THEO
CHEM

Characterize dynamic conformational space of human CCR5 extracellular domain by molecular modeling and molecular dynamics simulation

Shuqun Liu[a], Xiufan Shi[b], Ciquan Liu[b], Zhirong Sun[a],*

[a]*Department of Biological Sciences and Biotechnology, Institute of Bioinformatics, Tsinghua University, Beijing 100084, China*
[b]*Cellular and Molecular Evolutionary Key Laboratory, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China*

Abstract

The chemokine receptor CCR5 is the receptor for several chemokines and major coreceptor for R5 human immunodeficiency virus type-1 strains entry into cell. Three-dimensional models of CCR5 were built by using homology modeling approach and 1 ns molecular dynamics (MD) simulation, because studies of site-directed mutagenesis and chimeric receptors have indicated that the N-terminus (Nt) and extracellular loops (ECLs) of CCR5 are important for ligands binding and viral fusion and entry, special attention was focused on disulfide bond function, conformational flexibility, hydrogen bonding, electrostatic interactions, and solvent-accessible surface area of Nt and ECLs of this protein part. We found that the extracellular segments of CCR5 formed a well-packet globular domain with complex interactions occurred between them in a majority of time of MD simulation, but Nt region could protrude from this domain sometimes. The disulfide bond Cys20–Cys269 is essential in controlling specific orientation of Nt region and maintaining conformational integrity of extracellular domain. RMS comparison analysis between conformers revealed the ECL1 of CCR5 stays relative rigid, whereas the ECL2 and Nt are rather flexible. Solvent-accessible surface area calculations indicated that the charged residues within Nt and ECL2 are often exposed to solvent. Integrating these results with available experimental data, a two-step gp120-CCR5 binding mechanism was proposed. The dynamic interaction of CCR5 extracellular domain with gp120 was emphasized.
© 2004 Elsevier B.V. All rights reserved.

Keywords: CCR5; Molecular dynamics simulation; Extracellular domain; Disulfide bond; Interaction; Ligand binding

1. Introduction

Chemokines are a family of proinflammatory cytokines that attract and activate specific types of leukocytes via interaction with their specific receptors. This type of receptor belongs to the superfamily of seven transmembrane (TM) glycoproteins coupled to a G-protein signaling pathway [1]. Seven TM G-protein-coupled receptors (GPCR) have complex membrane topologies consisting of an N-terminal region, three extracellular and intracellular loops, and a cytoplasmic C-terminal tail. The CC chemokine receptor 5 (CCR5) is a member of such receptors that binds RANTES, MIP-1α, and MIP-1β [2] and is expressed by monocytes, memory T-lymphocytes, preferentially Th1 cells and NK cells [3–5]. It acts in concert with CD4 to associate with the envelope glycoprotein of HIV-1 R5 virus strains leading to fusion of viral and target cell membranes and subsequent viral entry [6–8]. R5 virus strains are largely responsible for virus transmission, and individuals who lack CCR5 due to a natural knock-out mutation in the CCR5 gene (ccr5 Δ32 allele) are highly resistant to HIV-1 infection [9,10]. The importance of CCR5 for viral entry and replication is further underscored by the observation that individuals heterozygous for the CCR5 D32 allele have a 2–4-year delayed progression to AIDS [8,11], most likely due to reduced expression levels of CCR5 [12,13].

Characterization of structural and functional determinants of CCR5 for its ligand binding activity and HIV-1

coreceptor function is essential for understanding the mechanisms of inflammatory diseases, HIV-1 viral entry and developing novel therapeutic agents against inflammatory diseases or HIV-1 infection. To date, CCR5 has been subjected to extensive studies using chimeric chemokine receptors [14–18], site-directed mutagenesis [19–21] and monoclonal antibodies (mAbs) competition binding experiments [22,23], which have revealed regions in CCR5, important for its interaction with HIV-1 and natural ligands, e.g. N-terminus (Nt) extracellular region plays a crucial role for high affinity binding of HIV-1 gp120, while ECL2 is important for natural ligands (RANTES and MIP-1β) binding, further, the extracellular loops (ECLs) are also important for inducing the conformational changes in gp120 that lead to membrane fusion. However, the CCR5 interactions with gp120 are a complex and dynamic process, which requires cooperativity between multiple extracellular domains and with multiple residues on the chemokine receptor. Although experimental data are important for understanding the interactions between coreceptor CCR5 and viral glycoprotein, structural model of CCR5 is indispensable to further explore the dynamic interaction process.

The CCR5 receptor belongs to the family of seven TM GPCRs. Unfortunately, structural information about GPCRs is difficult to get using experimental techniques because of enormous difficulties in preparation of samples suitable for subsequent X-ray, NMR or electron microscopic analysis [24]. Thus, despite of considerable efforts, there is no one 3D structure of CCR5 has been solved to atomic resolution by experimental methods to date. Fortunately, large amount of data is available for CCR5 from various biophysical, ligand binding and site-directed mutagenesis experiments, this information can be used to help building molecular model of CCR5. Furthermore, such CCR5 structural model can rationalize the phenomena observed in chimera and mutational experiments and further be employed in designing experiments destined to goal-oriented alteration of CCR5 binding activity. In this study, we have used an approach combining homology modeling and molecular dynamics (MD) simulation to propose plausible 3D structural models of CCR5, the conformational space feature of the extracellular domain was explored by a high temperature and 1 ns MD simulation. Analysis of the dynamic conformational characteristics and interactions of extracellular domain in 1 ns high temperature MD revealed that there are two primary conformer families and the Nt region and ECL2 are the most flexible structural elements. Combining the dynamic conformational features of extracellular domain of CCR5 with the available experimental data, a two-step mechanism of gp120 binding was proposed, which will shed light on the molecular mechanism of the HIV infection process from the angle of dynamic conformation transformation of Nt.

2. Materials and methods

2.1. Template-searching and assignment of transmembrane helix regions

The primary sequence of human CCR5 was taken from Swiss-Prot (P51681) [25]. The crystal structures of bovine rhodopsin (1F88a, 1F88b) at 2.80 Å resolution [26] were obtained as the modeling template of CCR5 from the ExPdb database using Swiss-PdbViewer (spdbv 3.7b2, http://www.expasy.ch/spdbv/) [27]. The assignment of transmembrane segments (TMS) of CCR5 was made based on the results proposed by others [28]. Sequence alignment between CCR5 and 1F88a is shown in Fig. 1, where the TMS and ECLs of CCR5 are labeled.

2.2. Homology modeling of CCR5

Transmembrane segments. The structures of TMS regions of human CCR5 were built by homology modeling based on the experimentally determined structure of the TM core of bovine rhodopsin using Homology module of InsightII (Accelrys Inc., San Diego, CA). The alignments between CCR5 and bovine rhodopsin were generated separately for TMS and loop regions. In the TMS,

Fig. 1. Sequence alignment between CCR5 and bovine rhodopsin. The potential TMS of CCR5 and ECLs were labeled. ' + ' and ' − ' indicate positive and negative scores for each equivalence, respectively. This alignment was generated by 3D-pssm program [45].

the alignments were manually adjusted and special care was taken to avoid gaps. Finally, residues in the membrane domain of bovine rhodopsin were replaced by corresponding residues of CCR5.

