

## Structure Modeling of the Chemokine Receptor CCR5: Implications for Ligand Binding and Selectivity

M. Germana Paterlini

Department of Medicinal Chemistry and Supercomputer Institute, University of Minnesota, Minneapolis, Minnesota 55455 USA

**ABSTRACT** The G-protein coupled receptor CCR5 is the main co-receptor for macrophage-tropic HIV-1 strains. I have built a structural model of the chemokine receptor CCR5 and used it to explain the binding and selectivity of the antagonist TAK779. Models of the extracellular (EC) domains of CCR5 have been constructed and used to rationalize current biological data on the binding of HIV-1 and chemokines. Residues spanning the transmembrane region of CCR5 have been modeled after rhodopsin, and their functional significance examined using the evolutionary trace method. The receptor cavity shares six residues with CC-chemokine receptors CCR1 through CCR4, while seven residues are unique to CCR5. The contribution of these residues to ligand binding and selectivity is tested by molecular docking simulations of TAK779 to CCR1, CCR2, and CCR5. TAK779 binds to CCR5 in the cavity formed by helices 1, 2, 3, and 7 with additional interactions with helices 5 and 6. TAK779 did not dock to either CCR1 or CCR2. The results are consistent with current site-directed mutagenesis data and with the observed selectivity of TAK779 for CCR5 over CCR1 and CCR2. The specific residues responsible for the observed selectivity are identified. The four EC regions of CCR5 have been modeled using constrained simulated annealing simulations. Applied dihedral angle constraints are representative of the secondary structure propensities of these regions. Tertiary interactions, in the form of distance constraints, are generated from available epitope mapping data. Analysis of the 250 simulated structures provides new insights to the design of experiments aimed at determining residue-residue contacts across the EC domains and for mapping CC-chemokines on the surface of the EC domains.

### INTRODUCTION

A critical step in cellular entry by HIV-1 is the binding of the viral envelope protein gp120 to chemokine receptors on surface of immune cells (Hoffman et al., 1999; Kolchinsky et al., 1999; Rizzuto et al., 1998; Wu et al., 1996). The CC-chemokine receptors CCR5 and CXCR4 are the major co-receptors for R5 and X4 virus strains, respectively (Berger et al., 1998; Robertson et al., 2000). CCR5 has been the target of anti-HIV-1 strategies for disrupting the interaction with the HIV-1 envelope protein gp120. Studies suggest that gp120 binds preferentially to the N-terminus of CCR5 (Dragic et al., 1998; Farzan et al., 1998), and that the second extracellular region is mostly responsible for the binding of the endogenous chemokines peptides (Samson et al., 1997). The two domains are not clearly separated, because some single point mutations in the N-terminus can also abolish the binding of chemokines (Blanpain et al., 1999a). HIV-1 and chemokines mainly interact with the extracellular (EC) regions of CCR5, while small synthetic ligands mostly bind to residues of the transmembrane (TM) region (Dragic et al., 2000).

Chemokine receptors belong to the rhodopsin family of G-protein coupled receptors (GPCR) characterized by a heptapeptidic helical fold (7-TM) spanning the plasma membrane (Fig. 1). Modeling of GPCR has at its disposal

the 2.8 Å resolution structure of bovine rhodopsin (Palczewski et al., 2000) whose 7-TM motif has been proven valid for other receptors in its family as well (Elling et al., 1997; Mizobe et al., 1996). However, local structural motifs, such as those seen in helices of rhodopsin that contain Gly and Pro (Palczewski et al., 2000), need to be reexamined in the context of CCR5. Given the relative simplicity of the helical fold and ~30% sequence identity with rhodopsin in the transmembrane core region, it should be possible to generate models where ~80% of the C $\alpha$  atoms are within 3.5 Å of their correct positions (Sanchez and Sali, 1999). Protein models obtained at such resolution have been able to correctly predict the location of binding sites and the size of the ligands (Sanchez and Sali, 1999). To this extent, theoretical GPCR models based on a low-resolution template (Unger et al., 1997; Baldwin et al., 1997) before the x-ray structure have been successfully used in designing experiments for structure validation (Lu and Hulme, 2000; Mizobe et al., 1996; Zhou et al., 1994), determining ligand binding sites (Metzger et al., 1996; Simpson et al., 1999; Strader et al., 1994), and formulating hypotheses on mechanisms of receptor activation (Ballesteros et al., 1998; Cotecchia et al., 2000; Flanagan et al., 1999; Javitch et al., 1997, 1998).

I have used the x-ray structure of rhodopsin, in combination with homology techniques, to derive a structural model of the transmembrane region (TM) of CCR5, and then applied the evolutionary trace method (Lichtarge et al., 1996) to compare the residues of the receptor cavity with those of either distantly or closely related homologs. Evolutionary tracing has been successful in identifying functionally important residues across different protein families (Johnson and Church, 2000; Sowa et al., 2000). When

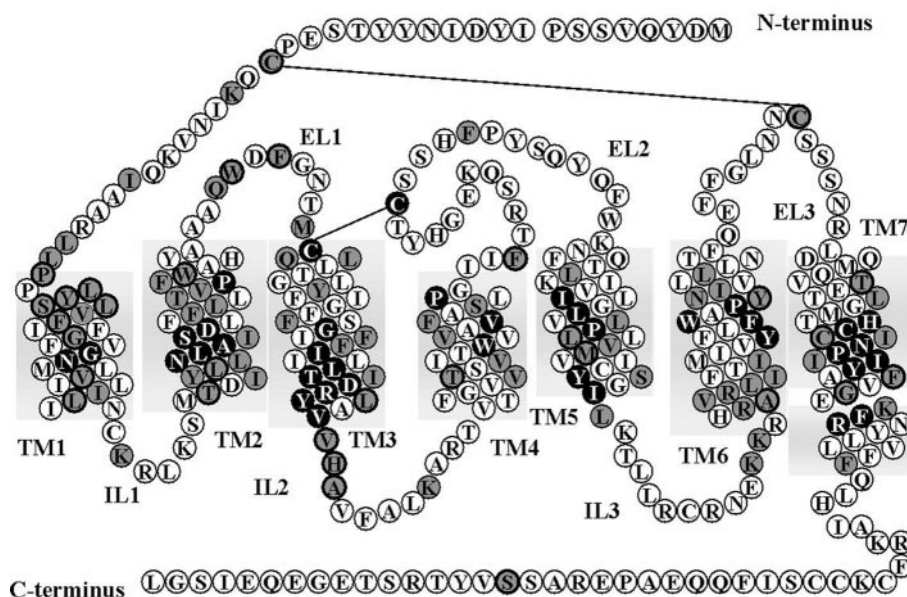
*Submitted February 7, 2002, and accepted for publication September 10, 2002.*

Address reprint requests to M. Germana Paterlini, Certusoft Inc., 7831 Glenroy Road, Suite 440, Minneapolis, MN 55439. Tel.: 952-921-0351; Fax: 952-921-0367; E-mail: germana@certusoft.com.

