on the structure of a CKR. The comparisons with decoy receptor groups revealed that the detected sites were biased on the intracellular side. In contrast, the sites detected from the comparisons with viral receptor groups were found on both the extracellular and intracellular sides. More sites were found in the ligand binding pocket in the analyses of the viral receptor groups, as compared to the decoy receptor groups. Some of the detected sites were located in the GPCR motifs. For example, the DRY motif of the decoy receptors was often degraded, although the motif of the viral receptors was basically conserved. The observations for the viral receptor groups suggested that the constraints in the pocket region are loose and that the sites on the intracellular side are different from those for the decoy receptors, which may be related to the constitutive signaling activity of the viral receptors.

Keywords: chemokine receptors, decoy receptors, viral receptors, GPCR, molecular evolution

INTRODUCTION

The members of the chemokine (CK) family play important roles in regulating cell migration against inflammation, immune surveillance, and oncogenesis in vertebrates (Zlotnik and Yoshie, 2000). The CKs are classified into four subfamilies: CC, CXC, CX3C, and XC, based on the cysteine positions in their motifs (Zlotnik and Yoshie, 2000). CKs exert their activities through binding to their corresponding receptors. Presently, more than 40 CKs and 18 chemokine receptors (CKRs) have been identified in the human genome: 10 CCRs, six CXCRs, one XCR, and one CX3CR (Nomiyama et al., 2011). The CKR homologs are widely distributed among the vertebrate genomes. For example, homologs have even been identified from sea lampreys, which are one of the most primitive vertebrates (Nomiyama et al., 2011). The amino acid sequence identities among the CKRs and the homologs range from 25 to 80%, and the CKRs constitute a subfamily in the class A G protein-coupled receptors (GPCRs). The CKRs have broad ligand specificities (Nomiyama et al., 2011), and each receptor is able to interact with several CKs, and vice versa. This binding promiscuity makes it difficult to develop drugs to pinpoint the specific function of each CKR. Among these receptors, only the structure of CXCR4 has been solved by X-ray crystallography (Wu et al., 2010). Like the otherGPCRs, this structure is characterized by the seven transmembrane (TM) helices, although T4 lysozyme was inserted within the intracellular loop (ICL) 3 between the TM helices 5 and 6, to stabilize the crystal. The extracellular cavity of CXCR4 is reportedly larger and wide open, as compared to those of other GPCR structures (Wu et al., 2010).

In addition to the traditional CKRs, five non-signaling CKR homologs have been identified in vertebrate genomes: CCR1 (also known as CCX-CKR), CCRL2, CCBP2 (D6), CXCR7, and DARC (Duffy antigen receptor; Graham, 2009; Leick et al., 2010; Naumann et al., 2010). They are called “decoy” or “silent” receptors, because they are able to bind to several CKs without ligand-induced signaling. Most of them are constitutively internalized with or without ligands, and only the receptors are recycled to the cell membrane. Their functions are considered to regulate inflammatory responses by controlling the volume of free extracellular CKs, through internalization and degradation (Bonecchi et al., 2010). Like the traditional CKRs, these decoy receptors show a
broad CK-binding spectrum. CCRL1 interacts with several homeostatic CC-type CKs (Comerford et al., 2006), whereas CCBP2 and DARC interact with inflammatory CKs (Graham, 2009). CXCR7 interacts with the dual-functional CXC-type CKs (Naumann et al., 2010) without activating G proteins (Thelen and Thelen, 2008). CCRL2 is known to be a multifunctional receptor (Yoshimura and Oppenheim, 2011). Like other decoy receptors, it regulates the amount of free CKs. At the same time, it functions as a receptor for an adipokine called chemerin, although the ligand binding does not induce signaling and the receptor is not internalized even after ligand engagement. DARC is the most distantly related to the CKRs among the five decoy receptors, and was originally identified as a malarial parasite receptor (Bonecchi et al., 2010). The receptor also binds to the CC- and CXC-type inflammatory CKs.

Chemokine receptors homologs have been detected in double stranded DNA viruses, such as herpesvirus and poxvirus. These viruses are considered to have gained these proteins by horizontal gene transfer during the course of evolution (Slinger et al., 2011). The viral receptors are constitutively active without ligands, although some of them can bind to CKs. We studied five groups of viral proteins as described below. E1 is derived only from equid herpesvirus 2 of γ-herpesvirinae, which interacts with CCL11 (Camarda et al., 1999). ORF74 is derived from several γ-herpesviruses, and interacts with a broad range of CXC-type CKs (Maussang et al., 2009). The β-herpesviruses also have several CKR homologs. Among them, UL33 is encoded by the genomes of various vertebrate viruses, although its ligands have not been identified (Grujthuijsen et al., 2002). On the other hand, the US27, US28, and vGPCRs, which share high sequence similarity, have only been identified in the primate β-herpesviruses (Sahagun-Ruiz et al., 2004). US28 is characterized as a receptor for CC-type ligands (Maussang et al., 2009). Several poxviruses, such as capripox virus, deerpox virus, and yatapox virus, also encode CKR homologs in their genomes. The receptors of poxviruses not only share high amino acid sequence similarity to CCRL8, but also the CCR8-like CK-binding profile; that is, high affinity to CCL1 (Najarro et al., 2009). These viral receptors are considered to contribute to the escape from and/or the perturbation of the host immune system, and are involved in inflammatory diseases and cancer (Slinger et al., 2011), although the mechanisms of these receptors in viral pathogenesis remain poorly understood.

The CKRs and their homologs have been classified into three functionally different types, from the viewpoints of ligand binding and signaling. The traditional CKRs bind their ligands, which induce signal transduction. The decoy receptors also bind ligands, although the process does not induce signal transduction. In contrast, the viral receptors exhibit signaling activity without ligand binding. The decoy receptors and the viral receptors are considered to have functionally differentiated after their divergence from the traditional CKRs, by gene duplication or horizontal gene transfer. Therefore, the functional differentiation of these three types is expected to have changed the functional constraints at the amino acid sites responsible for the ligand binding and/or signaling. If the sites involved in the functional differentiation can be identified, then the information about the sites would be helpful to understand the mechanisms for the signaling associated with ligand-induced conformational changes. Several different methods have been developed to detect the amino acid sites involved in the functional differentiation of homologous proteins from a multiple sequence alignment, and they are roughly classified into two types. One of them examines the difference in the evolutionary rate at each alignment site among the proteins with different functions (Gü, 1999; Simon et al., 2002), while the other compares the amino acid compositions at each alignment site among the proteins with different functions (Landgraf et al., 1999; Hannenhalli and Russell, 2000). We applied the latter method, the comparison of amino acid compositions, to identify the sites involved in the functional differentiation of CKR homologs. The difference in the amino acid composition at each alignment site between two groups (CKRs and decoy receptors, or CKRs and viral receptors) was calculated as the Kulback–Leibler (KL) information value (Hannenhalli and Russell, 2000; Ichihara et al., 2004). The sites with large KL information values were selected as the candidates for the functional differentiation. The amino acid residues corresponding to the selected sites were mapped on the tertiary structure of CXCR4. The comparison of the CKRs and decoy receptors revealed that the sites with large KL information values were concentrated on the cytosolic side of the CKR structure, with statistical significance. In contrast, there was no such bias in the distribution of the sites with large KL values between the CKRs and viral receptors. Based on the detected sites and the distribution of the corresponding residues on the tertiary structure, the underlying mechanisms for the functional divergence of the CKR homologs will be discussed.

**MATERIALS AND METHODS**

**AMINO ACID SEQUENCE DATA**

The amino acid sequences of the CKRs and their homologs, including decoy receptors and viral receptors, were collected by searching the non-redundant protein sequence database at the NCBI site with BLAST version 2.2.25 (Altschul et al., 1997). The amino acid sequence of human CXCR4 (GI number of NCBI: 1705894) was used as the query for the BLAST search. The sequence similarity search was also performed against the Ensembl\(^{2}\) and elephant shark genome project\(^{3}\) genome databases. When several amino acid sequences were almost identical, one of them was selected as the representative. The sequences used in this study are shown in Table 1.