Loops and N-terminus. There are four extracellular segments and four intracellular segments in CCR5, out of them there are six loop regions. The initial structures of these loops were built by a loop conformational generation program (InsightII/Homology) that generates alternate loop regions using random tweak method (Accelrys Inc., San Diego, CA). In addition to the loops, there are two long segments in the N- and C-terminus that do not have any conformational constraint. The initial structures of the N and C-terminuses were built as an extended peptide chain conformation using EndRepair function of Homology module of InsightII. Because in most of GPCRs the ECL1 and 2 are linked via S–S bridge [29], disulfide bond was introduced between Cys101 and Cys178. The model containing only a Cys101–Cys178 disulfide bond is denoted the SS model. Considering potential disulfide crosslink between Cys20 and Cys269 by which the Nt and loop3 are linked [30,31], another model was built by connecting Cys20 and Cys269 of the SS model, and is denoted the 2SS model. For these two models, loop regions and entire models of CCR5 were refined through gradually structural segments relax to relieve steric contacts: splice points repairing → Nt residues relaxing → loop side-chains relaxing → entire loops relaxing → TMS relaxing.

2.3. Molecular dynamics simulation of extracelluar domain of CCR5 models

All computer simulations were performed on a Silicon Graphics Fuel workstation. Energy minimization and MD were carried out with the commercial software package InsightII/Discover3 molecular simulation program version 2.98 (Accelrys Inc., San Diego, CA). The Consistent Valence Forcefield was adopted in calculations. To approximate the solvation, calculations were carried out with a distance dependent dielectric constant $1 \times r$; non-bond part of the energy calculations was carried out employing a group-based summation method for Van der Waals interactions with cut-off 13 Å, and for electrostatic interactions with cut-off 20 Å, spline width 1.0 Å and buffer width 1.0 Å, respectively. All hydrogen atoms were included in the calculations.

The SS and 2SS models were first subjected to energy minimization via 200 steepest descent steps followed by 500 conjugate gradient iterations. The loop structures generated by loop generation program are alterable to a large extent and Nt generated by EndRepair is a peptide chain with randomly extended conformation. So, in order to investigate the motion tendency in extracellular domain and search more effectively the conformational spaces of the Nt and ECLs of CCR5 that are important for ligand binding and HIV entry, a high temperature (1000 K) and a relative long

simulation time (1 ns) were employed with write frequency steps every 2 ps, and during the MD simulation, only residues in the extracellular domain were allowed to move. Finally, the extracellular domains of the representative models of CCR5 selected from previous high temperature MD were subjected to a simulated annealing procedure: the extracellular parts of the representative models of SS and 2SS CCR5 were separately solvated with a layer of waters 15 Å thick, the temperature of the systems was fast heated to 1000 K, cooled to 298 K up to 20 ps, and a constant temperature (298 K) MD for additional 200 ps, followed by energy minimization. During these processes the TM helices and the outer 5 Å thick water layers were fixed to prevent diffusion of the inner water molecules away from the solvation shell. The force field and non-bonded cutoff are the same as those in the high temperature MD calculations except for the dielectric value, which was set to a constant 1.

3. Results

3.1. Structural models of SS CCR5 and 2SS CCR5

Considering the first 40 ps of 1 ns MD was the equilibration phase, the 480 structures of the SS CCR5 and 2SS CCR5, respectively, were extracted from the MD trajectory in the last 960 ps at a time interval of every 2 ps, and the extracellular domains of all these structures were subsequently energy-minimized to optimize their conformations. For 2SS CCR5 models, the 480 structures were clustered into two primary conformational families according to the backbone structural similarity of extracellular domain. The representative structures for each family were analyzed according to the known experimental data and conformational differences of Nt regions, then, two structures A and B were selected, and the extracellular domains of which were embedded separately in 15 Å thick water layers and then were, respectively, subjected to simulated annealing procedure and another 200 ps normal temperature (298 K) MD simulations with write frequency steps every 1 ps. Fig. 2 shows the total energies of structures A and B extracted from the 200 ps normal temperature dynamics trajectory. Trajectory a reveals that total energy of representative structure A reaches a plateau with the range $\sim -48{,}105$ to $\sim -47{,}936$ kcal/mol after ~ 55 ps, and RMS calculation reveals the RMS value between the Cα atoms of extracellular parts of sampled structures within the last 145 ps trajectory is less than 1.2. Trajectory b shows that total energy of structure B reaches a plateau with the range $\sim -46{,}948$ to $\sim -46{,}770$ kcal/mol after 75 ps, and the RMSD value between Cα atoms of extracellular parts of sampled structures within the last 125 ps is less than 0.9. Finally, two structures with the lowest total energy in the MD trajectories a and b were separately selected as the possible models for 2SS CCR5 (Fig. 3, models A, B).
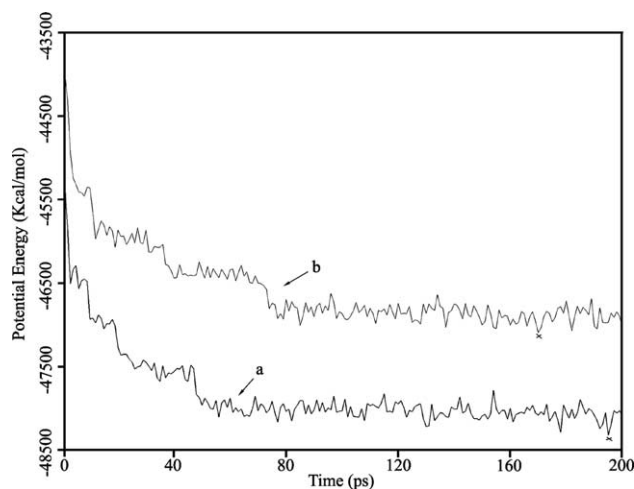
Fig. 2. The total energies of structures A and B extracted from the 200 ps normal temperature (298 K) dynamics trajectory. Trajectories a and b are energy of representative structures A and B during MD, respectively. Two structures with the lowest energy from the dynamics trajectories as labeled by cross were selected as possible structural models for 2SS CCR5.

For the 2SS CCR5 model A, the extracellular domain of CCR5 take on a compact globular state with Nt region adsorbing on the surface of globular domain. But for the model B, although the extracellular segments of ECL1–3 pack into a globular domain with a relative compact conformational state, the Nt 1–19 region stays away from and locates at the top of this globular domain. The distinct difference between these two models exists mainly in conformational variance of Nt regions, it is worthy to point out that the total energy of model A is also much lower than that of model B, about 1157 kcal/mol. For SS CCR5 models, the 480 structures extracted from the last 960 ps of high temperature MD have the similar conformational feature with the ECL1–3 and Nt region separately forming two inequable domains. The extracellular domain of a representative structure of SS CCR5 was embedded in 15 thick water layer and then subjected to simulated annealing procedure and an additional 200 ps normal temperature (298 K) MD simulation as treated with 2SS CCR5, and the structure with the lowest energy was shown in Fig. 3 (model C). Unlike the models A and B, the Nt region of this SS CCR5 model neither closely contacts with ECL1–3 nor projects from globular domain formed by ECL1–3, it forms another small globular domain itself and lies beside ECL1–3 domain.