© 2002 by the Biophysical Society

0006-3495/02/12/3012/20 \$2.00

FIGURE 1 Schematic representation of the CCR5 sequence. Gray rectangles outline the residues in the 7-TM region, TM1 through TM7, and helix 8. EL and IL denote extracellular and intracellular loop regions, respectively. Disulfide bridges between C20 and C269 and C101 and C178 are shown as lines connecting these cysteines. Gray circles denote conservation of strong groups in CCR1 through CCR5. Gray circles with heavy outlines denote identical residues in CCR1 through CCR5. Black circles denote highly conserved residues in the rhodopsin family of GPCR. For ease of comparison with other GPCR, residues are numbered using the highly conserved residues as reference (Ballesteros and Weinstein, 1995). N: 1.50 (TM1); D: 2.50 (TM2); R: 3.50 (TM3); W: 4.50 (TM4); P: 5.50 (TM5); P: 6.50 (TM6); P: 7.50 (TM7).



applied to CCR5, the tracing readily isolates residues of the receptor cavity that can either bind small synthetic ligands or function as selectivity filters among closely related CC-chemokine receptors. To test this hypothesis, I have docked the antagonist TAK779 (Baba et al., 1999) to CCR5 and compared its binding mode with available experimental data. The selectivity of TAK779 for CCR5 has been previously measured against closely related receptors and the binding site has been probed using site-directed mutagenesis (Baba et al., 1999; Dragic et al., 2000). Simulations have been carried out using an automated docking protocol (Meng et al., 1992) used in docking simulations of opioid ligands (Subramanian et al., 1998, 2000). In that work, the docking procedure, in conjunction with mutagenesis data, provided binding modes in agreement with structure-activity relationship (SAR) of opioid ligands and explained, in part, the observed selectivity. In this work I show that the docking simulations not only are consistent with the mutagenesis data, but also suggest a rationale for the selectivity of TAK779 for CCR5, as compared with the chemokine receptors CCR1 and CCR2. The results show a conserved region of the binding pocket that may function as a binding locus for chemokine antagonists that share a common ammonium group. Nonconserved residues in helices TM3, TM5, and TM6 would impart receptor selectivity. The docking results outline specific residues that can be targeted by site-directed mutagenesis studies for validation of the proposed binding mode.

The importance of the extracellular (EC) regions in binding and recognition of chemokines and the HIV-1 envelope glycoprotein gp120 makes it paramount to include such regions in the modeling effort. Close proximity among the EC domains is likely, due to the presence of two disulfide

bonds, one linking the N-terminus to the third extracellular loop (EL), EL3, and a second between EL1 and EL2 (Blanpain et al., 1999b) (Fig. 1). Indirect information on the 3-D arrangement of the extracellular regions can be inferred from the mapping of epitope residues recognized by monoclonal antibodies (mAbs). The antibody mAb-2D7 recognizes residues in both EL1 and EL2 (Lee et al., 1999), while mAbs PA9 and PA14 bind to amino acids in both the N-terminus and EL2 (Olson et al., 1999).

Modeling of GPCR has been mostly applied to residues in the transmembrane region, and considerably less work has been done on the N-terminus and loops connecting the TM helices (Luo et al., 1997; Paterlini et al., 1997; Moro et al., 1999; Odile-Colson et al., 1998; Pogozheva et al., 1998). Modeling of the extracellular regions of GPCR is challenging because of limited homology with known structures and of limitations of current loop modeling techniques. The combination of low sequence identity with rhodopsin and short loop length makes it difficult to match the EC regions with homologous segments of known structure using database search programs (Altschul et al., 1997). The three extracellular loops are generally 10–30 residues in length, and thus behind the applicability of current loop modeling tools (Van Vlijmen and Karplus, 1997; Li et al., 1999; Fiser et al., 2000). Loop modeling is typically applied to isolated loops of globular proteins, while loops of GPCR most likely contain numerous tertiary interactions, as demonstrated by rhodopsin. The tertiary fold of a protein is difficult to predict computationally without the aid of geometrical constraints (Zhang et al., 2002a). In the case of membrane proteins, distance constraints are usually obtained using EPR and fluorescence spectroscopies, or by disulfide bond engineering. For example, EPR spectroscopy has been ex-

tensively used to map interactions among the three intracellular loops of rhodopsin (Hubbell et al., 2000). Spectroscopy-derived distances are not available for CCR5. However, indirect distance information can be inferred from epitope mapping results. A previous study on the sensitivity of mAbs to the tertiary structure of their targets has shown that antibodies recognize a region within a 15 Å radius of the epitope residues (Burrill et al., 1998). I have, therefore, transformed the epitope mapping data of CCR5 into distance constraints between the different EC regions to create tertiary contacts among the EC domains.

I have combined homology information, tertiary constraints derived from available experimental data, and ab initio computational tools to model the 95 residues of the extracellular regions of CCR5, which vary in length from a 32-residue-long N-terminus to the 13 residues of EL1. Secondary structure prediction methods were first used to obtain secondary structure propensities of the individual extracellular region. Additional searches for homologous sequences in the PDB protein data bank (Berman et al., 2000) provided validation of the prediction results and initial structural templates for each of the four domains. Conformations of the extracellular regions generated in this fashion were subjected to constrained simulated annealing simulations and analyzed based on conformational energy, structural variability, and average physical characteristics.

The addition of tertiary constraints results in clusters of critical aromatic and acidic residues previously implicated in the binding of chemokines and gp120 (Blanpain et al., 1999a; Howard et al., 1999; Rabut et al., 1998; Zhou et al., 2000). The EC models are used to explain the sensitivity of these residues to mutation and suggest specific residue-residue interactions that could be validated by site-directed mutagenesis and sites for mapping the interaction of the chemokines with CCR5.

## MATERIALS AND METHODS

### CCR5 receptor model

The CCR5 sequence (Accession number P51681) was scored against the SWISSPROT database (Bairoch and Apweiler, 2000) using BLAST2.0 (Altschul et al., 1997) with the BLOSUM62 matrix and standard parameters in the Biology Workbench (<http://workbench.sdsc.edu>) (Subramaniam, 1998). Sequences were aligned using CLUSTALW (Thompson et al., 1994). The suite of programs PERSCAN (Donnelly et al., 1994) was used to identify residues located in the transmembrane from a set of 50 nonredundant sequences with the highest homology to CCR5 (see Table 1). Sequence identity of the 50 sequences with CCR5 was between 28% and 99% for residues in the transmembrane region.

The length of the CCR5 helices was determined by first considering a minimum length of 18 residues in the core of the lipid bilayer. Up to seven additional residues at each end were then added to allow extension of the helices into the lipid polar head regions (White, 1994). Criteria for finding the end of the helices were based either on the  $\alpha$ -helical index (AP index) (Donnelly et al., 1994), calculated using the 50 sequences in Table 1, or on sequence analysis of these regions.

Criteria from sequence analysis, homology to rhodopsin, and the presence of either Gly or Pro near the ends were used to determine the length

of the TM helices. Glycine and proline residues are indicators of helix termination or initiation in globular helices (Prieto and Serrano, 1997; Viguera and Serrano, 1999) and have a similar function in rhodopsin as well (Palczewski et al., 2000). Residues modeled in the  $\alpha$ -helical conformation are shown in Fig. 1.

The set of 50 unique sequences with the highest sequence homology to CCR5 (Table 1) was also used to compare the orientation of the 7-TM bundle of CCR5 with that of the average template (Baldwin et al., 1997) and of rhodopsin. Calculations of the helix orientations using criteria based on the periodicity of conserved residues (Donnelly et al., 1994) were within 20° of those obtained using periodicity of substitutions (Overington et al., 1992; Donnelly et al., 1994).

Structural homologs of sequences in the EC regions were obtained by searching the PDB database (Berman et al., 2000) using BLASTP2.0 or PSI-BLAST with either PAM30 or PAM70 matrices. Expectation values of the hits varied from  $E = 0.0002$  to  $E = 107$ . The searches produced from 3 to 12 structures for each sequence searched. Sequences found in this fashion were then used in MODELLER4 (Sali and Blundell, 1993) using the "model" routine to generate initial loop structures. Structures 11RK (residues 1150–1163), 1PFC (residues 390–401), 1UWB (residues 274: B-283:B), and 2BPA (residues 120:2–131:2) were used as templates for the N-terminus. Initial conformations of EL1 made use of structures 1TIA (residues 12–19), 1BGY (residues 31:O-50:O) and 1TMF (residues 206–212 of chain 1). Structures used for EL2 were 1AVQ (residues 178–186 of chain A), 1HRA (residues 32–34), and 2T51 (residues 248–257). Structures used for EL3 were 1ARP (residues 112–128), 17AJ (residues 266–273) and 2MEV (residues 140–146). Intracellular loop regions were built using the Search Loop module of the InsightII program (InsightII, 2002. Accelrys Inc., San Diego, CA).