**AMINO ACID SEQUENCE ALIGNMENT AND PHYLOGENETIC ANALYSIS**

A multiple amino acid sequence alignment was produced with the alignment software MAFFT, version 6.857 (Katoh and Toh, 2008). At first, 444 traditional CKRs were aligned. This result was manually refined, based on information about the secondary structures. Subsequently, 178 sequences consisting of decoy and viral receptors were added to the CKR alignment one by one, using the profile option of Clustal W (version 1.83; Thompson et al., 1994). Based on the alignment, an unrooted molecular phylogenetic tree was constructed by the neighbor-joining (NJ) method (Saitou and Nei, 1987). The genetic distance between

---

1http://www.ncbi.nlm.nih.gov/BLAST/
2http://www.ensembl.org/index.html
3http://esharkgenome.imcb.a-star.edu.sg/
## Table 1 | The sequences of the CKRs, decoy receptors, and viral receptors.

| CKRs          | Sequence 1 | Sequence 2 | Sequence 3 | Sequence 4 | Sequence 5 | Sequence 6 | Sequence 7 | Sequence 8 | Sequence 9 | Sequence 10 | Sequence 11 | Sequence 12 | Sequence 13 | Sequence 14 | Sequence 15 | Sequence 16 |
|---------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| CCR1          | 14546498   | 32221578   | 297206879  | 3023506   | 85718627   | 118150789  | 48675909   | 283837817  | 19422405  |
| CCR2 AND CCR5 | 10240454   | 28134358   | 209863062  | 84370370  | 126341460  |
| CCR3          | 1705896    | 19043302   | 2851566    | 11020491  | 10120491   | 281343586  | 209863082  | 84370370   | 126341640  |
| CCR4          | 1705894    | 297206879  | 3023506    | 85718627  | 118150789  | 48675909   | 283837817  | 19422405  |
| CCR5          | 1352335    | 1354236    | 41054914   | 296223926 | 281343586  | 209863082  | 84370370   | 126341460  |
| CCR6          | 332235484  | 332435368  | 297206879  | 3023506   | 85718627   | 118150789  | 48675909   | 283837817  |
| CCR7          | 1705892    | 149288986  | 205830369  | 26314164  | 6831505    | 53976357   | 28134587   | 55976357   | 281343586  | 209863084  | 48675903   | 30222943   |
| CCR8          | 1705894    | 297206879  | 3023506    | 85718627  | 118150789  | 48675909   | 283837817  | 19422405  |
| CCR9          | 1705894    | 297206879  | 3023506    | 85718627  | 118150789  | 48675909   | 283837817  | 19422405  |
| CXCR1 AND CXCR2 | 108936015 | 2494362   | 2494362   | 11029570  | 11029570   | 11029570   | 11029570   | 11029570   | 11029570   | 11029570   | 11029570   | 11029570   | 11029570   | 11029570   | 11029570   |

(Continued)
| Table 1 | Continued |
|--------|------------|
| **CXCR3** | 2829400 22253776 33226565 297710303 14976513 31127645 147966731 75072906 75072909 14976513 75072906 298327759 291407697 |
| **CXCR4** | 46577576 3913205 3023448 114152796 7507 4809 3023451 75072692 46577575 75072692 169154030 58272235 58272236 149730555 114149257 2494971 2494970 149637056 327260636 |
| **CXCR5** | 416718 311264026 29141297 473955058 301788472 291173052 416719 461630 297690401 332837885 332837885 |
| **CXCR6** | 312181 29622502 29141297 73956058 301788472 291781352 416719 461630 297690401 332837885 332837885 |
| **CX3CR1** | 1351394 29671678 332215591 226324297 100041508 29622841 28135282 73900295 122 1236266 149729043 |
| **CCR1** | 1145984 291383957 2988214 15782209 120341368 149632305 326923007 113963667 22404500 327260636 |
| **CCL1** | 12416046 55621142 29761939 100049361 292629075 149221508 147966731 30171760 73990094 544463 |
| **CCL2** | 1145984 291383957 2988214 15782209 120341368 149632305 326923007 113963667 22404500 327260636 |
| **CCBP2** | 20459469 14560363 29871505 75075026 296225035 48679505 115493626 1322063 30178046 73985695 281313590 |
| **CCR7** | 11501395 10888289 297671505 7507506 296225035 48679505 115493626 1322063 30178046 73985695 281313590 |
| **DARC** | 67416970 27734275 29966360 22734274 299623939 29397675 293341477 27734283 149755929 160332362 |
| **Viral receptors** | 124736385 124736389 124736381 9628003 124736365 124736377 124736369 124736395 124736381 124736383 124736379 |

Decoy receptors

| **CCRL1** | 124736385 124736389 124736381 9628003 124736365 124736377 124736369 124736395 124736381 124736383 124736379 |
| **CCRL2** | 124736385 124736389 124736381 9628003 124736365 124736377 124736369 124736395 124736381 124736383 124736379 |
| **CCBP2** | 124736385 124736389 124736381 9628003 124736365 124736377 124736369 124736395 124736381 124736383 124736379 |

Viral receptors

(Continued)
The sequences obtained from the NCBI database are indicated by the GI numbers. The sequences with names starting with AAVX were obtained from the elephant shark genome project database, while the names stating with ENS indicate sequences obtained from the Ensembl database.

The sequences obtained from the NCBI database are indicated by the GI numbers. The sequences with names starting with AAVX were obtained from the elephant shark genome project database, while the names stating with ENS indicate sequences obtained from the Ensembl database.

**CALCULATION OF THE KULLBACK–LEIBLER INFORMATION VALUE**

The multiple alignment thus obtained was reconstructed into 10 alignments, each consisting of two groups, the traditional CKRs and one of the decoy receptor groups or viral receptor groups. We then calculated the amino acid compositions of the two groups at each alignment site, according to the multiple alignment. We used the method adopted in PSI-BLAST (Altschul et al., 1997) to estimate the site-specific amino acid composition. The weighting method of Henikoff and Henikoff (1994) was used for the residue count. The weight for the pseudocount \( \beta \), was set to 0.1. For the calculation of the pseudocount, \( \lambda_u \), a parameter for ungapped BLAST, was calculated for each alignment by the Newton–Raphson method (Ewens and Grant, 2001). When more than half of the sequences had gaps at an alignment site, the calculation of the site-specific amino acid composition and the following investigation were skipped. Next, the difference in the amino acid composition between two groups at each alignment site was calculated as the KL information value. The KL information value is defined as follows:

\[
\sum_{i=1}^{20} p(i) \log \frac{p(i)}{q(i)}
\]

where \( p \) and \( q \) are the site-specific amino acid residue compositions for the two groups, which are estimated by the method described above. The parameter \( i \) indicates that the summation is obtained over 20 amino acid residues. KL information does not satisfy one of the distance axioms, symmetry. To satisfy this condition, the KL information was modified as follows:

\[
\sum_{i=1}^{20} p(i) \log \frac{p(i)}{q(i)} + \sum_{i=1}^{20} q(i) \log \frac{q(i)}{p(i)}
\]

Formula 2 was used to predict the sites subjected to different functional constraints between the two groups. In this study, the functional constraint at a site of a protein sequence is defined as the extent of intolerance to mutation at the site, due to a reduction of the protein function by the mutation. This is a special case of the cumulative relative entropy developed by Hannenhalli and Russell (2000), which is applicable to an alignment consisting of multiple groups. When the KL information value of an alignment site was located in the top 5% of the distribution of KL information values for all of the sites, the site was regarded as a site under different constraints between the groups (Ichihara et al., 2004).
Among them, the sites that fell in the gap region of CXCR4 in the alignment were neglected, because the subsequent analyses were performed based on mapping onto the CXCR4 structure.