The CCR5 models A and B have been deposited in Protein Data Bank. 1ND8 is for the 2SS CCR5 model B with the Nt region projecting from extracellular domain, and 1NE0 for the 2SS model A with the Nt region adsorbing on the surface of extracellular domain.

3.2. Analysis of MD data

The function of disulfide bond linking the Nt and ECL3. It has been known that disulfide bonds linking together the ECLs and Nt are necessary for ligand binding and coreceptor activation by mutation data [30], and the disulfide bond function has been presumed to maintain the structural integrity of extracellular domain [31]. To further investigate how the disulfide bond between C20 and C269 maintains such conformational integrity, the movement trajectories of Cys20 dihedral angel phi vs psi for 2SS CCR5 and SS CCR5 model during high temperature MD simulation were calculated, respectively (Fig. 4), Fig. 4a,b show that the motion range of dihedral angel phi is mainly restricted to −170 to −60° with psi angel −90–150°, but Fig. 4b shows that the phi angel in SS CCR5 has more opportunities to go into the range of −60–110° than phi angle in 2SS CCR5, this is due to the introduction of disulfide bonds into SS CCR5 constrains the local conformation motion of Cys20. Since Cys20 locates at the two-thirds point of the whole Nt region, we conjectured that the local conformation change of Cys20 was essential in influencing the entire conformational state of Nt region. To further prove this, the geometry center of Nt 1–19 residue region was defined as a pseudo-atom, the geometry center of membrane interface binding residues Arg31, Tyr89, Leu103, Phe166, Val199, Phe260 and Ala278 was defined as another pseudo-atom, and the distances between these two pseudo-atoms for SS and 2SS CCR5 model in the 1 ns MD were calculated and shown in Fig. 5. Two unexpected features were found in Fig. 5: (i) the distance vibration amplitude of 2SS CCR5 geometry center is wider than that of SS CCR5; (ii) almost entire trajectory of 2SS CCR5 stays on the top of the trajectory of SS CCR5. The wider vibration amplitude infers that rather than contracted, on the contrary, the motion range of Nt 1–19 region is expanded by introducing C20–C269 disulfide bond. The top location suggests that the C20–C269 disulfide bond keeps the Nt 1–19 region farther away from cell membrane surface. Through the animation of the 1 ns MD process, we found that the Nt region of SS CCR5 formed a compact globular domain with movement scope restricting at the one side of ECLs domain throughout the last 960 ps of MD, which explained the narrow vibration amplitude of distance change trajectory of SS CCR5 shown in Fig. 5. However, the Nt region of 2SS CCR5 did not form a separate structural domain itself, it was linked to the ECL3 via disulfide bond and formed a bigger globular domain with ECL1–3 in the most of MD process, but sometimes the Nt 1–19 region disengaged from ECLs and extended into the outer space of this domain. The peak regions (380–510 ps) in the trajectory for 2SS CCR5 shown in Fig. 5 correspond to these conformers. These results permit the following conclusions to clarify the significances of disulfide bond Cys20–Cys269 in maintaining conformational integrity of extracellular domain. (i) Avoiding the formation of a separate Nt domain and causing ECLs and Nt together to form a whole extracellular structural domain; (ii) increasing structural flexibility of Nt 1–19 region by constraining local movement of Cys20; (iii) keeping Nt 1–19 region stay on
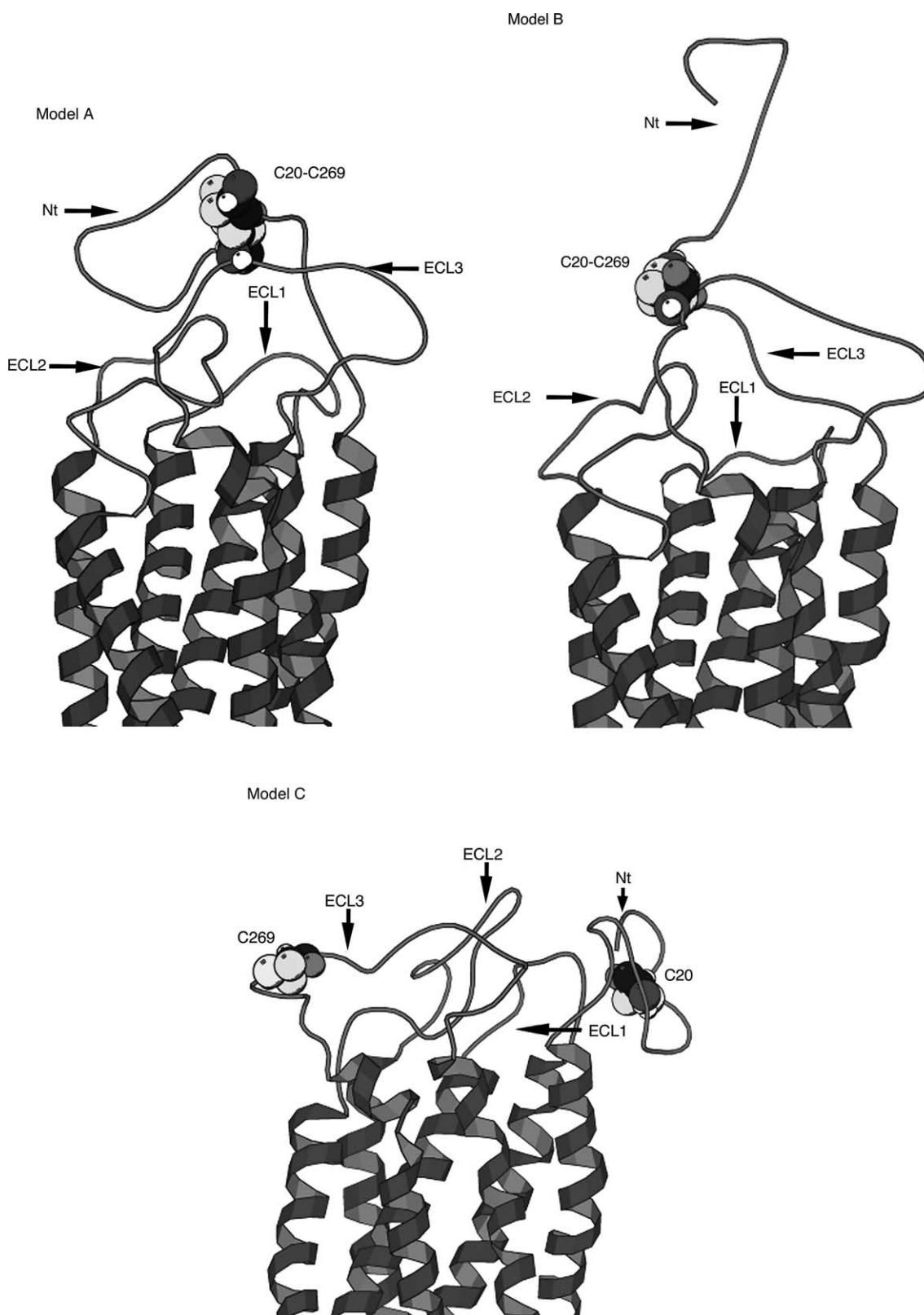
Fig. 3. The representative structural models for 2SS CCR5 (models A, B) and SS CCR5 (model C). The Nt regions and ECLs 1–3 are labeled. The side chains of Cys20 and Cys269 are rendered as CPK. Model A shows the Nt region adsorbs on the surface of extracellular domain formed by ECL1–3. Model B shows that Nt region extends into outer space of extracellular domain. Model C shows Nt region forms a separate domain and lies beside the domain formed by ECL1–3. The figure was generated by using the MOLSCRIPT program [46].
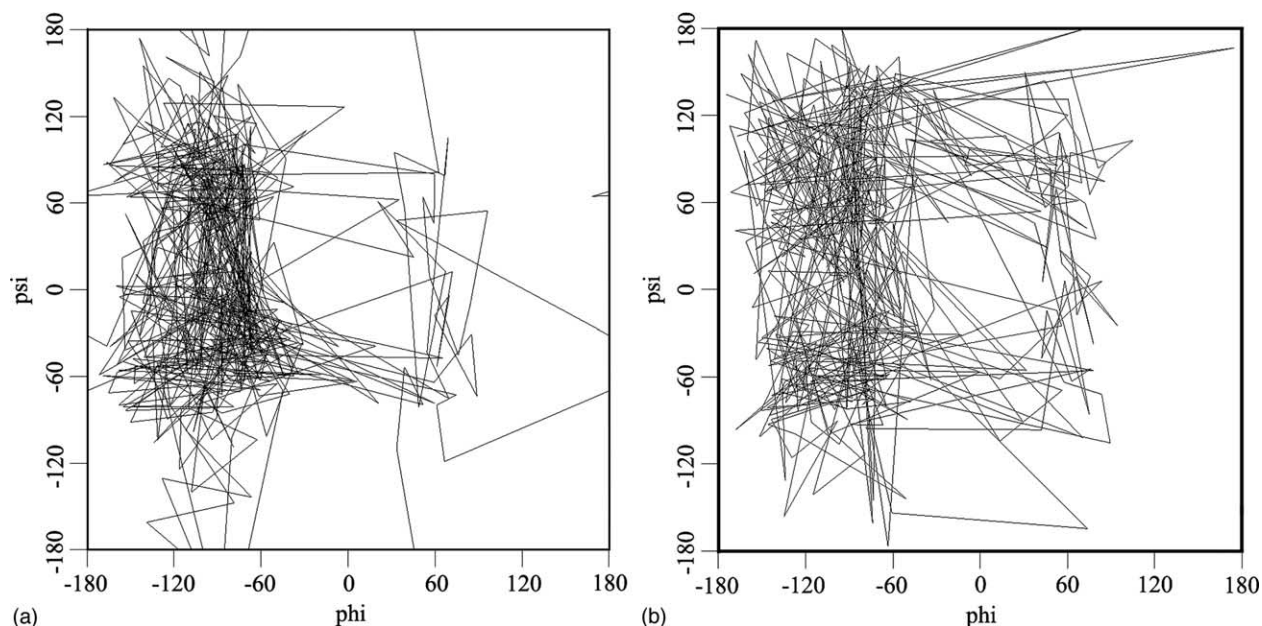
Fig. 4. Motion trajectory of Cys20 dihedral angel phi vs psi for 2SS CCR5 (a) and SS CCR5 (b) model in 1 nm MD simulation. (a) The motion scope of dihedral angel phi for 2SS CCR5 model is mainly restricted to -170 to $-60°$, and psi angel is restricted to $-90-150°$. (b) Although he motion scope of dihedral angel phi for SS CCR5 model is restricted to -170 to $-60°$ and psi angel to $-90-150°$, the phi angel in SS CCR5 has more opportunities to go into the range of $-60-110°$ than the phi angle in 2SS CCR5.