Prediction of the secondary structure of the loops residues was obtained using the protein prediction tool PELE, as provided by the Biology Workbench (Subramaniam, 1998). The method compares predictions from seven major algorithms and determines secondary structure based on the minimum consensus for each amino acid.

Side chain rotamers were generated using a backbone-dependent rotamer procedure (Dunbrack and Karplus, 1993). The AMBER5.0 (Case et al., 1995) suite of programs was used for energy minimization and simulated annealing simulations. The model of the transmembrane region was subjected to energy minimization followed by 500 ps of molecular dynamics (MD) simulation at 300 K, during which positional restraints applied to backbone atoms were gradually lowered from 2 to 0.1 kcal/mol-Å<sup>2</sup>. The structural quality of the model was checked using PROCHECK with 94% of the residues in the most favored allowed helical region and remaining 6% in the additionally allowed helical region (Laskowski et al., 1993). Side chain  $\chi_1$  and  $\chi_2$  angles were found in their most favorable regions without stereochemical conflicts.

Structural models of the TM regions of receptors CCR2 and CCR1 were constructed from the CCR5 template by replacing the nonconserved side chains using SQWRL (Dunbrack and Karplus, 1993), followed by energy minimization.

Side chain accessibility areas of the amino acids were calculated with the program ACCESS as a percentage compared with that measured for that residue in an extended Ala-Xxx-Ala tripeptide (Hubbard and Thornton, 1993).

Structures of the EC regions were generated using a simulated annealing procedure consisting of fast heating to 1200 K, cooling to 300 K up to 30 ps, and constant temperature for an additional 70 ps, followed by energy minimization. The backbone dihedral angle  $\phi$  (C-N-C <sub>$\alpha$ -C) of all residues except Gly were restricted to negative values during annealing and additional constraints were imposed to maintain trans- $\omega$  bonds for non-Pro residues and the correct chirality of the C <sub>$\alpha$  carbons. Constraints made use of a flat well potential function with a force constant of 32 kcal/mol/deg<sup>2</sup> for angle constraints and 10 kcal/mol-Å<sup>2</sup> for distance constraints. A similar protocol was used to generate a structure of the intracellular loops. Residues in the transmembrane region were kept in their initial conformation using the belly option of SANDER during the annealing procedure. A set</sub></sub>

of 250 structures of the EC regions was obtained in this fashion. Clustering of structures based on RMSD was obtained using PADRE (Stahl, M. T., and W. P. Walters. 1995. PADRE. Population analysis and duplicate removal (available by ftp from ccl.osc.edu).

## Modeling of TAK779

The geometry of TAK779 was optimized using molecular orbital *ab initio* methods at the HF/6-31G\* level using the Gaussian98 program package (Frisch et al., 2001). Partial atomic charges were obtained using the restrained electrostatic potential (RESP) charge fitting formalism (Bayly et al., 1993). Force field parameters were adapted from similar chemical groups found in the Cornell et al. (1995) force field (parm96.dat of AMBER5.0).

## Ligand docking

Docking of the minimum energy structures of TAK779 to the model-built TM domain of CCR5, CCR1, and CCR2 was performed using the automated docking procedure DOCK3.5 (Meng et al., 1992) by following a previously described protocol (Subramanian et al., 1998). The docking cavity was generated using SPHGEN by generating spheres from the solvent-accessible molecular surface of the receptor cavity. Clusters containing 121 spheres for CCR5, 90 for CCR1, and 74 for CCR2 were generated in this fashion. A maximum of six steric overlaps were allowed during the generation of docking orientations. The predicted orientations were scored individually based on the empirical evaluation of the electrostatic and van der Waals energy contributions.

Refinement of the initial receptor-ligand complex was obtained in *vacuo* energy minimization (0.001 kcal/mol rms deviation) followed by MD simulations up to 1 ns. A 3.0 kcal/mol positional constraint was applied to the receptor backbone atoms, the cutoff for nonbonded interactions was 8 Å, and the dielectric constant was 4. The temperature of the system was maintained at 298 K with a 0.2 ps coupling constant.

## RESULTS

### Modeling and characterization of the transmembrane region of CCR5

The model of the transmembrane region of CCR5 is based on the x-ray structure of rhodopsin (Palczewski et al., 2000) and makes use of a previous sequence analysis of 493 GPCR (Baldwin et al., 1997) to determine the average orientation of the helices in the 7-TM bundle with respect to the lipid environment and to outline possible discrepancies between rhodopsin and CCR5. Homology between CCR5 and rhodopsin is lower near the extracellular medium, and variations in these regions may result in a different length and/or different orientation of the helices at these ends. The sequence of CCR5 is schematically shown in Fig. 1, where highly conserved residues are numbered according to the convention of Ballesteros and Weinstein (1995). A set of 50 sequences with the highest sequence identity to CCR5 was used in the comparison (Table 1).

Calculations of the lipid-facing direction of helices TM1, TM3, TM4, and TM6 were in good agreement with those obtained from a larger set and from the x-ray structure of rhodopsin. The orientation of residues at the extracellular end of TM2 and TM5, while in agreement with that ob-

tained by Baldwin et al. (1997), deviated by  $\sim 100^\circ$  from rhodopsin. This discrepancy originates from bulges in TM2 and TM5 that alter the helical periodicity, thus exposing residues to the lipid environment that would otherwise face the bundle's interior. The bulge in TM5 of rhodopsin is near a conserved proline, flanked by large hydrophobic residues in the F<sub>212</sub>IIP<sub>LIV</sub><sub>218</sub> region. The corresponding CCR5 fragment, L<sub>203</sub>VL<sub>PLL</sub>V<sub>209</sub>, has a similar amino acid composition and is likely to adopt a similar conformation. By using the TM5-rhodopsin template, residues K191 and T195 of TM5 orient toward residue T259 of TM6. They would not face each other if TM5 is modeled as an ideal helix. Previous studies have shown that residues at these three positions are proximal to each other, as they can form interhelical disulfide bonds when mutated to Cys in rhodopsin and the tachykinin NK1 receptors (Elling et al., 2000; Struthers et al., 1999). Similarly, replacement with His creates a zinc binding site in both the NK1 and kappa opioid receptors (Thirstrup et al., 1996; Elling et al., 1997). The NK1, kappa opioid, and CCR5 receptors all have an amino acid composition similar to rhodopsin near the conserved Pro. Helix 5 of CCR5 was therefore modeled using the rhodopsin template.