**STATISTICAL EVALUATION FOR BIAS IN THE SPATIAL DISTRIBUTION OF THE SITES UNDER DIFFERENT CONSTRAINTS**

We examined the statistical significance for the bias in the positions of the selected sites by the KL information values on the reference CXCR4 structure (PDB ID: 3ODU), using the following procedure. At first, we calculated the geometric center of the three extracellular loops (ECLs) and the N-terminal region, and that of the three ICLs. The coordinates of the Cα atoms were used for the calculation. The C-terminal region (residues 303–328) was not used in the calculation of the geometric center of the intracellular side, since this region extended into the cytosolic region. The chimeric lysozyme region was also neglected from the calculation. A unit vector on the axis connecting the two geometric centers, which originated from the midpoint between the geometric centers toward the geometric center of the extracellular side, was calculated. The inner product between the unit vector and a vector from the midpoint to the Cα atom of every residue, except for those of the chimeric lysozyme region, was then calculated. The inner product score indicated the projected position of the residue on the axis (see Figure 1). The positive or negative score corresponded to the extracellular or cytosolic location of a residue, respectively, relative to the geometric center. The distribution of the inner product scores for the residues selected by the KL information values was compared with those of the remaining residues by the two-sided t-test. The null hypothesis was the same for all of the tests: the average of the residues corresponding to the sites with large KL information values is the same as that of the remaining residues. For the statistical test, the function in the statistical computing software R, “t-test,” was used for the evaluation.

**RESIDUE INDICATION**

The sites of each group selected by the KL information values are indicated on the corresponding sites of CXCR4 in this study. When the site has the number based on Ballesteros–Weinstein nomenclature (Ballesteros and Weinstein, 1995), the figure is also shown in the superscript. In this notation, the first digit indicates the number of the TM helix, and the following digit is the position counted from the most conserved site in each TM, to which the number 50 is assigned.

**RESULTS**

**THE PHYLOGENETIC ANALYSIS**

The multiple alignment of 622 sequences was constructed, which is downloadable from the URL: http://seala.cbrc.jp/~toh/suppl.html. The alignment of the representative sequences is shown in Figure 2. Based on the alignment, the phylogenetic tree of the CKRs and the decoy and viral receptors was constructed (Figure 3). Several clusters with high bootstrap probability (>80%) were identified in the tree, which included five decoy receptor groups and five viral receptor groups. The decoy receptor groups are referred to as CCRL1, CCRL2, CCBP2, CXCR7, and DARC, according to the constituent receptors. The numbers in each group were 23, 15, 15, 24, and 15, respectively. On the other hand, the viral receptor groups are referred to as E1, ORF74, UL33, bHV, and pox. The first three groups were named according to the constituent receptors. The bHV group consists US28, US27, and vGPCRs. Pox is a group of receptors derived from poxviruses. The numbers in each viral group were 18, 14, 19, 16, and 19. The evolutionary relationships between the CKRs and the decoy and viral receptors shown in the figure were roughly similar to those reported previously (Rosenkilde et al., 2001; Zlotnik et al., 2006). Murphy et al. (2000) suggest that the evolutionary rates of the CKRs are faster than those of the other GPCRs, because of the immune functions of CKRs. The long branch lengths suggested that the evolutionary rates of the receptors belonging to CCRL2, DARC, ORF74, UL33, bHV, and pox may be higher than those of the traditional CKRs, although we refrained from further examination of evolutionary rate accelerations in this study. In the subsequent analyses, each group of the decoy and viral receptors thus obtained was compared with the group of the traditional CKRs.

**DETECTION OF SITES WITH LARGE KL INFORMATION VALUES**

The differences in the amino acid composition at each alignment site were examined between the traditional CKRs and each group of decoy and viral receptors. The sites with large KL information values in the top 5% are summarized in Table 2. The residues corresponding to such sites were mapped on the structure of CXCR4 (Figure 4). As shown in Table 2, about 11 ~ 14

---

**Figure 1** Projection of a residue on the axis connecting the intracellular and extracellular sides of the receptor. The structure of CXCR4 is shown by the ribbon model. The membrane spanning helices indicated by the structural element page for CXCR4 in GPCRDB (http://www.gpcr.org/7tm/) are colored yellow. The sphere colored cyan indicates the geometric center of the alpha carbons of the membrane spanning helices. The red axis connects the geometric center of extracellular loops and the N-terminal loop and that of the intracellular loops. The midpoint of the axis is indicated by a filled sphere colored red. The distance between the cyan and red spheres is close (3.26 Å). That is, the midpoint is considered to roughly reflect the geometric center of the transmembrane helices. How to take the orthogonal projection of an amino acid residue to the axis is shown by using Residue X. Consider a vector from the midpoint to the Cα atom of the residue. By taking an inner product between the vector and a unit vector, which runs along the axis and is originated from the midpoint. The projected point is obtained by taking the inner product.
FIGURE 2 | Multiple amino acid sequence alignment of CKRs, decoy, and viral receptors. The GI number and protein name of the representative protein from each group are shown at the left side of the aligned amino acid sequence. In the case of CXCR4, the residue number of the first residue of the aligned sequence is shown after the protein name, and the TM regions described in the GPCRDB (http://www.gpcr.org/7tm/) are indicated by underlines. The corresponding sites for x.50 of Ballesteros–Weinstein nomenclature are colored blue. Four motifs, TxP, DRY, CWxP, and NPxxY, are enclosed by red line.
sites were selected from each group with the comparison of CKRs, and they included the sites in the sequences for GPCRs or the CKR-specific motif. Several sites that have been experimentally identified to be important for ligand binding or signaling were also selected. In addition, many uncharacterized sites were detected.

The DRY (Asp-Arg-Tyr) motif of the GPCRs is conserved as the sequence DRYLAIV in the traditional CKRs, from the end of TM3 to ICL2 (Graham, 2009). The motif is related to signal transduction, through interactions with G proteins. The conserved R134 is involved in the interchanges between the inactive and active conformations of GPCRs. In the inactive conformation, this Arg
interacts with its neighboring residue, D133\textsuperscript{3.49}, but in the active conformation, the residue interacts with Y219\textsuperscript{5.58} (Holst et al., 2010). The sites in the DRY region of the DRYLAIV sequence were only detected from the analyses with the decoy receptor groups, CCRL2, CCBP2, and DARC. In addition, Y219\textsuperscript{5.58} was also detected from the analysis with the DARC group. On the other hand, the sites in the LAIV\textsuperscript{3.52}~\textsuperscript{3.55} region of the DRYLAIV sequence were detected from the examinations with the decoy and viral receptor groups, CCBP2, UL33, and βHV. The CWxP motif is located in the middle of TM6. This W252\textsuperscript{6.48} is believed to function as a micro-switch in the receptor activation mechanism, and P254\textsuperscript{6.50} creates a kink in this helix, around which TM7 performs its rigid body movements during activation (Nygaard et al., 2009). The corresponding sites of this motif were detected from the analyses of two decoy receptor groups, CCRL1 and CCBL2, but not from those of any viral receptor group. The fifth site of the NPxxY\textsubscript{5-6} motif in TM7, Y302\textsuperscript{7.53} functions in the interchange of an inactive rotamer conformation (Nygaard et al., 2009). The sites of this motif were detected from the investigations with the decoy and viral receptor groups, ORF74 and UL33. The TxP motif of TM2 is known as a specific motif of the traditional CKRs. It is known that the TxP motif in TM2 is specific for traditional CKRs (Govaerts et al., 2001). The third site of the TxP motif, P92\textsuperscript{2.58}, bends the helix, which determines the intra-helical location that is involved in the receptor activation. The sites of the motif were detected from the analyses of two viral receptor groups, the ORF74 and pox groups, but not from the assessment with any decoy receptor group. In addition, several sites corresponding to highly conserved positions in GPCRs, which are denoted as x.50 by the Ballesteros–Weinstein nomenclature, such as N56\textsuperscript{1.50} and D84\textsuperscript{2.50}, were detected from analyses of several groups (see Table 2). Table 2 also shows the other important residues experimentally identified as having binding or signaling functions.