the top surface of extracellular domain and offering Nt opportunity to extend into outer space of extracellular domain. Thus, the 2SS CCR5 was considered as a more reasonable structural model than SS CCR5.

Conformational flexibility of extracellular segments. The extracellular domain includes Nt, ECL1, ECL2 and ECL3. In this study, the initial structures of Nt and ECL1–3 were build as extended peptide chain and alterable loop conformation, respectively. To evaluate the motion range of ECLs in 2SS CCR5, the geometry centers of ECL1, ECL2 and ECL3 were, respectively,

defined as pseudo-atoms pseatom1, 2 and 3, the geometry center of membrane interface binding residues (see *The function of disulfide bond linking the Nt and ECL3*) was defined as pseudo-atom X, then the distance trajectory between pseatom1 and X, pseatom2 and X, as well as pseatom3 and X were calculated, respectively, in the 1 ns MD (Fig. 6). Fig. 6 shows that the ECL2 behaves the biggest vibration amplitude with the value scope 9–17, and ECL1 behaves the smallest vibration amplitude with value cope 6–10, indicating that ECL2 has the maximal freedom of motion and ECL1 has the minimal ones.
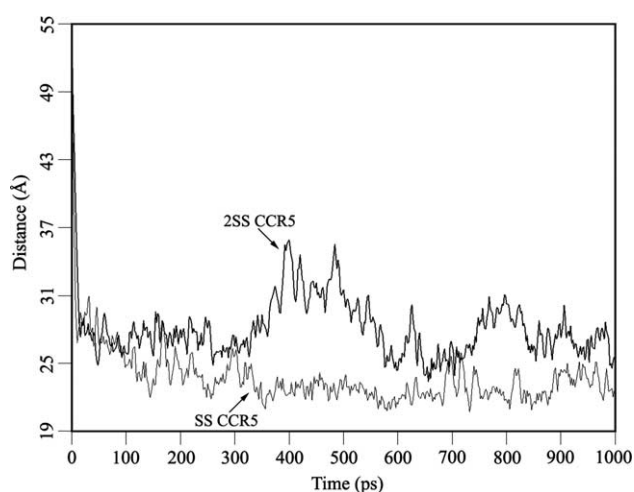
Fig. 5. Distance trajectory between geometry center of Nt 1–19 region and that of membrane interface binding residues for SS CCR5 and 2SS CCR5 model in 1 ns MD simulation. The maximum of vibration amplitude for 2SS CCR5 is about 12 Å, while for SS CCR5 is about 7 Å. The trajectory of 2SS CCR5 stays on the top of the trajectory of SS CCR5.
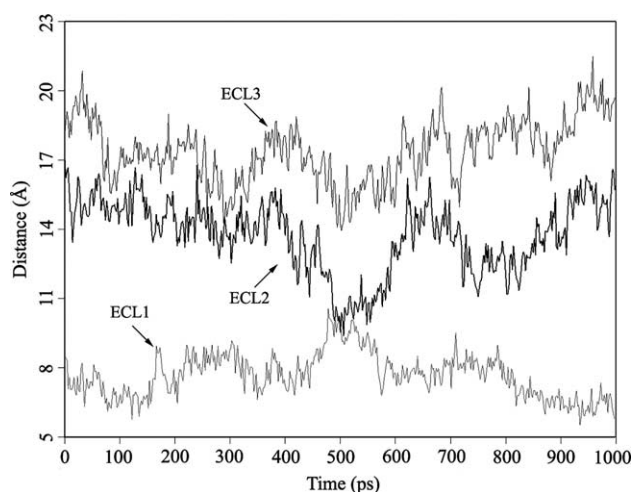