Replacement of the rhodopsin side chains with those of CCR5 resulted in steric conflict between residues T82 and V83 of TM2 and L107 in TM3. Furthermore, the orientation of the C-terminal residues of TM2 did not agree with that predicted by the Fourier transform analysis of the 50 helices or with the previous sequence analysis of GPCR (Baldwin et al., 1997). Closer analysis of helix 2 showed that T82 and V83 replace residues G89 and G90 of rhodopsin at a location where TM2 and TM3 cross and interact with the closest packing. This type of interaction, where large side chains in one helix pack against the Gly backbone of the other helix, is often found in membrane proteins (Javadpour et al., 1999). In rhodopsin, packing is achieved by a local distortion of the helical conformation at these sites. Replacement of the two glycines with the larger CCR5 side chains within this structural motif results in atomic overlap. The extracellular portion of TM2 was therefore modeled by changing the dihedral angles of the L<sub>80</sub>-P<sub>84</sub> fragment according to average values observed in Pro-containing  $\alpha$ -helices (Némety et al., 1992; Sankararamkrishnan and Vishveshwara, 1992). The changes in the torsion angles relieved the atomic overlap between TM2 and TM3 and resulted in an orientation of the C-end residues of TM2, in agreement with both the Fourier analysis and the previous sequence analysis of GPCR (Baldwin et al., 1997). A comparison of dihedral angles in this region between CCR5 and rhodopsin is given in Table 2. The modeled TM2 helix is likely to be valid for those GPCRs that contain the highly conserved TM2-Pro and lack the Gly-Gly motif. Templates for helix 7 and its adjoining helix 8 were taken from the x-ray structure of rhodopsin, based on high sequence homology in both helical regions. Helix 7 of rhodopsin is irregular, containing four residues in the 3<sub>10</sub>

**TABLE 1 G-protein coupled receptor with high sequence homology to human CCR5**

Receptor	% Identity*	Database Reference†
Chemokines		
CCR5, human	100	CKR5_HUMAN
CCR5, rhesus macaque	87	CKR5_MACMU
CCR5, sooty mangabey	87	CKR5_CERTO
CCR5, green monkey	86	CKR5_CERAE
CCR5, chimpanzee	98	CKR5_PANTR
CCR5, gorilla	98	CKR5_GORGO
CCR5, baboon	99	CKR5_PAPHA
CCR5, rat	84	CKR5_RAT
CCR5, mouse	84	CKR5_MOUSE
CCR1, human	66	CKR1_HUMAN
CCR1, rhesus macaque	64	CKR1_MACMU
CCR1, mouse	63	CKR1_MOUSE
CCR2, human	86	CKR2_HUMAN
CCR2, mouse	86	CKR2_MOUSE
CCR3, human	63	CKR3_HUMAN
CCR3, resus macaque	63	CKR3_MACMU
CCR3, green monkey	63	CKR3_CERAE
CCR3, mouse	68	CKR3_MOUSE
CCR4, human	60	CKR4_HUMAN
CCR4, mouse	59	CKR4_MOUSE
CCR6, human	40	CKR6_HUMAN
CCR7, human	40	CKR7_HUMAN
CCR7, mouse	40	CKR7_MOUSE
CCR8, human	49	CKR8_HUMAN
GPR13, human	50	GPRD_HUMAN
GPR13, rat	56	GPRD_RAT
GPR-96, human	43	GC96_HUMAN
VQ31, capripoxvirus	42	VQ31_CAP
CXCR2, human	40	IL8B_HUMAN
CXCR3-human	39	CCR3_HUMAN
CXCR4-human	37	CCR4_HUMAN
GPR2, human	35	GPR2_HUMAN
RDC1, human	33	RDC1_HUMAN
EBV-induced	32	EBI2_HUMAN
Bradykinin		
B2	29	BRB2_HUMAN
Angiotensin		
Type 1B, human	36	AG2S_HUMAN
Type 2, human	33	AG22_HUMAN
Type 1-A	35	AG2R_HUMAN
Type 1-B	35	AG2S_MOUSE
Orphan	32	APJ_HUMA
Opioid		
$\kappa$ , human	30	OPRK_HUMAN
$\mu$ , human	28	OPRM_HUMAN
$\delta$ , human	28	OPRD_HUMAN
Nociceptin, human	29	OPRX_HUMAN
Somatostatin		
SSR1, human	32	SSR1_HUMAN
SSR3, human	32	SSR3_HUMAN
SSR4, human	34	SSR4_HUMAN
SSR5, human	33	SSR5_HUMAN
Purinergic		
P2Y5, human	30	P2Y5_HUMAN
P2Y9, human	28	P2Y9_HUMAN

\*% Identity for 203 residues of the transmembrane region.

†SWISSPROT (Bairoch and Apweiler, 2000).

conformations, flanked by  $\alpha$ -helical segments. The model of the transmembrane region of CCR5 is shown in Fig. 2.

**TABLE 2 Backbone dihedral angle comparison between modeled CCR5 and rhodopsin at the extracellular end of the second transmembrane helix**

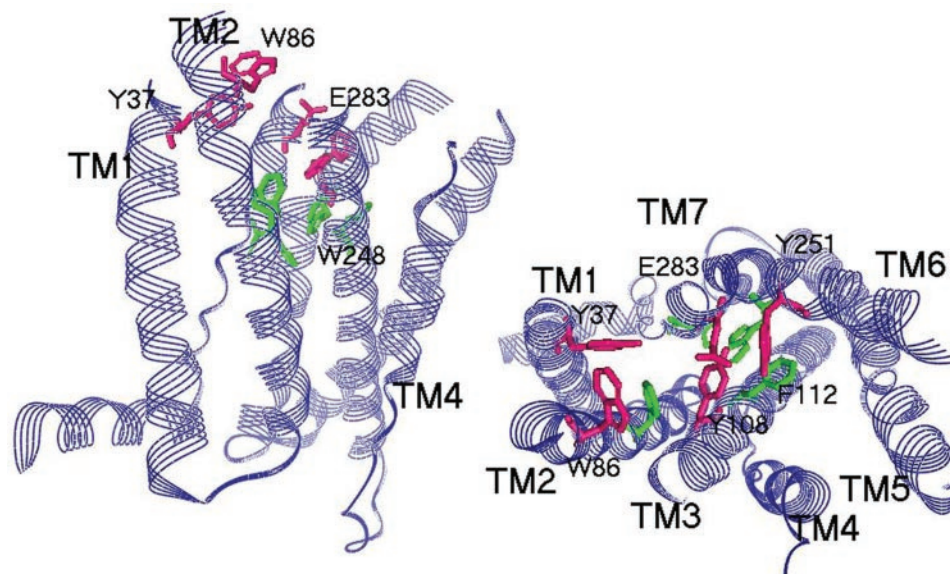
Residue	CCR5		Residue	Rhodopsin*	
	$\phi$	$\psi$		$\phi$	$\psi$
Leu-80	-63	-36	Val-87	-47	-39
Leu-81	-72	-18	Phe-88	-102	-49
Thr-82	-87	-22	Gly-89	-61	-58
Val-83	-59	-49	Gly-90	-92	-58
Pro-84	-51	-32	Phe-91	-59	-24

Angles are in degrees.

\*Palczewski et al., 2000.

The peptide subfamily of GPCR (Kolakowski, 1994) contains receptors that bind a disparate array of peptides, such as angiotensin, bradykinin, opioid, somatostatin, and chemokines. I have used evolutionary tracing (Lichtarge et al., 1996) to locate functionally important residues in this receptor family based on sequence conservation with the goal of uncovering amino acids in CCR5 that are responsible for binding and selectivity of ligands. Evolutionary trace analysis was restricted to the transmembrane regions, where the accuracy of the model template is sufficient for this type of analysis. The sequences of 50 highly homologous receptors (Table 1) were chosen for the comparison and dendrograms were constructed by including only residues in the transmembrane region. The resulting phylogenetic tree (Fig. 3 A) is similar to the one generated with the full sequence (not shown), thus suggesting a functional specificity for the TM residues. Results from the tracing were visualized using surface area accessibility plots of the TM regions (Fig. 3 B). The periodic pattern mirrors the helical periodicity, with maxima occurring at residues on the outside of the TM bundle, and minima at residues located either at helical interfaces or the interior. The plot readily identifies conserved and CCR5-specific residues of the receptor cavity. The minima observed for residues TM1-N48, TM2-D76, TM3-I116, TM5-L207, TM6-Y244, and TM7-N293 correspond to a network of interacting residues that define the bottom of the receptor cavity. Of these, N48 (1.50, using the numbering of Ballesteros and Weinstein, 1995), D76 (2.50), and N293 (7.49) are highly conserved in the rhodopsin family of GPCR (36% identity cutoff in Fig. 3 A), either Phe or Tyr is found at position Y244 (6.44), and large hydrophobic residues occupy positions I116 (3.40) and L207 (5.51) (Fig. 3 B). The trace obtained with a 76% identity cutoff shows only one additional conserved residue, TM1-Y37 (1.39). Three additional conserved residues at 92% identity, P34 (1.36), W86 (2.60), and C290 (7.46), are found in both chemokines and angiotensin receptors. The highly conserved TM6-Trp (6.48) of GPCR, W248 in CCR5, appears only in the trace at 92% identity, because of sequence variations in CCR7 and the purinergic receptor. An additional six residues are conserved in receptors CCR1 through

FIGURE 2 Structural model of the transmembrane and intracellular regions of CCR5. Shown in red are aromatic and acid residues conserved in receptors CCR1 through CCR5 (Y37 (1.39), W86 (2.60), Y108 (3.32), Y251 (6.51), and E283 (7.39)). An adjacent cluster consisting of conserved (F79 (2.53), W248 (6.48)) and CCR5-specific (H289 (7.45), F112 (3.36)) aromatic residues is shown in green. Residues Y37, W86, Y108, and E283 are part of the binding pocket for TAK-779.