We examined which sites were commonly selected from the comparisons. No site was shared in all of the comparisons. Furthermore, there was no site commonly detected from the analyses with all of the decoy receptor groups or all of the viral receptor groups. However, several sites were detected from the different comparisons. For example, the sites corresponding to D74\textsuperscript{2.40}, D84\textsuperscript{2.50}, R134\textsuperscript{2.50}, A141\textsuperscript{3.57}, T142\textsuperscript{3.58}, S144\textsuperscript{3.60}, C218\textsuperscript{5.57}, K230\textsuperscript{6.26}, T241\textsuperscript{6.37}, C251\textsuperscript{6.47}, G306\textsuperscript{8.47}, and K308\textsuperscript{8.49} were detected from at least two assessments with decoy receptor groups. Most of these sites are located in ICL2, 3, and the C-terminal region. Among these sites, D84\textsuperscript{2.50}, A141\textsuperscript{3.57}, C218\textsuperscript{5.57}, and K308\textsuperscript{8.49} were also detected from at least one analysis with the viral receptor group. W94\textsuperscript{2.60}, W102 (ECL1), L136\textsuperscript{3.52}, H140\textsuperscript{3.56}, G207\textsuperscript{5.46}, L208\textsuperscript{5.47}, and K308\textsuperscript{8.49} were detected from at least two analyses of the viral receptor groups.
Table 2 | Selected sites with large KL information values.

| Residue (CXCR4) | Position | Region | B and W | Remarks | KL value | Frequency (%) | Reference |
|-----------------|----------|--------|---------|---------|----------|---------------|-----------|
| **A. (Decoy receptors)** | | | | | | | |
| | | | | | | | |
| **CCRL1** | | | | | | | |
| H203 | Extra | TM5 | 5.42 | Pocket | 7.03 | E | 93.3 | (L) Scholten et al. (2012) |
| C251 | Extra | TM6 | 6.47 | CWxP | 6.71 | C | 39.7 | F | 31.7 | (S) Nygaard et al. (2009) |
| A307 | Intra | C | 8.48 | NpxYYsF | 7.29 | E | 33.8 | V | 32.0 | |
| K308 | Intra | C | 8.49 | NpxYYsF | 8.07 | K | 73.0 | S | 86.7 | (L) Scholten et al. (2012) |
| | | | | | | | | |
| Y121 | Intra | TM3 | 3.37 | Pocket | 8.50 | Y | 670 | |
| L125 | Intra | TM3 | 3.41 | | 7.19 | L | 55.6 | F | 38.0 | |
| T142 | Intra | ICL2 | 3.58 | | 6.67 | T | 33.2 | V | 32.2 | |
| S144 | Intra | ICL2 | 3.60 | | 6.80 | A | 479 | |
| K230 | Intra | ICL3 | 6.26 | | 7.74 | R | 35.4 | |
| G231 | Intra | ICL3 | 6.27 | | 8.26 | N | 39.9 | |
| R235 | Intra | ICL3 | 6.31 | | 7.29 | H | 39.0 | |
| T241* | Intra | ICL3 | 6.37 | | 8.78 | I | 675 | |
| I261 | Extra | TM6 | 6.57 | | 8.35 | L | 471 | |
| L317 | Intra | C | 8.58 | | 7.44 | L | 44.9 | |
| **CCRL2** | | | | | | | |
| T73 | Intra | ICL1 | 2.39 | | 9.69 | T | 91.9 | |
| D84 | Intra | TM2 | 2.50 | | 8.20 | D | 95.4 | |
| D133 | Intra | ICL2 | 3.49 | DRY | 8.08 | D | 88.1 | |
| Y190 | Extra | ECL2 | – | Pocket | 9.14 | Y | 475 | |
| A237* | Intra | ICL3 | 6.33 | | 9.72 | A | 79.1 | |
| C251 | Extra | TM6 | 6.47 | CWxP | 9.27 | C | 39.9 | F | 31.5 | (S) Nygaard et al. (2009) |
| F292 | Extra | TM7 | 7.43 | | 10.56 | M | 72.8 | |
| G306 | Intra | C | 8.47 | NpxYYsF | 11.94 | G | 96.0 | |
| K308 | Intra | C | 8.49 | NpxYYsF | 8.96 | K | 73.8 | |
| | | | | | | | | |
| E31 | Extra | N | – | Pocket | 9.72 | Y | 78.0 | |

(Continued)
Table 2 | Continued

| Residue (CXCR4) | Position | Region | K and W | Remarks | KL value | Frequency (%) | Reference |
|----------------|----------|--------|---------|---------|----------|---------------|-----------|
|                |          |        |         |         |          |               | CKRs group |
| A141           | Intra    | ICL2   | 3.57    |         | 8.63     |               | A 82.2    |
| I215           | Intra    | TM5    | 5.54    |         | 8.31     |               | M 72.6    |
| I223           | Intra    | ICL3   | 5.62    |         | 8.16     |               | I 40.0    |
|                |          |        |         |         |          |               | R 66.8    |
|                |          |        |         |         |          |               |          |
| **CCBP2**      |          |        |         |         |          |               |          |
| R134           | Intra    | ICL2   | 3.50    | DRY     | 10.75    |               | R 99.0    |
|                |          |        |         |         |          |               | K 65.9    |
| A137           | Intra    | ICL2   | 3.53    | DRY     | 9.22     |               | A 77.1    |
|                |          |        |         |         |          |               | E 53.4    |
| G306           | Intra    | C      | 8.47    | NPxxY56F| 9.35     |               | G 96.2    |
|                |          |        |         |         |          |               | S 53.7    |
| V59            | Intra    | TM1    | 1.53    |         | 7.61     |               | V 94.3    |
|                |          |        |         |         |          |               | L 81.3    |
| T142           | Intra    | ICL2   | 3.58    |         | 10.12    |               | T 33.2    |
|                |          |        |         |         |          |               | Q 57.0    |
| S144           | Intra    | ICL2   | 3.60    |         | 7.54     |               | A 48.0    |
|                |          |        |         |         |          |               | H 33.5    |
| A152           | Intra    | ICL2   | 4.41    |         | 7.65     |               | K 60.6    |
| Y157           | Intra    | TM4    | 4.46    |         | 7.47     |               | C 49.1    |
|                |          |        |         |         |          |               | S 36.1    |
| S224           | Intra    | ICL3   | 5.63    |         | 7.64     |               |            |
| K230           | Intra    | ICL3   | 6.26    |         | 7.89     |               | C 80.7    |
|                |          |        |         |         |          |               | R 36.1    |
| Q233           | Intra    | ICL3   | 6.29    |         | 11.12    |               | K 36.1    |
|                |          |        |         |         |          |               | G 91.9    |
| **CXCR7**      |          |        |         |         |          |               |          |
| D74            | Intra    | ICL1   | 2.40    |         | 11.53    |               | D 72.0    |
|                |          |        |         |         |          |               | H 90.8    |
| F87            | Extra    | TM2    | 2.53    |         | 10.25    |               | F 75.4    |
|                |          |        |         |         |          |               | V 84.5    |
| G306           | Intra    | C      | 8.47    | NPxxY56F| 11.52    |               | G 96.2    |
|                |          |        |         |         |          |               | N 66.8    |
| K38            | Extra    | N      | 1.32    |         | 10.59    |               |            |
| G55            | Intra    | TM1    | 1.49    |         | 11.17    |               | Y 86.7    |
|                |          |        |         |         |          |               | G 97.8    |
| M63            | Intra    | TM1    | 1.57    |         | 10.87    |               | A 91.1    |
|                |          |        |         |         |          |               | L 49.7    |
| M72*           | Intra    | ICL1   | 2.38    |         | 10.57    |               | M 41.0    |
| L86            | Extra    | TM2    | 2.52    |         | 9.89     |               | E 54.2    |
|                |          |        |         |         |          |               | D 37.3    |
| A141           | Intra    | ICL2   | 3.57    |         | 11.00    |               | C 55.5    |
|                |          |        |         |         |          |               | W 38.4    |
| C218*          | Intra    | ICL3   | 5.57    |         | 10.28    |               | C 94.0    |
|                |          |        |         |         |          |               | F 79.9    |