Fig. 6. The distance trajectory between geometry center of ECL1–3 and that of membrane interface binding residues for 2SS CCR5 model in 1 ns MD. The maximums of vibration amplitude for ELC1–3 are about 4, 8 and 6 Å, respectively.

The RMS comparisons between corresponding Cα atoms of extracellular domain of 480 conformers extracted from the 1 ns MD were performed to further assess the conformational flexibility of extracellular segments. For Nt regions and ECL1–3 the RMS value scopes are 0–17.2, 0–7.5, 0–13.1 and 0–9.22, respectively. In combination the RMS values with freedom of motion of the extracellular domain presented above, we concluded that ECL1 and Nt region represent the most rigid and the most flexible extracellular segments of 2SS CCR5, respectively, and ECL2 is the most flexible loop among ECL1–3.

Interaction between extracellular segments. At the beginning of high temperature MD simulation, the Nt part was an isolated structural segment that was not involved in any H-bonding and electrostatic interaction with other ECLs, and the residues within Nt possessed of the large solvent-accessible surface. However, along with the time lapsing, the Nt gradually closed to the ECLs, about 40 ps later, the Nt was adsorbed on the surface of ECLs with which a globular domain was formed. Then, the dynamic and complex interactions were detected between Nt and ECLs as well as between ECLs during the whole last 960 ps MD. For the 2SS CCR5 model, several interactions including H-bond, electrostatic, and Van der Waals contact between extracellular segments were investigated, the solvent-accessible surface area of extracellular segments was also calculated. The results are: (i) the high frequency H-bonds could be detected between Nt region and ECL1 (Gln27-Ala92, Gln27-Asn98), ECL2 (Asn24-Ser179, Asn24-Gln186, Asn24-Ser180, Lys26-Thr177, Lys26-Ser180, Lys26-Tyr176, Gln27-Thr177), ECL3 (Met1-Ser272, Met1-Asn268, Asn13-Asn268, Tyr15-Asn267, Cys20-Ser270, Gln21-Asn268, Gln21-Ser270, Lys22-Ser270, Lys22-Ser271, Val25-Asn273, Lys26-Arg274, Lys26-Asp276), between ECL1 and ECL2 (Thr99-Tyr176, Gln102-Ile165, Gln102-Glu172, Gln102-Thr177) and between ECL2 and ECL3 (His181-Asn268, Ser185-Glu262, Gln188-Glu262, Trp190-Glu262). The low frequency H-bonds were observed between Nt and ECL1 (Lys22-Thr99, Ile28-Ala90), ECL2 (Gln4-Lys171, Gln4-Thr177, Asn24-Gln188, Gln21-Ser179, Lys22-His181, Gln27-His181, Ile28-Tyr176), ECL3 (Ser6-Asn267, Asn13-Asn267, Gln21-Ser272, Lys22-Ser272, Asn24-Ser271, Gln27-Asn273, Gln27-Leu275, Ile28-Arg274), between ECL1 and ECL2 (Asn98-His175, Met100-Leu174, Cys101-Gln186, Gln102-Ser179), and between ECL2 and ECL3 (Ser180-Asn267, Ser180-Ser271, His181-Ser272, Tyr184-Glu262, Ser185-Asn267, Gln188-Phe264). In addition, large numbers of instantaneous H-bonding interactions could be observed between extracellular segments. (ii) Electrostatic interactions could often be detected between Nt and ECL2 (Lys22-Glu172, Asp2-Lys171, Lys26-Glu172, Asp11-Lys191), ECL3 (Lys26-Asp276, Asp2-Arg274), and between ECL2 and ECL3 (Lys191-Glu262). (iii) Van der Waals contacts were measured with pairwise interresidual

distance less than 2.5, we found that residues Asp2, Asn24, Lys26, Gln27 in Nt could often fall into close contact with residues Tyr176, Ser179, Tyr176, Thr177 in ECL2, respectively, other large numbers of contacts were also observed in residue pairs of Nt-ECL3, ECL2–3, and 1–2, only few contacts were observed between ECL1 and 3. (iv) The residues that usually exposed to solvent with more than 70% of the total solvent-accessible surface area are: Met1, Asp2, Tyr3, Gln4, Pro8, Ile9, Asp11, Asn13, Tyr14, Tyr15, Glu18 in Nt, Ala92, Gln93, Phe96 in ECL1, Arg168, Gln170, Lys171, Glu172, Leu174, Tyr176, Lys191, Asn192 in ECL2, Phe264, Ser270, Ser271, Asn273, Leu275 in ECL3. These residues could fall into interaction with water molecules, some of them might also be the potential interaction sites with ligands and HIV-1 envelope glycoprotein. Together with the help of disulfide bridges between Cys20–Cys269 and Cys101–Cys178, the various interactions described above cause the extracellular segments to form a complete globular structure domain.

4. Discussions

The fact that none of CC chemokine receptors has been crystallized and no experimental 3D structures determined to date made it difficult for us to construct the very valid 3D model of CCR5 using the method of homology modeling. However, although bovine rhodopsin has low sequence homology with CCR5 (the identity is 20.6%), it belongs to the GPCR family and its crystal structure displays an explicit conformational feature of a bundle of seven TM α-helices shared by other GPCRs, and the sequence identity in the TMS between CCR5 and rhodopsin is ~30%, which should be possible to generate models where ~80% of the Cα atoms are within 3.5 Å of their correct positions [32]. So, it is reasonable to adopt the seven TMS of bovine rhodopsin as templates to construct the structure conserved regions (seven TMS) of CCR5. However, the PRO residue in TMS has long been known to affect helical stability and steric conflict in the helix structure [33]. In this study, the local structural motif containing PRO residues in TMS1, 2, 4–7 were re-examined and re-relaxed until the steric conflicts are removed and chirality of PRO residues are correct. At the same time, the TMS of CCR5 models were fixed but not embedded in lipid layer during the MD simulation, which can be explained as following reasons: (i) the large and densely packed 7-helix bundles in GPCRs is significantly screened from lipids and, therefore, protein–lipid contacts do not play a major role in determined folding of this domain [34], also, though extracellular domain has contacts with polar headgroups of lipids, they contribute little to conformation formation of extracellular protein part [35]; (ii) the conformational flexibility of multi-helical membrane bundle is rather lower than that of the extracellular domain [36]; (iii) there are no precise models

of the lipid bilayer suitable for including into the system under study.