CCR5 (residues L33 (1.35), Y108 (3.32), L203 (5.47), Y251 (6.51), E283 (7.39), and H289 (7.45)), and the remaining seven residues are unique to CCR5 (Fig. 3 B).

#### Automated docking simulations of TAK779 to CCR5

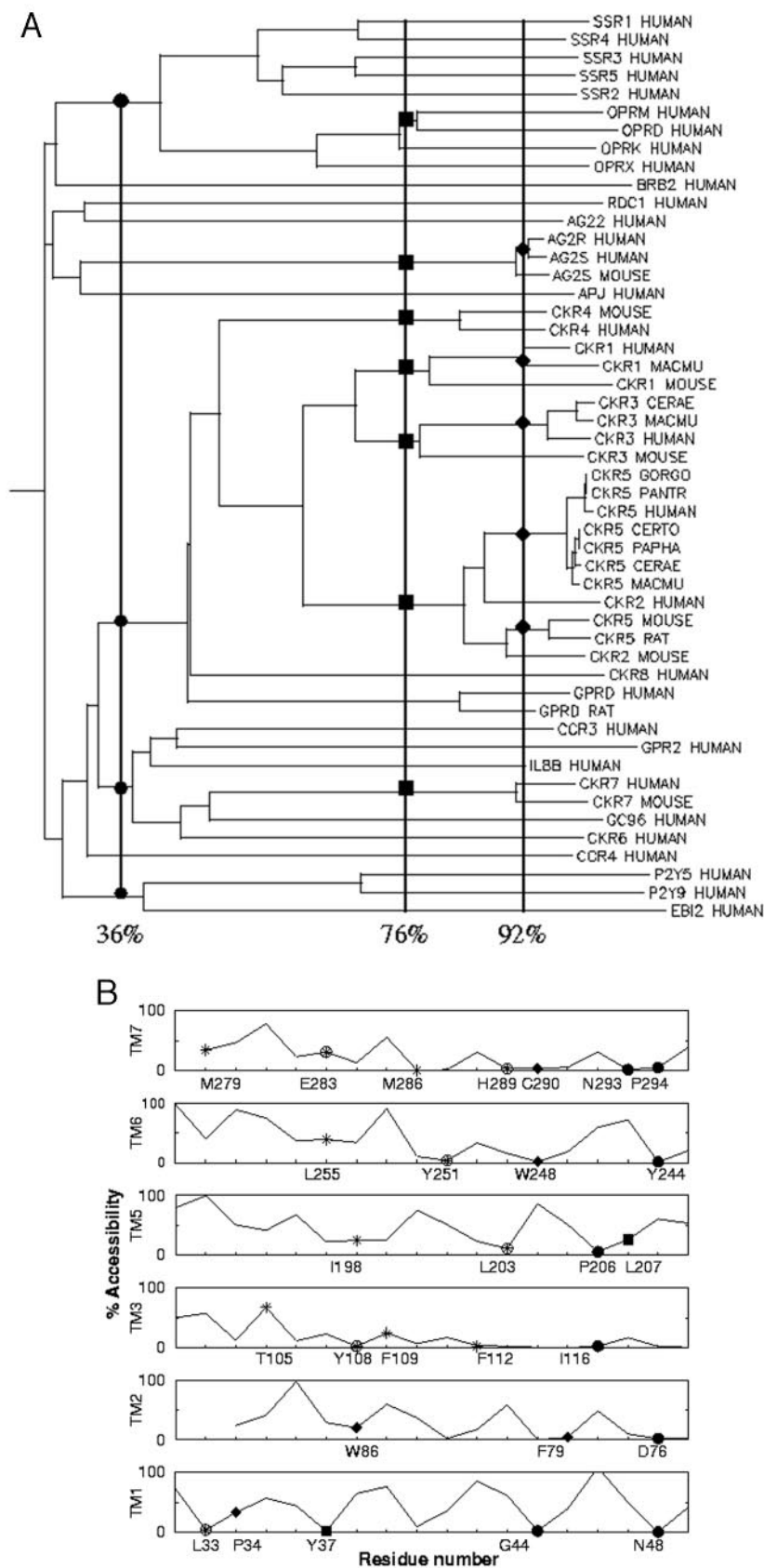
The antagonist TAK779 (Fig. 4) binds with high affinity to CCR5, but its affinity is ~20-fold lower for CCR2 and it does not bind to CCR1 (Baba et al., 1999). I have used a previous automated docking protocol (Subramanian et al., 1998, 2000) to simulate the binding of TAK779 to CCR5, CCR2, and CCR1. Computed docking orientations were sorted based on their docking scoring energy. Docking configurations of TAK779 with energies within 10 kcal/mol from the energy minimum showed interactions of the benzyl-pyran-ammonium group with helices in TM1, TM2, and TM7, and close contact of the ammonium nitrogen with E283. The methylphenyl-benzocycloheptenyl moiety mainly interacted with residues in TM3, TM5, and TM6.

A docked structure with the shortest distance between the basic nitrogen and the carboxyl oxygen atoms of E283 was selected from the ensemble of low-energy orientations and used in MD simulations to further refine the receptor-ligand complex. The resulting docking configuration (Fig. 5 and Table 3) was compared with a previous site-directed mutagenesis study of the binding of TAK779 to CCR5. Residues of the binding pocket were defined using a 5 Å cutoff from TAK779. As shown in Table 3, the docking mode contains several residues implicated in the binding of TAK779 (Dragic et al., 2000). In that study, putative binding residues were identified based on a change of 20% or higher in the efficacy of TAK779 to block HIV-1 entry. Comparison with the docking results shows that the major-

ity of such residues have been found by the docking procedure. Such residues include E283 in TM7 and neighboring aromatic residues TM1-Y37, TM2-W86, and TM3-Y108. Interactions of TAK779 with TM5 and TM6 included TM5-T195 and I198 in TM5, and Y251, N252, and L255 in TM6. By comparison, mutagenesis data suggest ligand interaction at I198 (20% change), but a negligible contribution from T195 and L255 (Table 3). Data are not available for Y251 and N252 because mutation of these two amino acids to Ala resulted in poor receptor expression (Dragic et al., 2000).