(Continued)
### Table 2 | Continued

| Residue (CXCR4) | Position | Region | B and W | Remarks | KL value | Frequency (%) | Reference |
|-----------------|----------|--------|---------|---------|----------|--------------|-----------|
| **CKRs group**  |          |        |         |         |          |              |           |
| K236            | Intra    | ICL3   | 6.32    | 10.01   | K        | 56.4         | S 54.7    |
|                 |          |        |         |         | S        |              |           |
| L238*           | Intra    | ICL3   | 6.34    | 10.15   | V        | 32.9 I 31.4  | R 60.5    |
|                 |          |        |         |         | R        |              |           |
| L244            | Intra    | TM6    | 6.40    | 11.42   | V        | 54.0         | Y 92.6    |
|                 |          |        |         |         | Y        |              |           |
| K271            | Extra    | ECL3   | –       | 10.02   | F        | 79.8         |           |
| **DARC**        |          |        |         |         |          |              |           |
| N56             | Intra    | TM1    | 1.50    | 11.51   | N        | 98.8         | S 78.2    |
|                 |          |        |         |         | S        |              |           |
| D74             | Intra    | ICL1   | 2.40    | 11.91   | D        | 72.3         | R 44.9 W 30.1|
|                 |          |        |         |         | R        |              |           |
| D84             | Intra    | TM2    | 2.50    | 11.61   | D        | 95.5         | S 73.8    |
|                 |          |        |         |         | S        |              |           |
| Y116            | Extra    | TM3    | 3.32    | 11.54   | Y        | 57.9         | W 63.8    |
|                 |          |        |         |         | W        |              |           |
| R134            | Intra    | ICL2   | 3.50    | 12.36   | R        | 98.9         | G 61.7    |
|                 |          |        |         |         | G        |              |           |
| Y135            | Intra    | ICL2   | 3.51    | 13.42   | Y        | 92.6         | P 83.1    |
|                 |          |        |         |         | P        |              |           |
| Y219*           | Intra    | ICL3   | 5.58    | 12.72   | Y        | 96.2         |           |
|                 |          |        |         |         | Y        |              |           |
| Y302            | Intra    | TM7    | 7.53    | 11.86   | Y        | 95.7         | L 75.4    |
|                 |          |        |         |         | L        |              |           |
| F309            | Intra    | C      | 8.50    | 12.49   | F        | 98.7         | A 54.7    |
|                 |          |        |         |         | A        |              |           |
| V214            | Intra    | TM5    | 5.53    | 11.34   | V        | 52.4         | P 93.5    |
|                 |          |        |         |         | P        |              |           |
| C219*           | Intra    | ICL3   | 5.57    | 12.38   | C        | 94.1         | L 63.6    |
|                 |          |        |         |         | L        |              |           |
| T241*           | Intra    | ICL3   | 6.37    | 13.28   | I        | 67.3         | W 75.7    |
|                 |          |        |         |         | W        |              |           |
| L246            | Intra    | TM6    | 6.42    | 12.77   | W        | 96.0         |           |
|                 |          |        |         |         | W        |              |           |
| C296            | Intra    | TM7    | 7.47    | 11.56   | C        | 88.7         | V 64.4    |
|                 |          |        |         |         | V        |              |           |
| **B. (Viral receptors)** | | | | | | | |
| E1              | Extra    | TM3    | 3.32    | Pocket  | 7.62     | Y 56.1     |             |
|                 |          |        |         |         | Y        |              | (L) Schoiten et al. (2012) |
| Q96             | Intra    | ICL1   | 1.60    | 6.34    | M        | 83.5         |             |
|                 |          |        |         |         | M        |              | (L) Surgand et al. (2006) |
| A95             | Extra    | TM2    | 2.61    | 8.18    | A        | 66.0         |             |
|                 |          |        |         |         | M        |              | (L) Schoiten et al. (2012) |
| V99             | Extra    | TM2    | 2.65    | 6.66    | A        | 34.8         |             |
|                 |          |        |         |         | G        |              | (S) Rosenkilde et al. (2008) |
| N106            | Extra    | ECL1   | 3.22    | 8.48    | I        | 71.1         |             |
|                 |          |        |         |         | I        |              | (Continued) |
| Residue (CXCR4) | Position | Region | B and W | Remarks | KL value | Frequency (%) | Reference |
|----------------|----------|--------|---------|---------|----------|---------------|-----------|
| **CKRs group** |          |        |         |         |          |               |           |
| S123           | Intra TM3| 3.39   | Pocket  |         | 8.02     | G 49.6        | S 36.4    |
| G207           | Extra TM6| 5.46   | Pocket  |         | 7.77     | G 90.7        | S 84.3    |
| C220           | Intra ICL3| 5.59   |         |         | 6.62     | Y 473         | W 41.2    |
| G231           | Intra ICL3| 6.27   |         |         | 9.44     | N 39.6        | P 84.4    |
| **ORF74**      |          |        |         |         |          |               |           |
| D84            | Intra TM2| 2.50   |         |         | 11.25    | D 95.6        | (S) Rosenkilde et al. (2008) |
| P92            | Extra TM2| 2.58   | TxP     |         | 11.63    | P 98.4        | (L) Govaerts et al. (2001) |
| W94            | Extra TM2| 2.60   | Pocket  |         | 9.67     | W 74.7        | (S) Wu et al. (2010) |
| V112           | Extra TM3| 3.28   | Pocket  |         | 9.31     | V 373         | (L) Scholten et al. (2012) |
| W161           | Intra TM4| 4.50   |         |         | 9.93     | W 99.3        | (C) Ballesteros and Weinstein (1995) |
| N298           | Intra TM7| 7.49   | NPxYx5xF|        | 11.57    | N 94.9        | (S) Rosenkilde et al. (2008) |
| F304           | Intra C  | 8.45   | NPxYx5xF|        | 9.17     | F 94.6        | (S) Nygaard et al. (2010) |
| A307           | Intra C  | 8.48   | NPxYx5xF|        | 10.63    | E 33.4        | V 32.2    |
| K308           | Intra C  | 8.49   | NPxYx5xF|        |          | K 73.8        | (L) Scholten et al. (2012) |
| Y76 *          | Intra ICL1| 2.42   |         | Y       | 8.84     | F 60.6        | 32.7      |
| A83            | Intra C  | 2.49   |         | A       | 10.43    | A 55.3        | S 40.1    |
| H140 *         | Intra ICL2| 3.56   |         | H       | 9.90     | F 50.0        |           |
| A237 *         | Intra ICL3| 6.33   |         | A       | 9.31     | V 65.4        | I 30.2    |
| **UL33**       |          |        |         |         |          |               |           |
| L120           | Extra TM3| 3.36   | Pocket  |         | 13.59    | F 70.9        | (L) Surgand et al. (2006) |
| L136           | Intra ICL2| 3.52   | DRY     |         | 11.63    | L 66.3        |           |
| V139           | Intra ICL2| 3.55   | DRY     |         | 12.66    | V 90.5        |           |
| L208           | Extra TM6| 5.47   | Pocket  |         | 13.48    | F 70.1        | (L) Holst et al. (2010) |
| A291           | Extra TM7| 7.42   |         |         | 11.64    | A 60.0        | G 31.0    |
| K308           | Intra C  | 8.49   | NPxYx5xF|        | 11.88    | K 73.3        | (L) Scholten et al. (2012) |
| G55            | Intra TM1| 1.49   |         |         | 12.83    | G 97.9        | L 46.4    |