The initial structures of extracellular domain were built as alterable loops for ECLs and an extended peptide chain for Nt. However, the structures thus obtained might be considered only as a crude approximation of these protein parts. Exhaustive conformational search is required to access as many as possible structures of these protein parts, to do this in a limited time interval, a high temperature (1000 K) was used in the MD simulation because molecular conformation have more chances to get over higher energy barrier under such temperature. For the finally selected representative 2SS CCR5 structural models A and B, the geometrical and stereochemical features have been calculated for validation by the ProCheck program at http://pdbdep.protein.osaka-u.ac.jp/validate/, the results show that more than 96% torsions angles are within the expected Ramachandran regions, all bond distances and angles lie within the allowable range about the standard dictionary values, and the atom chirality of these two models is also correct, which indicates that our CCR5 models are reasonable in geometry and stereochemistry. At the same time, in the case of observation on a relative large time scale, the high temperature MD simulation for CCR5 model could provide a wealth of useful information, e.g. rationalization of the available experimental data, understanding the relationship between CCR5 structure and function, exploring the HIV infection mechanism.

Huang et al. have built CCR5 models in complex with MIP-1β and RANTES to explore the residues critical for interactions between CCR5 and its natural ligands [37], Paterlini built a structural model of CCR5 to explain the binding and selectivity of the antagonist TAK779 [38], Arseniev et al. also explored the conformational characteristic of CCR5 extracellular domain by structural modeling of CCR5 [35]. A common feature shared by these studies above is that the authors explore CCR5 function from the static structural models. Here in this study, because studies with experimental methods on CCR5 have revealed that extracellular segments are essentially involved in interaction with chemokines and HIV viral glycoprotein, our primary objective is to investigate the general tendencies in the dynamic behaviour of the extracellular domain of CCR5 through structural modeling and high temperature MD simulation, and to explain CCR5 function from the dynamic conformational character and structural flexibility of extracellular domain.

In addition to the disulfide bond Cys101–Cys178 linking ECL1 with ECL2, another disulfide bridge is predicted to exist between Cys20 in Nt and Cys269 in ECL3 [39]. Using site-directed mutagenesis, Genoud et al. [30] mutated the four extracellular cysteines of CCR5 single or in combination to investigate their role in ligand binding and ability to function as a viral coreceptor, they found all cysteine mutants were unable to bind detectable

levels of MIP-1β, and did not respond functionally to CCR5 agonists. Blanpain et al. [31] found that the mutations of the four cysteine residues had a significant impact on viral entry, and the coreceptor activity of mutant C269A was only 10–20% of that of wild-type CCR5. These results indicate from the angle of biochemistry that the disulfide bonds of CCR5 are required for maintaining the structural integrity necessary for ligand binding and coreceptor function. Here, we further explored the function of extracellular disulfide bonds Cys20–Cys269 from the angle of CCR5 structure, and found that Nt region will form a isolated globular domain locating at the side of ECLs with less conformational flexibility if there was no disulfide bond linkage between Nt and ECL3, while the Nt region could attached on the ECLs with which to form a whole extracellular domain if the disulfide bond between Cys20 and Cys269 existed. Meanwhile, the orientation of Nt 1–19 region was on the top of the extracellular domain and its conformational flexibility was enhanced by the restriction of local freedom of motion of Cys20. The appropriate orientation and strong flexibility of Nt make it ready to interact with CC chemokines and HIV gp120. Our MD simulations for CCR5 rationalize previous mutation experiments [30,31, 39], and show disulfide bonds playing important roles in maintaining the structural organization of the receptor outer domains.

Although the ECL1–3 and Nt are the flexible conformational regions comparing with rigid TMS core of CCR5, the interaction networks formed by disulfide bonds, hydrogen bonds, salt bridges and Van der Waal contacts make these extracellular segments combine into one well-packed globular domain. However, here we should stress that the interactions between extracellular segments are dynamic and complex: on one hand, the complex interactions maintain the conformational integrity of the whole extracellular domains, and the compact packing in extracellular domain is preferable for its activities because it significantly decreases global flexibility of this protein part and facilitates selection of the conformer suited for binding; on the other hand, the dynamic interactions endow individual structure of extracellular segments (especially for Nt 1–19 region and ECL2) with strong local flexibility that is involved in essential functional roles of CCR5.

Nt segment of CCR5 contains 30 amino acid residues, among them there are two positively and three negatively charged and, moreover, all the negative charges are within the Nt 1–19 region, while the positive charges are close to the TM domain. Using site-directed alanine-scanning mutagenesis, Dragic et al. [19] showed that substitutions of the negatively charged aspartic acid residues at positions 2 and 11 (D2A and D11A) and a glutamic acid residue at position 18 (E18A), individually or in combination, impaired CCR5-mediated HIV-1 entry for M-tropic strains, these mutations also impaired Env-mediated membrane

fusion and the gp120-CCR5 interaction. Blanpain et al. [40] also investigated the role of specific residues in the CCR5 Nt 1–19 region for chemokine and HIV-1 gp120 binding by using of a panel of truncation and alanine-scanning mutants, and found the truncations of the Nt region resulted in a progressive decrease of the binding affinity for chemokines and R5 Env protein, mutants lacking residues 2–13 exhibited fairly weak responses to high concentrations (500 nM) of RANTES or MIP-1β, and the negatively charged (D2, D11, E18) and aromatic (Y3, Y10) residues were identified to play an important role in both chemokine and Env high affinity binding. However, substitution of the CCR5 amino-terminal domain with the corresponding region from divergent receptors, such as CCR1, CCR2b, CXCR2 or CXCR4, had little effect on chemokine binding and coreceptor activation [18,41]. Although these chemokine receptors share little primary sequence identity, they contain similar motifs of hydrophobic and charged residues, which suggest that the role of the Nt is not receptor specific, but structurally required, the structural organization and conformational state of Nt region shared by other receptors played an important role in the high affinity binding of chemokines and gp120. Interestingly, our 1 ns MD simulation of 2SS CCR5 model displayed that the Nt region was the most flexible part among the extracellular segments, and it could adopt two clearly distinct conformational states with different free energies. One state with the lower free energy was that it was adsorbed on the surface of ECLs during a majority of time of MD simulation (Fig. 3, model A), another state with a higher potential energy was that it could project from extracellular domain and extend into the outer space of this domain (Fig. 3, model B) at times. Comparing the later conformational state of 2SS CCR5 with the former, we found that the solvent-accessible surfaces of the residues 1–19 were clearly increased when Nt protruded from extracellular domain. The residues that possessed the most increasing amount of accessible surfaces are: Met1 (≥70%); Val5, Ser6, Pro8 (40–50%); Asp2, Gln4, Tyr10, Asp11, Ser17 (30–40%); and Tyr3, Ser7, Ile9, Tyr14, Tyr15, Gln30 (20–30%). Some of these amino acids are located in unfavorable environment because their hydrophobic side chains are exposed to solvent and, therefore, tend to change polarity of their microenvironment to the more favorable one via bind to chemokines or virus, or adsorb on other part of extracellular domain again. On the other hand, the exposure of negatively charged amino acids increases their opportunities to interact with positively charged surface of ligands or gp120 [42,43]. Thus, we speculate the Nt region behaves like an antenna in vivo, which is able to feel their microenvironment and then recognize or interact with its ligands, or, it can be used by HIV-1 envelope glycoprotein.