The majority of the CCR5 residues that form the binding pocket of TAK779 are also found in receptors CCR1 through CCR4, suggesting that only a few, nonconserved residues are responsible for the selectivity of TAK779. Specifically, selectivity could originate from some of the seven CCR5-unique residues identified by the trace analysis (Fig. 3 B). This hypothesis was tested by docking simulations of TAK779 to CCR2 and CCR1. Structural models of receptors CCR2 and CCR1 were constructed from the CCR5 template by replacing the nonconserved side chains using the backbone-dependent rotamer procedure of Dunbrack and Karplus (1993). Docking of TAK779 to CCR2 resulted in several orientations where the ammonium group of TAK779 was <6 Å from the conserved glutamate in TM7 (7.39). However, most of these structures lay above the TM region without making appreciable interactions with the CCR2 binding pocket. The results suggest that the transmembrane region of CCR2 cannot accommodate TAK779. Manual docking of TAK779 in an orientation similar to that of Fig. 5 resulted in atomic overlap, due to steric clash between TAK779 and R206 in TM5 and F116 and Y124 in TM3. CCR5 contains smaller amino acids at these positions. Position 5.42 is Arg (R206) in CCR2 and Ile (I198) in CCR5, while residues F116 (3.28) and Y124

FIGURE 3 Analysis of the transmembrane region of CCR5. (A) Sequence identity dendrogram of the 50 GPCR in Table 1. Only residues of the transmembrane region were included in the comparison. Sequences are denoted using their SWISSPROT database entry names. Percent identity values indicate partition identity cutoff used in the trace analysis. The dendrogram was generated using the DRAWGRAM option in the Biology Workbench (<http://workbench.sdsc.edu>). (B) Solvent accessibility surface area plots of residues in the TM region. Sequences are shown in the direction from the extracellular (*left*) to the intracellular (*right*) side. Residues that orient toward the receptor cavity are labeled according to the percent sequence identity using the same symbols as in A. Residues of TM4 are not part of the receptor cavity. Circled star symbols refer to residues common to CCR5 and receptors CCR1 through CCR4. Star symbols denote residues unique to CCR5. Surface areas are given as a percentage compared to the same residues in an Ala-Xxx-Ala peptide (Hubbard and Thornton, 1993).



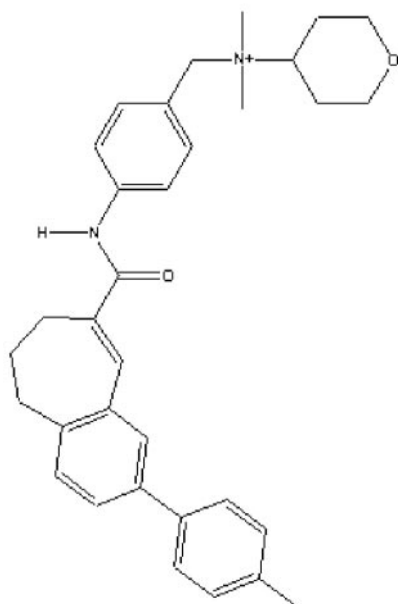


FIGURE 4 Chemical structure of TAK779 (*N,N*-dimethyl-*N*-4[[[2-(4-methylphenyl)-6,7-dihydro-5H-benzocyclohepten-8-yl]carbon-yl]benzyl]-tetrahydro-2H-pyran) (Baba et al., 1999).

(3.36) of CCR2 correspond to L104 and F112 of CCR5, respectively. Docking simulations of TAK779 to CCR1 resulted in some orientations similar to those of CCR5, but the docking scoring energy was high, suggesting residual atomic clashes in the binding region. MD simulations of TAK779 to CCR1 starting from a docked orientation similar to that of Fig. 5 resulted in the methylphenyl-benzocycloheptenyl group outside the pocket defined by the TM3, TM5, and TM6 helices. Amino acid comparison shows that

TABLE 3 Residues of CCR5 within a 5 Å cutoff of TAK779, as found in Fig. 5, and comparison with previous site-directed mutagenesis results

CCR5	% Viral Entry*
WT	0
L33	40
Y37	75
W86	80
L104	15
Y108	100
F112	10
T195	5
I198	20
L255	5
E283	30

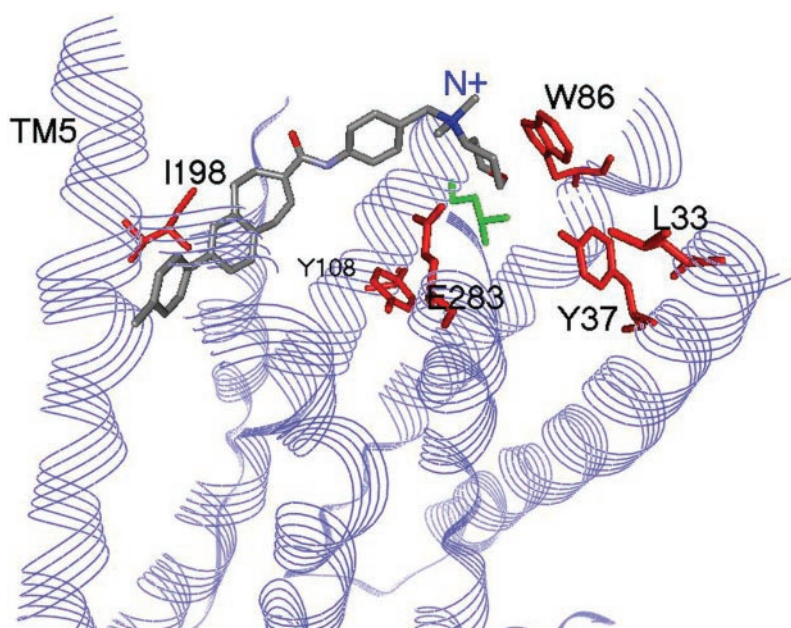
\*% Change in the efficacy of TAK779 to inhibit viral entry upon mutation of these residues to alanine (Dragic et al., 2000).

two tyrosines at positions 3.33 and 3.37 of CCR1 (Y114 and Y118, respectively) are replaced by Phe in CCR5 (F109 and F113, respectively). Other positions involve residues of similar size, such as 6.55 (L255 in CCR5 and I259 in CCR1) and 5.42 (I198 in CCR5 and L203 in CCR1). Apparently, replacing Phe with Tyr is sufficient for the observed displacement from the pocket of CCR1 during the MD simulation.

### Modeling and characterization of extracellular and intracellular regions

Modeling of the extracellular regions was performed in two steps. In the first step, information on secondary structure propensity was gathered. In the second step, each region was assembled sequentially into the model. Tertiary con-

FIGURE 5 Orientation of TAK779 in the TM region of CCR5 after 1 ns MD simulations. Side chains within 5 Å of TAK779 are shown with a stick representation. Residues in red have previously been implicated in TAK779 binding (Dragic et al., 2000). Residues in green are side chains whose substitution with Ala did not interfere with TAK779 antiviral activity.





straints were then introduced and the receptor was subjected to a series of simulated annealing simulations with appropriate distance and dihedral angle constraints. Secondary structure information was obtained from applications of prediction methods and by searching for structural templates in the PDB database (Berman et al., 2000).

Secondary structure propensities were obtained by comparing predictions from seven major algorithms, as provided by the Protein Structure Prediction (PELE) module of the Biology Workbench (<http://workbench.sdsc.edu>). Structure was assigned based on the joint prediction for each amino acid. Database searches included seven amino acids at either ends of the loop region to allow stemming of the loops from the helix ends. The searches produced from 3 to 14 hits for each of the loops investigated. In the majority of cases, sequence homology searches produced hits for only portions of the loops, with the average length of the fragment being  $11 \pm 5$  residues. Interestingly, the matches were from either loops or solvent-exposed regions of their respective x-ray structures. The fragments found by the database search were also used to confirm or reject the secondary prediction results. If agreement was found, dihedral angle constraints were added to restrict the matched regions to either the  $\alpha$ -helical or  $\beta$ -sheet conformation during the simulations.