(Continued)
| Residue (CXCR4) | Position | Region | B and W | Remarks | KL value | Frequency (%) | Reference |
|-----------------|----------|--------|---------|---------|----------|---------------|-----------|
| W102            | Extra    | ECL1   | Pocket  | 12.73   | W        | 96.3          | L 34.3    |
|                 |          |        |         |         | L        | 34.3          |           |
| A141            | Intra    | ICL2   | 3.57    | 11.49   | A        | 82.0          | R 83.8    |
| G207            | Extra    | TM5    | 5.46    | 15.89   | G        | 90.4          | W 96.7    |
| C218 *          | Intra    | ICL3   | 5.57    | 11.64   | C        | 93.6          | F 96.2    |
| I222 *          | Intra    | ICL3   | 5.61    | 11.47   | I        | 75.4          | F 96.2    |
| Y256            | Extra    | TM6    | 6.52    | 11.30   | N        | 77.3          | V 48.6    |
|                 | Intra    | TM7    | 7.47    | 12.72   | C        | 88.6          | L 59.0    |
| **pHV**         |          |        |         |         |          |               |           |
| T73 *           | Intra    | ICL1   | 2.39    | 9.04    | T        | 92.5          | (L) Scholten et al. (2012) |
|                 |          |        |         |         | S        | 50.1          |           |
|                 |          |        |         |         | L        | 66.1          |           |
|                 |          |        |         |         | S        | 32.8          |           |
| L136            | Intra    | ICL2   | 3.52    | DRY     | 10.76    | L 66.1        | S 32.8    |
|                 |          |        |         |         |          |               |           |
| D171            | Extra    | TM4    | 4.60    | Pocket  | 8.92     | Y 47.7        | (L) Tian et al. (2005) |
|                 |          |        |         |         |          |               |           |
| Y190            | Extra    | ECL2   | –       | Pocket  | 11.26    | Y 47.3        | (S) Zhou et al. (2001) |
|                 | Intra    | TM7    | 7.34    | 9.65    | A        | 76.5          | F 37.0    |
|                 | Extra    | ECL3   | –       |         | 12.37    | C 96.2        |           |
|                 |          |        |         |         |          |               |           |
| **pox**         |          |        |         |         |          |               |           |
| C28             | Extra    | N      | –       | 9.87    | C        | 90.9          | (C) Wu et al. (2010) |
|                 | Intra    | N      | –       | 9.87    | P        | 70.4          | (L) Scholten et al. (2012) |
|                 |          |        |         |         | I        | 62.2          |           |
| P42             | Extra    | TM1    | 1.36    | 7.91    | P        | 70.4          | (L) Scholten et al. (2012) |
|                 |          |        |         |         | I        | 62.2          |           |
| T90             | Extra    | TM2    | 2.56    | TnP     | 6.79     | T 68.6        | (S) Govaerts et al. (2001) |
|                 |          |        |         |         |          |               | (S) Alvarez Arias et al. (2003) |
| W94             | Extra    | TM2    | 2.60    | Pocket  | 10.13    | W 74.6        | (L) Scholten et al. (2012) |
|                 |          |        |         |         | I        | 39.4          | (S) Rosenkilde et al. (2010) |
| L208            | Extra    | TM5    | 5.47    | Pocket  | 8.06     | F 70.8        | (S) Holst et al. (2010) |
|                 |          |        |         |         | M        | 67.6          |           |
| F248            | Intra    | TM6    | 6.44    | Pocket  | 9.82     | F 83.3        | T 33.8    |
|                 |          |        |         |         | S        | 49.9          | (S) Deupi and Standfuss (2011) |
|                 |          |        |         |         |          |               | (L) Surgand et al. (2006) |
| P27             | Extra    | N      | –       | 7.38    | P        | 56.2          | D 46.0    |
|                 |          |        |         |         | D        | 46.0          | E 30.8    |

(Continued)
Table 2 | Continued

| Residue (CXCR4) | Position | Region | B and W | Remarks | KL value | Frequency (%) | Reference |
|-----------------|----------|--------|---------|---------|----------|--------------|-----------|
| E31             | Extra N  | –      | Pocket  |         | 11.70    | Y            | 95.2      |
| L61             | Intra TM1| 1.55   |         |         | 6.93     | L            | 41.2      |
| I215            | Intra TM5| 5.54   |         |         | 728      | M            | 73.4      |
| I221 *          | Intra ICL3| 5.60  |         |         | 6.16     | L            | 65.3      |
| S227            | Intra ICL3| –     |         |         | 6.53     | L            | 45.4      |
| E277            | Extra ECL3| 728   |         |         | 6.16     | S            | 33.3      |

(All) The list of the sites detected from the comparisons with five decoy receptor groups. (B) The list of the sites detected from the comparisons with five viral receptors. Each row corresponds to a site with a large KL information value. The first column indicates the residue type and the residue number of CXCR4, to which the selected site corresponded. "*" Indicates a site located within 5 Å from the DRY motif in the CXCR4 structure. The second column indicates whether the site is located on the extracellular or intracellular side. The location was determined for the t-test (see Materials and Methods). The third column shows the position of the site in the primary structure of a GPCR (N-terminus, TM, ICL, ECL, and C-terminus). The fourth column indicates the site by the Ballesteros–Weinstein nomenclature. The fifth column provides remarks about the site, such as experiments and motifs. The "pocket" in this column was calculated at the CASTp site (Liang et al., 1998). A blank entry in the fifth column means that the site has not been characterized yet. The sixth column indicates the KL value of each site. The seventh column indicates the entry in the fifth column means that the site has not been characterized yet. The sixth column indicates the KL value of each site. The seventh column indicates the frequencies of the residues at each site. The upper half of the column indicates the frequencies for CKRs, and the lower half indicates the frequencies for the group under comparison. Only the residues with frequencies greater than 30% are shown. The eighth column indicates whether the site is involved in ligand binding (L), signaling (S), or conservation (C). The literature for experimental evidence or observations of the characteristics is also shown in the column, although the experiments were not always performed with the receptors under consideration, but with the homologs in the CKR family.

None of them, except for K308, was detected from the analyses of any decoy receptor group.

**STATISTICAL TEST FOR THE SPATIAL BIAS OF THE SITES WITH LARGE KL INFORMATION VALUES**

As shown in Figure 4, the distribution of the sites selected from the analyses with the decoy receptor groups seemed to be biased toward the cytosolic side of the CKR structure. In contrast, there did not seem to be any trends in the distribution of the sites obtained from the analyses with the viral receptor groups. To quantitatively examine the observations, the residues corresponding to the selected sites and the remaining residues were projected on the axis connecting the center of gravity of the ECLs including the N-terminal region, and that of the ICLs (see Figure 5). Based on the projection on the axis, t-tests were performed as described in the Section “Materials and Methods.”

The results of the t-tests are summarized in Table 3. As shown in this table, the null hypothesis was rejected in three cases of the analyses with decoy receptor groups, CCRL1, CCBP2, and DARC, under the significance level of 5%. To examine the bias further, the one-sided t-test was applied to the observations about the decoy receptor groups. The null hypothesis was the same as that of the two-sided test, but the alternative hypothesis was that the average of the residue with the large KL value is smaller than that of the remaining residues. We found that the null hypothesis was rejected for four cases with decoy receptor groups, CCRL1, CCBP2, DARC, and CXCR7 (data not shown). That is, the distribution of the residues corresponding to the sites with large KL information values of the decoy receptor groups, except for CCRL2, was biased toward the intracellular side of the receptor. The two-sided t-test was also applied to the analyses of the viral receptor groups. In all cases, the null hypothesis was not rejected. This result suggested that the residues selected by the KL information values of the viral receptors were distributed on both the extracellular and intracellular sides.