ECL2 contains 32 amino acid residues, out of them there are one negative charge and six positive charges. By using CCR5-CCR2b chimeras, Samson et al. [18] have shown that the ECL2 of CCR5 is the major determinant for chemokine binding specificity because the introduction of CCR5 ECL2 into a CCR2b background was able to confer high affinity binding of MIP-1α, as well as functional responses to MIP-1β and RANTES. In addition, a number of point mutations for the charged residues within the ECL2 loop were found to dramatically affect CCR5 function [22,44]. Moreover, through mAbs competition binding experiments, Lee et al. [23] found ECL2-specific mAbs were more efficient in blocking chemokine binding than mAbs directed to other extracellular segments, and Nt mAbs blocked gp120-CCR5 binding more effectively than ECL2 mAbs, surprisingly, ECL2 mAbs were more potent inhibitors of viral infection than Nt mAbs. These results imply that the binding sites of chemokines and gp120 on CCR5 are distinct but overlapping, and suggest that Nt is more important for gp120 binding while the ECL2 is more important for inducing conformational changes in envelope glycoprotein that lead to membrane fusion and virus entry. In this study, the MD simulation showed that the ECL2 is the most flexible ECL among the ECLs 1–3. Although the interactions of ECL2 with other ECLs and Nt restrain its conformational changes to a certain spatial range, the stronger conformational flexibility provides more motion freedom for ECL2 than for other ECLs, which endow ECL2 more opportunities to involve in ligand bindings. The calculation of solvent-accessible surfaces of positively charged residues for 480 frames indicate that R168, K171 and K191 usually expose more than 60% surface areas to solvent, moreover, the side chains of residues169–174 and 192–195 often pointed away from the backbone of ECL2, such conformational state seems to be advantageous for ECL2 to interact with natural ligands or gp120. Interestingly, the long-range hydrogen bonds, Van der Waal contacts and electrostatic interactions between Nt and ECL2 could be often observed in the 1 nm MD process, and the Coulomb force between the negatively charged residues in Nt and positively charged residue in ECL2 could draw Nt close to ECL2. We speculated these interactions might be involved in the functional roles of reciprocal communications and cooperation between Nt and ECL2 in chemokines bindings, gp120 binding and Evn-mediated membrane fusion.

5. Conclusion

In summary, through a high temperature and long time MD simulation, we found that the extracellular domain of CCR5 adopted a compact globular conformational state as a whole, and such conformation was maintained by a complex interaction network formed by disulfide bonds, hydrogen bonds, Van der Waals force and salt bridges between extracellular segments. Particularly, the disulfide linkage of Cys20–Cys269 between Nt and ECL3 is essential in maintaining appropriate conformational state of Nt 1–19 region that is necessary for chemokines and gp120 binding. In addition, the RMS comparisons indicate that Nt 1–19 and

ECL2 are both of the most flexible protein parts of CCR5, and the solvent-accessible surface area calculations reveal that large numbers of residues in Nt and ECL2 are exposed at receptor surface. Integrating these results with available experimental data permits us to put forward a two-step gp120-binding mechanism: the Nt 1–19 region first projects from the extracellular domain and adopts an appropriate conformational state being ready to recognize envelope glycoprotein; then, the binding of gp120 to Nt induces the conformational change of Nt, which makes gp120 further interact with ECL2 of CCR5; finally, the interactions occurred between ECL2 and gp120 induce conformational changes in envelope glycoprotein and lead to membrane fusion and virus entry. Here we should emphasize that the conformational changes of Nt and ECL2, as well as the interactions between CCR5 and gp120, are a concerted and dynamic process. In the absence of experimental determined 3D structure of CCR5, our 2SS CCR5 models derived from MD explained the structural basis of available biological data, and the two-step gp120-binding mechanism may aid the design of selective inhibitors that would specially antagonize HIV-1 coreceptor activity without impairing the natural ligands binding abilities. However, to further investigate how gp120 binds to CCR5 Nt region and how ECL2 interacts with gp120 leading to membrane fusion, the studies of docking gp120-CD4 complex to the CCR5 and building the complex models of CCR5-gp120 need to be done.