Application of secondary structure prediction to the 32-residue N-terminal region produced an extended conformation for residues 9 to 14 and a helical conformation for residues 26 to 31. Searches for sequence homologs produced fragments from 14 unique PDB structures. Segments that matched residues 26–31 were found to be helical (2 of 2 structures). Region 9–14 was matched by structures containing amino acid classified as being either in an extended or bend region (4 structures) and it was modeled as an extended structure based on secondary structure prediction. The remaining residues of the N-terminus mostly matched unstructured segments and the dihedral angles of such residues were not constrained during the simulations. Two prolines, P34 and P35, introduce a helix break between the predicted N-terminus 26–31 helix and TM1.

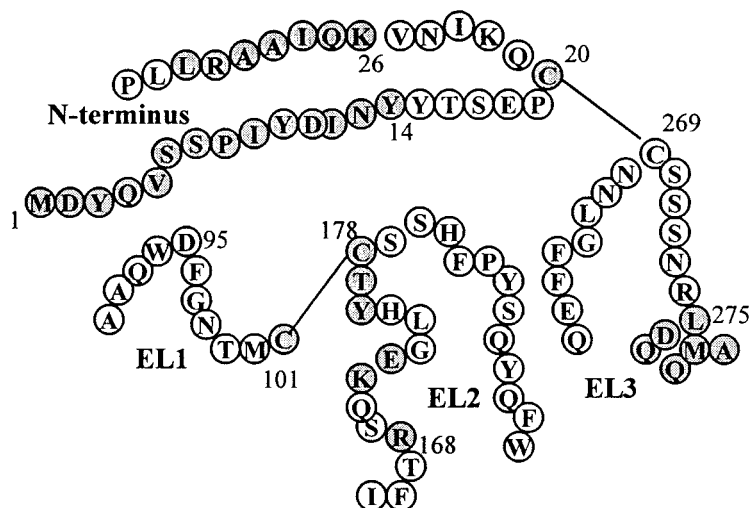
The first extracellular loop, connecting TM2 with TM3, comprises residues Y89 to M100 of the model. Application of the prediction algorithm resulted in an extension of the TM2 helix to A91. It also showed that the  $W_{94}DFGNT_{99}$  region was unstructured. Database searches resulted in one hit for region  $F_{85}-A_{91}$ , corresponding to a helical fragment and three matches for  $W_{94}-T_{99}$  corresponding to either coil or bend regions of these proteins. The loop structures varied greatly among the three hits and dihedral angle constraints were not applied to the  $W_{94}-T_{99}$  fragment of EL1.

The disulfide bridge between C101 and C178 (Blanpain et al., 1999b) separates EL2 in two segments, with the N-terminal residues linking TM3 and TM4, while residues C-terminal from C178 connect TM4 to TM5 (Fig. 1). Homology searches did not produce clear secondary structure

preference when searches were done either on the entire EL2 region or only the N-terminal or C-terminal fragments. Application of homology searches and prediction methods to EL3 resulted in a helical conformation for the  $L_{275}DQAMQ_{280}$  fragment, thus extending the TM7 helix into the extracellular domain by six residues. The database search provided four hits for the  $L_{275}-Q_{280}$  region, three of which were helical. Residues 261–274 of EL3 were matched by four fragments, all of which were in loop regions of these proteins.

Because of the scarcity and limited match of the database results, the fragments were used to generate initial templates and loop optimization was carried out using a constrained simulated annealing protocol. Loops were inserted one at the time by choosing conformations free of steric overlap among the different regions. Dihedral angle constraints were used to restrain the residues to the predicted secondary structures during the simulated annealing procedure. Tertiary constraints were added in the form of distance constraints between residues in different loops, based on available epitope mapping results, summarized in Table 4. Distance constraints  $r \leq 15 \text{ \AA}$  were used between the  $C_{\alpha}$  atoms of epitope residues (Burritt et al., 1998). Constraints were applied between N-terminal residues (region 1–13) and EL2 (residues 168, 176–177), and between EL1-D95 and K171-E172 in EL2. A more generous constraint of  $r \leq 25 \text{ \AA}$  was added between the  $C_{\alpha}$  atoms of D95 and those of residues 1–13 of CCR5 to account for the observed interaction between the first 13 amino acids of the N-terminus and D95 (Hill et al., 1998). Specific side chain-side chain interactions were not obtained in that study. The  $25 \text{ \AA}$  cutoff corresponds to the average diameter of a globular protein of the same size as the extracellular domain of CCR5 (Harpaz et al., 1994). This cutoff assumes that the observed sensitivity upon mutation of D95 may arise not only from direct side chain-side chain contact, but also from structural changes in the globular fold of these regions. As shown in rhodopsin, the EC domains of GPCRs are likely to form a compact folding motif. By comparison, the EC domain of rhodopsin is similar in length to that of CCR5 and the distance between EL1 and the first 13 N-terminus residues varies between  $\sim 15$  and  $\sim 25 \text{ \AA}$ . The  $25 \text{ \AA}$  distance constraint between the N-terminus and EL1 simply imposes compactness between these two regions, without creating specific residue pair interactions that are not available at this time. The set of geometrical constraints for the 95-residue extracellular domain consisted of a total of 29 distance constraints and 124 ( $\phi$ ,  $\psi$ ) dihedral angle constraints (Table 4).

The three intracellular loops of CCR5 range from 6 to 10 residues in length (Fig. 1). Residues in IL1 were modeled after rhodopsin, whose sequence homology with CCR5 is 83%. The second and third loops, however, have either low homology or differ in length from those of rhodopsin. These loops were therefore modeled by searching through a loop

**TABLE 4** Geometrical constraints used in modeling simulations of the extracellular loop regions\*

	Residues*	Distance <sup>#</sup> and dihedral angle constraints <sup>§</sup>	Reference
N-terminus	M <sub>1</sub> DYQVSSPIYDIN <sub>13</sub> I <sub>9</sub> YDINY <sub>14</sub> C <sub>20</sub> K <sub>26</sub> QIAARL <sub>32</sub> D <sub>2</sub> YQ <sub>4</sub> , S <sub>7</sub> P <sub>8</sub> , N <sub>13</sub>	C <sub>α</sub> (D95)-C <sub>α</sub> (M <sub>1</sub> -Y <sub>13</sub> ) ≤ 25 Å E S-S bond with C269 H C <sub>α</sub> (X)-C <sub>α</sub> (Y176) ≤ 15 Å C <sub>α</sub> (X)-C <sub>α</sub> (T177) ≤ 15 Å	Hill et al., 1998 Olson et al., 1999 Olson et al., 1999
EL1	D <sub>2</sub> D <sub>95</sub>	C <sub>α</sub> (D2)-C <sub>α</sub> (R168) ≤ 15 Å C <sub>α</sub> (D95)-C <sub>α</sub> (K171) ≤ 15 Å C <sub>α</sub> (D95)-C <sub>α</sub> (E172) ≤ 15 Å C <sub>α</sub> (D95)-C <sub>α</sub> (M <sub>1</sub> -Y <sub>13</sub> ) ≤ 25 Å	Olson et al., 1999 Lee et al., 1999 Lee et al., 1999 Hill et al., 1998
EL2	C <sub>101</sub> R <sub>168</sub> K <sub>171</sub> E <sub>172</sub> Y <sub>176</sub> T <sub>177</sub>	S-S bond with C178 C <sub>α</sub> (D2)-C <sub>α</sub> (R168) ≤ 15 Å C <sub>α</sub> (D95)-C <sub>α</sub> (K171) ≤ 15 Å C <sub>α</sub> (D95) C <sub>α</sub> (E172) ≤ 15 Å C <sub>α</sub> (X)-C <sub>α</sub> (Y176) ≤ 15 Å C <sub>α</sub> (X)-C <sub>α</sub> (T177) ≤ 15 Å	Olson et al., 1999 Lee et al., 1999 Lee et al., 1999 Olson et al., 1999 Olson et al., 1999
EL3	C <sub>178</sub> L <sub>275</sub> DQAMQ <sub>280</sub> C <sub>269</sub>	S-S bond with C101 H S-S bond with C20	

\*Residues are denoted by their one-letter code. Subscripts indicate the residue number of the first and last residue in the fragments.