**DISCUSSION**

**DECOY RECEPTORS**

The difference in the amino acid composition at an alignment site between two receptor groups, as evaluated by the KL information value, was considered to reflect the difference in the functional constraints at the site between the groups. As described above, decoy receptors are able to bind to CKs, but do not induce signaling. The sites detected by the KL information value would reflect the distribution of the residues corresponding to the sites with large KL information values of the decoy receptor groups, except for CCRL2, was biased toward the intracellular side of the receptor. The two-sided t-test was also applied to the analyses of the viral receptor groups. In all cases, the null hypothesis was not rejected. This result suggested that the residues selected by the KL information values of the viral receptors were distributed on both the extracellular and intracellular sides.
FIGURE 4 | Mapping of the sites with large KL information values on the CXCR4 structure. The sites detected from the comparisons with (A) five decoy receptor groups and with (B) five viral receptor groups are mapped on the main chain structure of CXCR4. The residues corresponding to the detected sites with information about function and/or motif are depicted by space filling models, and are indicated according to the Ballesteros–Weinstein nomenclature. The corresponding amino acid residue types and numbers of CXCR4 are also shown in parentheses. On the other hand, the sites without any information are indicated by line models. The four motif regions are indicated by gray surface models. The residues that mapped on the extracellular side are colored blue, and those that mapped on the intracellular side are colored red.
FIGURE 5 | Projection of the residues on the axis connecting the intracellular and extracellular sides of the receptor. (A) Projections corresponding to the five comparisons with the decoy receptor groups. (B) Projections corresponding to the five comparisons with the viral receptor groups. The axis generation method and the projection method are described in the Section “Materials and Methods.” In each figure, the vertical line indicates the axis. Horizontal lines on the right side of the axis indicate the projected positions of the residues corresponding to the sites selected based on the KL information values. Horizontal lines on the left side indicate the projected positions of the remaining residues. The filled circles indicate the midpoint of the axis.

Table 3 | Results of t-tests.

| DECOY RECEPTORS | VIRAL RECEPTORS |
|-----------------|-----------------|
| CCRL1           | 3.87 × 10⁻³     |
| CCRL2           | 0.142           |
| CCBP2           | 3.37 × 10⁻⁷     |
| CXCR7           | 0.066           |
| DARC            | 1.51 × 10⁻⁴     |
| E1              | 0.981           |
| ORF74           | 0.080           |
| UL33            | 0.098           |
| βHV             | 0.308           |
| pox             | 0.144           |

The p-value for each two-sided t-test is shown. The details of the tests are described in the Section “Materials and Methods.”

were located on the intracellular side. The test with the CCRL2 group was the only one that did not suggest a statistically significant bias in the distribution of the detected sites. As described above, CCRL2 is also able to bind to chemerin (Yoshimura and Oppenheim, 2011). The adaptation to the new ligand may have introduced the change in the functional constraints on the extracellular side, which may be the reason why the null hypothesis was not rejected. This observation suggested that the functional divergence of CCRL2 was induced under different selective pressure, as compared to the other decoy receptors after gene duplication. CCRL2 forms a gene cluster together with the genes for CCR1, 2, 3, and 5 in several mammalian genomes (Nomiyama et al., 2011). The close relationship of CCRL2 to these CKRs and its distant relationship to the other decoy receptors in the phylogenetic tree (Figure 3) were consistent with the conservation of the gene orders in the genomes, although the bootstrap probabilities for the relationships were not always high. The evolutionary relationship and the conserved gene order, together with the acquisition of binding activity to a new ligand, suggested a unique evolutionary position of CCRL2 relative to the other decoy receptors.

The lack of signal transduction activity in the decoy receptors is attributed to the degeneration of the DRY motif (Comerford et al., 2007). Our study suggested that the degenerations of other motifs and functional residues may also be related to functional changes. For example, two decoy groups, CCRL1 and CXCR7, contained the typical DRY motif. However, the sites in other motifs that are related to the conformational change associated with the active-inactive switch had large KL information values in these decoy receptors. This observation suggested that the constraints for the residue conservation at the sites in the traditional CKRs are looser in the two decoy receptor groups (see Table 2). In addition to the motif sites, the highly conserved sites in the TM regions of GPCRs, including the traditional CKRs (x.50 in the Ballesteros–Weinstein nomenclature), had large KL information values in the analyses with several decoy receptor groups. The use of different amino acid residues at such sites may lead to functional and/or structural changes. Several sites with uncharacterized functional relationships also showed large KL information values. Most of them were
found in ICLs 2 and 3. As these loops are considered to interact with G proteins, the sites detected on the loops may be involved in the loss of the signaling function of the decoy receptors.

**VIRAL RECEPTORS**

We anticipated that the sites detected from the analyses with the viral receptor groups would be found on the extracellular side, since viral receptors exhibit signaling activity without ligand binding. As described above, however, the sites with large KL information values were found not only on the extracellular side, but also on the intracellular side. We examined the detected sites from the different viewpoint. CASTp is a program to identify pocket regions in a given tertiary structure. When we applied CASTp to the coordinates of CXCR4, the pocket region corresponding to the ligand binding cavities of GPCRs was reported with the highest score. The residues consisting of the pocket region were mainly projected on the extracellular side of the axis (see Figure 1), although some residues were projected on the intracellular side. The numbers of detected sites located in the pocket regions of the five decoy receptor groups were 2, 2, 0, 1, and 1, whereas 3, 2, 5, 3, and 4 sites were located in the pocket regions of the five viral receptor groups (see Table 2). The number of sites was transformed into the ratio to the total number of detected sites for each receptor group. The one-sided t-test showed that the difference in the ratios between the decoy and viral receptor groups was statistically significant (p-value = 0.003864). That is, more sites were detected in the pocket region in the viral receptor groups, as compared to the decoy receptor groups. As shown in Table 2, in addition, about half of the sites in the pocket region have been characterized as being involved in ligand recognition. These sites are often occupied by conserved, bulky amino acid residues in CKRs. The result suggested that the functional constraints at the ligand binding region are different between the viral receptors and the traditional CKRs, as we first expected.

The sites in the DRY motif were not detected in any of the viral receptor groups. This motif was basically conserved in the viral receptors, except for the ORF74 group. A previous study reported that ORF74 performs signal transduction, despite the fact that the DRY motif is changed to DTW (Rosenkilde et al., 2005). They also showed that the introduction of the DRY sequence into ORF74 induces functional reduction. In our study, the sequences collected as the ORF74 group showed variations in this region. For example, equid herpesvirus 2 has DTW, whereas the rodent and primate herpesviruses have xRC or xRY. Each variation includes the residues identical to those of the original DRY motif, which may have reduced the KL information value and led to the failure in the detection of the sites. Instead, the sites in the TxP and NPxxY motifs and the sites spatially surrounding the DRY motif were detected from the analysis of the ORF74 group (see Table 2). The amino acid replacements in the two motifs, which are considered to be involved in the conformational change, and those of the residues near the DRY motif may have contributed to the maintenance of the signaling activity of ORF74, despite the deviation from the typical DRY motif. In contrast, no sites in any motif were detected from the comparison with the E1 group. The E1 receptor reportedly lacks constitutive signaling activity (Rosenkilde et al., 2008). The conservation of the motifs suggested the difference in the signaling functions between the E1 group and other viral receptor groups.