Acknowledgements

References

[1] P.M. Murphy, Annu. Rev. Immunol. 12 (1994) 593.
[2] M. Samson, O. Labbe, C. Mollereau, G. Vassart, M. Parmentier, Biochemistry 35 (1996) 3362.
[3] B. Lee, M. Sharron, L.J. Montaner, D. Weissman, R.W. Doms, Proc. Natl Acad. Sci. USA 96 (1999) 5215.
[4] R. Bonecchi, G. Bianchi, P.P. Bordignon, R. Lang, A. Borsatti, S. Sozzani, P. Allavena, P.A. Gray, A. Mantovani, F. Snigaglia, J. Exp. Med. 187 (1998) 129.
[5] F. Sallusto, D. Lenig, C.R. Mackay, A. Lanzavecchia, J. Exp. Med. 187 (1998) 875.
[6] G. Alkhatib, C. Combadiere, C.C. Broder, Y. Feng, P.E. Kennedy, P.M. Murphy, E.A. Berger, Science 272 (1996) 1955.
[7] A. Trkola, T. Dragic, J. Arthos, J.M. Binley, W.C. Olson, G.P. Allaway, C. Cheng-Mayer, J. Robinson, P.J. Maddon, J.P. Moore, Nature 384 (1996) 184.
[8] L. Wu, N.P. Gerard, R. Wyatt, H. Choe, C. Parolin, N. Ruffing, A. Borsetti, A.A. Cardoso, E. Desjardin, W. Newman, C. Gerard, J. Sodroski, Nature 384 (1996) 179.
[9] M. Dean, M. Carrington, C. Winkler, G.A. Huttley, M.W. Smith, R. Allikmets, J.J. Goedert, S.P. Buchbinder, E. Vittinghoff, E. Gomperts, S. Donfield, D. Vlahov, R. Kaslow, A. Saah, C. Rinaldo, R. Detels, S.J. O'Brien, Science 273 (1996) 1856.
[10] M. Samson, F. Libert, B.J. Doranz, J. Rucker, C. Liesnard, C.M. Farber, S. Saragosti, C. Lapoumeroulie, J. Cogniaux, C. Forceille, G. Muyldermans, C. Verhofstede, G. Burtonboy, M. Georges, T. Imai, S. Rana, Y. Yi, R.J. Smyth, R.G. Collman, R.W. Doms, G. Vassart, M. Parmentier, Nature 382 (1996) 722.
[11] J. Eugen-Olsen, A.K. Iverson, P. Garred, U. Koppelhus, C. Pedersen, T.L. Benfield, A.M. Sorensen, T. Katzenstein, E. Dickmeiss, J. Gerstoft, P. Skinhoj, A. Svejgaard, J.O. Nielsen, B. Hofmann, AIDS 11 (1997) 305.
[12] W.A. Paxton, R. Liu, S. Kang, L. Wu, T.R. Gingeras, N.R. Landau, C.R. Mackay, Virology 244 (1998) 66.
[13] L. Wu, W.A. Paxton, N. Kassam, N. Ruffing, J.B. Rottman, N. Sullivan, H. Choe, J. Sodroski, W. Newman, R.A. Koup, C.R. Mackay, J. Exp. Med. 185 (1997) 1681.
[14] J. Gosling, F.S. Monteclaro, R.E. Atchison, H. Arai, C.L. Tsou, M.A. Goldsmith, I.F. Charo, Proc. Natl Acad. Sci. USA 94 (1997) 5061.
[15] R.E. Atchison, J. Gosling, F.S. Monteclaro, C. Franci, L. Digilio, I.F. Charo, M.A. Goldsmith, Science 274 (1996) 1924.
[16] L. Picard, G. Simmons, C.A. Power, A. Meyer, R.A. Weiss, P.R. Clapham, J. Virol. 71 (1997) 5003.
[17] G. Alkhatib, S.S. Ahuja, D. Light, S. Mummidi, E.A. Berger, S.K. Ahuja, J. Biol. Chem. 272 (1997) 19771.
[18] M. Samson, G. LaRosa, F. Libert, P. Paindavoine, M. Detheux, G. Vassart, J. Biol. Chem. 272 (1997) 24934.
[19] T. Dragic, A. Trkola, S.W. Lin, K.A. Nagashima, F. Kajumo, L. Zhao, W.C. Olson, L. Wu, C.R. Mackay, G.P. Allaway, T.P. Sakmar, J.P. Moore, P.J. Maddon, J. Virol. 72 (1998) 279.
[20] G.E. Rabut, J.A. Konner, F. Kajumo, J.P. Moore, T. Dragic, J. Virol. 72 (1998) 3464.
[21] M. Farzan, H. Choe, L. Vaca, K. Martin, Y. Sun, E. Desjardins, N. Ruffing, L. Wu, R. Wyatt, N. Gerard, C. Gerard, J. Sodroski, J. Virol. 72 (1998) 1160.
[22] J.M. Navenot, Z.X. Wang, J.O. Trent, J.L. Murray, Q.X. Hu, L. DeLeeuw, P.S. Moore, Y. Chang, S.C. Peiper, J. Mol. Biol. 313 (2001) 1181.
[23] B. Lee, M. Sharron, C. Blanpain, B.J. Doranz, J. Vakili, P. Setoh, E. Berg, G. Liu, H.R. Guy, S.R. Durell, M. Parmentier, C.N. Chang, K. Price, M. Tsang, R.W. Doms, J. Biol. Chem. 274 (1999) 9617.
[24] J.E. Walker, M. Saraste, Curr. Opin. Struct. Biol. 6 (1996) 457.
[25] A. Bairoch, R. Apweiler, Nucleic Acids Res. 28 (2000) 45.
[26] K. Palczewski, T. Kumasaka, T. Hori, C.A. Behnke, H. Motoshima, B.A. Fox, I. Le Trong, D.C. Teller, T. Okada, R.E. Stenkamp, M. Yamamoto, M. Miyano, Science 289 (2000) 739.
[27] N. Guex, M.C. Peitsch, Electrophoresis 18 (1997) 2714.
[28] C.J. Raport, J. Gosling, V.L. Schweickart, P.W. Gray, I.F. Charo, J. Biol. Chem. 271 (1996) 17161.
[29] T.H. Ji, M. Grossmann, I. Ji, J. Biol. Chem. 273 (1998) 7299.
[30] S. Genoud, F. Kajumo, Y. Guo, D. Thompson, T. Dragic, J. Virol. 73 (1999) 1645.
[31] C. Blanpain, B. Lee, J. Vakili, B.J. Doranz, C. Govaerts, I. Migeotte, M. Sharron, V. Dupriezi, G. Vassart, R.W. Doms, M. Parmentier, J. Biol. Chem. 274 (1999) 18902.
[32] R. Sahchez, A. Sali, J. Comput. Phys. 151 (1999) 388.
[33] F.S. Cordes, N. Bright, M.S.P. Sansom, J. Mol. Biol. 323 (2002) 951.
[34] R.A.F. Reithmeier, Curr. Opin. Struct. Biol. 5 (1995) 491.
[35] R. Efremov, F. Legret, G. Vergoten, A. Capron, G.M. Bahr, A.S. Arseniev, J. Biomol. Struct. Dyn. 16 (1998) 77.
[36] N. Grigorieff, T.A. Ceska, K.H. Downing, J.M. Baldwin, R. Henderson, J. Mol. Biol. 256 (1996) 393.
[37] N.M. Zhou, Z.W. Luo, J.W. Hall, J.S. Luo, X.B. Han, Z.W. Huang, Eur. J. Immunol. 30 (2000) 164.
[38] M.G. Paterlini, Biophys. J. 83 (2002) 3012.

[39] R. Horuk, TIPS 15 (1994) 159.

[40] B. Blanpain, B.J. Doranz, J. Vakili, J. Rucker, C. Govaerts, S.S.W. Baikc, O. Lorthioir, I. Migeotte, F. Libert, F. Baleux, G. Vassart, R.W. Doms, M. Parmentier, J. Biol. Chem. 274 (1999) 34719.

[41] S.S. Baik, R.W. Doms, B.J. Doranz, Virology 259 (1999) 267.

[42] P.D. Kwong, R. Wyatt, J. Robinson, R.W. Sweet, J. Sodroski, W.A. Hendrickson, Nature 393 (1998) 648.

[43] C.D. Rizzuto, R. Wyatt, N. Hernandez-Ramos, Y. Sun, P.D. Kwong, W.A. Hendrickson, J.A. Sodroski, Science 280 (1998) 1949.

[44] T.N. Wells, C.A. Power, A.E. Proudfoot, Trends Pharmacol. Sci. 19 (1998) 376.

[45] L.A. Kelley, R.M. MacCallum, M.J.E. Sternberg, J. Mol. Biol. 299 (2000) 499.

[46] P.J. Kraulis, J. Appl. Crystollogr. 24 (1991) 946.