We had not expected to detect the sites on the intracellular side from the comparisons with the viral receptor groups, since these receptors exhibit signaling activity without ligand binding. However, quite a few sites with large KL information values were also found on the intracellular side. As described above, the overlap of the selected sites between the decoy receptors and the viral receptors was small. The difference in the selected sites on the intracellular side between the viral receptor groups and the decoy receptor groups may be basically related to the difference in the activities of the receptor groups. That is, the sites of the viral receptor groups under the constraint to maintain the signaling without ligand binding may be different from the sites of the decoy receptor groups, where the functional constraints may have been weakened due to the loss of the signaling activity.

**CONCLUSION**

We have identified the alignment sites (and the corresponding amino acid residues) that may be responsible for the functional changes from CKRs to decoy receptors or viral receptors. The distributions of the identified residues on the tertiary structure seemed to reflect the functional differences. This prediction could be examined by an experimental study, such as amino acid replacement, or a computational study with molecular dynamic simulations. Such studies could provide deep insights into the mechanism of GPCR signaling through conformational changes. The experimental and computational confirmations of our prediction remain as future endeavors.

**ACKNOWLEDGMENTS**

Hiromi Daiyasu was supported by a Grant-in-Aid for Scientific Research and MEXT SPIRE Supercomputational Life Science. Hiroyuki Toh was supported by the Target Tanpaku program.

**REFERENCES**

Adachi, J., and Hasegawa, M. (1996). **MOLEPHY (Programs for Molecular Phylogenetics), Version 2.3b3.** Tokyo: Institute of Statistical Mathematics.

Alshul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402.

Alvarez Arias, D., Navenot, J. M., Zhang, W. B., Broach, J., and Peiper, S. C. (2003). Constitutive activation of CCR5 and CCR2 induced by conformational changes in the conserved TXP motif in transmembrane helix 2. *J. Biol. Chem.* 278, 36513–36521.

Ballesteros, J. A., and Weinstein, H. (1995). Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. *Methods Neurosci.* 25, 366–428.

Bonecchi, R., Savino, B., Borroni, E. M., Mantovani, A., and Locati, M. (2010). Chemokine decoy receptors: structure-function and biological properties. *Curr. Top. Microbiol. Immunol.* 341, 15–36.

Camarda, G., Spinetti, G., Bernardini, G., Mair, C., Davis-Poynter, N., Capogrossi, M. C., and Napolitano, M. (1999). The equine herpesvirus 2 E1 open reading frame encodes a functional chemokine receptor. *J. Virol.* 73, 9843–9848.
Functional divergence of CKR homologs

Choi, W. T., Tian, S., Dong, C. Z., Kumar, S., Liu, D., Madani, N., An, I., Sodroski, J. G., and Huang, Z. (2005). Unique ligand binding sites on CXCR4 probed by a chemical biology approach: implications for the design of selective human immunodeficiency virus type 1 inhibitors. J. Virol. 79, 15398–15404.

Comerford, I., Litchfield, W., Harata-Holst, B., Nygaard, N., Comerford, I., Milasta, S., Morrow, V., Litchfield, W., Harata-Holst, B., and Schwartz, T. W. (2010). A conserved aromatic lock for the tryptophan rotamer switch in TM-VI of seven-transmembrane receptors. J. Biol. Chem. 285, 3973–3985.

Ichihara, H., Dayiau, H., and Toh, H. (2004). How does a topological inversion change the evolutionary constraints on membrane proteins? Protein Eng. Des. Sel. 17, 235–244.

Jones, D. T., Taylor, W. R., and Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. Comput. Appl. Biosci. 8, 275–282.

Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment. Comput. Appl. Biosci. 18, 209–218.

Katoh, K., and Toh, H. (2008). Recent progress in the MAFFT multiple sequence alignment program. Brief. Bioinform. 9, 266–289.

Landgraf, R., Fisher, D., and Eisenberg, D. (1999). Analysis of helix conformation by profile–profile alignment. Proteins 35, 153–166.

Le Maire, C., Carrel, J.-L., and Schwartz, T. W. (2009). Ligand binding and micro-switches in 7TM receptor structures. Trends Pharmacol. Sci. 30, 249–259.

Lee, Y., Nibbs, R. J., and McColl, S. (2006). The rat a constitutive fashion.

Kaptein, S. J., Broers, J. L., Leurs, R., Bach, A., Engelstoft, M. S., Hébert, C. A., Horuk, R., I. F., Hébert, C. A., Horuk, R., and Vink, C. (2002). The rat chemokine receptor family: novel regulators of immune and inflammatory processes. Eur. J. Immunol. 32, 1327–1325.

Murphy, P. M., Bagnoli, M., Charo, I. F., Hébert, C. A., Horuk, R., Mita, S., Miller, L. H., Oppenheim, I. J., and Power, C. A. (2000). International Union of Pharmacology. XXII. Nomenclature for chemokine receptors. Pharmacol. Rev. 52, 145–176.

Najarro, P., Gobser, C., Hollinshead, M., Fox, J., Pease, J., and Smith, G. L. (2006). Yaba-like disease virus chemokine receptor 7L, a CCR8 orthologue. J. Gen. Virol. 87( Pt 4), 809–816.

Naumann, U., Cameroni, E., Pruenster, M., Mahabaleshwar, H., Raz, E., Zerwe, H. G., Rot, A., and Thelen, M. (2010). CXCR7 functions as a scavenger for CXCL12 and CXCL11. PLoS ONE 5, e9175. doi:10.1371/journal.pone.0009175

Nimaiyama, H., Osada, N., and Yoshie, O. (2011). A family tree of vertebrate chemokine receptors for a unified nomenclature. Dev. Comp. Immunol. 35, 705–715.

Nygard, R., Frimurer, T. M., Holst, B., Rønken, M., and Schwartz, T. W. (2009). Ligand binding and micro-switches in 7TM receptor structures. Trends Pharmacol. Sci. 30, 249–259.

Thelen, M., Thelen, S. (2008). CXCR7, CXCR4 and CXCL12: an eccentric triad. Neuroimmunol. 198, 9–13.

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680.

Tian, S., Choi, W. T., Liu, D., Pesavento, Y., Wang, Y., An, J., Sodroski, J. G., and Huang, Z. (2005). Distinct functional sites for human immunodeficiency virus type 1 and stromal cell-derived factor 1alpha on CXCR4 transmembrane helical domains. J. Virol. 79, 12667–12673.

Wu, B., Chien, E. Y., Mol, C. D., Fanalli, G., Liu, W., Katritch, V., Abagyan, R., Brooun, A., Wells, P., Bi, F. C., Hamel, D. J., Kuhn, P., Handel, T. M., Cherezov, V., and Stevens, R. C. (2010). Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. Science 330, 1066–1071.

Yoshimura, T., and Oppenheim, J. J. (2004). Simian cytomegalovirus encodes five rapidly evolving chemokine receptor homologues.

Chemokine receptor homologues.
Zhou, N., Luo, Z., Luo, J., Liu, D., Hall, J. W., Pomerantz, R. J., and Huang, Z. (2001). Structural and functional characterization of human CXCR4 as a chemokine receptor and HIV-1 co-receptor by mutagenesis and molecular modeling studies. *J. Biol. Chem.* 276, 42826–42833.

Zlotnik, A., and Yoshie, O. (2000). Chemokines: a new classification system and their role in immunity. *Immunity* 12, 121–127.

Zlotnik, A., Yoshie, O., and Nomiyama, H. (2006). The chemokine and chemokine receptor superfamilies and their molecular evolution. *Genome Biol.* 7, 243.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 06 June 2012; paper pending published: 17 June 2012; accepted: 05 July 2012; published online: 26 July 2012.

Citation: Daiyasu H, Nemoto W and Toh H (2012) Evolutionary analysis of functional divergence among chemokine receptors, decoy receptors, and viral receptors. *Front. Microbio.* 3:264. doi: 10.3389/fmicb.2012.00264

This article was submitted to Frontiers in Virology, a specialty of Frontiers in Microbiology. Copyright © 2012 Daiyasu, Nemoto and Toh. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and subject to any copyright notices concerning any third-party graphics etc.