<sup>#</sup>C<sub>α</sub>(X) denotes the C<sub>α</sub> atom of the residues D<sub>2</sub>YQ<sub>4</sub>, S<sub>7</sub>P<sub>8</sub>, N<sub>13</sub>.

<sup>§</sup>H = ( $\phi = -65^\circ$ ,  $\psi = -40^\circ$ ), E = ( $\phi = -115^\circ$ ,  $\psi = -135^\circ$ ). Additional conformational constraints were  $\phi < 0$  for all residues except Gly and  $\omega = 180^\circ$  for non-Pro residues.

database for segments with similar end-end distance (InsightII, 2002. Accerlys Inc., San Diego, CA). The receptor model was terminated after helix 8 (Palczewski et al., 2000), which has high homology with rhodopsin. Residues 302 to 352 were not modeled because of low homology to rhodopsin and lack of experimental data. Initial loop structures were then subjected to a simulated annealing protocol, as described in the Methods.

The accuracy of modeled loop structures is typically assessed against the actual x-ray structure of the test proteins (Fiser et al., 2000). Unfortunately, such comparison is not possible for CCR5, because of the poor homology with

rhodopsin in the extracellular domains. A previous study of single loops by Fiser et al. (2000) has shown that, given an adequate sampling of the conformational space, the quality of the structures can be inferred from 1) the correlation between the energy of the models and the RMSD, and 2) by the low structural variability among the low-energy structures. These two criteria were used here in the analysis of the simulated EC domains.

The bell-shaped energy profile of the simulated structures (Fig. 6 A) indicates statistically significant sampling of the conformational space. A plot of the RMSD from the lowest energy structure versus energy gave a Pearson coefficient

$r = 0.42$  (Figure 6 B). The backbone atoms of the TM regions and EC domains were used in the superimposition. Structural variability was calculated as the average RMSD of the nine structures with the lowest conformational energies. Fig. 7 shows two low-energy structures representative of low variability ( $E = -398.8$  kcal/mol) and high variability ( $E = -384.2$  kcal/mol). Variability was small for five of these structures, suggesting that one dominant native conformation may have been obtained from the simulations ( $E = -398.8$  kcal/mol in Fig. 7). However, the two criteria for assessing the quality of the EC domains are only partially fulfilled by the modeled structure because of the weak-to-moderate correlation between energy and RMSD, and the presence of low-energy conformations with high variability (Fig. 6, *inset*).

When each EC domain was superimposed individually, rms deviations from the mean coordinates ranged up to 9 Å for the N-terminus, 5 Å for EL1, 7 Å for EL2, and 6 Å for EL3. The structures were then clustered based on structural similarity to find major structural groups. RMSD cutoffs were increased from 2.0 Å until a significant number of structures were represented in each cluster. Clustering of EL1 structures with a 3.5 Å cutoff resulted in 60% of the structures in six clusters containing 14 to 53 loops each. About 50% of EL2 structures were found in nine clusters with 9 to 18 structures each when using a 6.5 Å cutoff. Fifty percent of the EL3 structures clustered in eight groups containing 9 to 27 structures each. Clustering of the N-terminus structures with a 7 Å cutoff resulted in 30% of the structures in five clusters containing 17 to 23 structures each. Results show that, with the exception of the shorter EL1, the clusters contain structures with large structural variations among them.

I further analyzed the simulated structures in terms of their average properties to capture global characteristics that may not be apparent from either the lowest energy structures or individual loop conformations. Commonalities in the global fold of the EC domains were found by examining average properties of solvent accessibility and frequency of contact interactions among the domains.

The solvent-accessible surface area of the EC amino acids, averaged over the 250 structures, is shown in Fig. 8 A. Approximately 60% of the N-terminus residues have solvent accessibility >40%. On the contrary, EL3 is clearly the most buried of the loops. The disulfide bridge between C101 and C178 limits the solvent accessibility of six residues near C178 in EL2, where, on average, the C-end of this loop is more solvent-exposed than its N-end portion. The low accessibility of D95 in EL1 arises from interaction of this side chain with residues of the N-terminus.

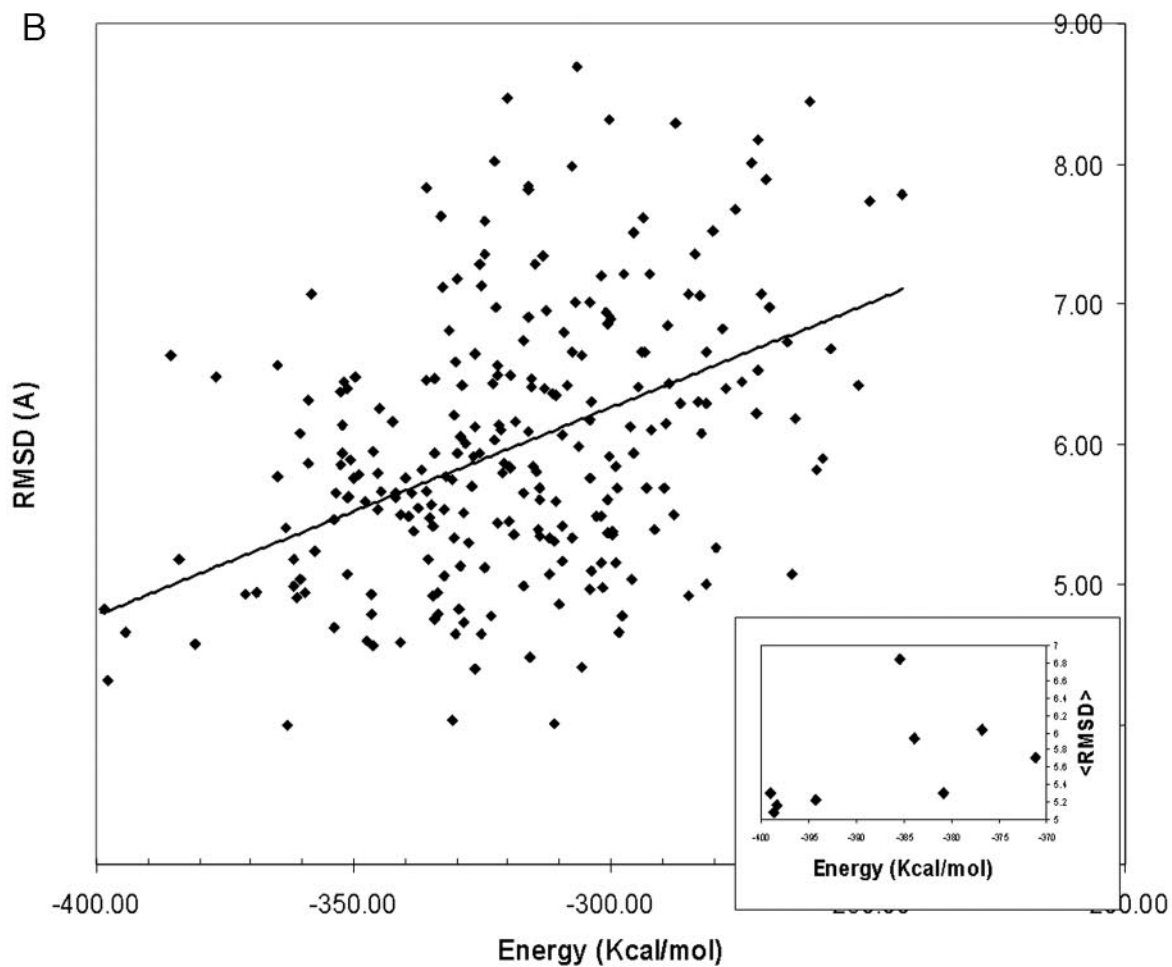
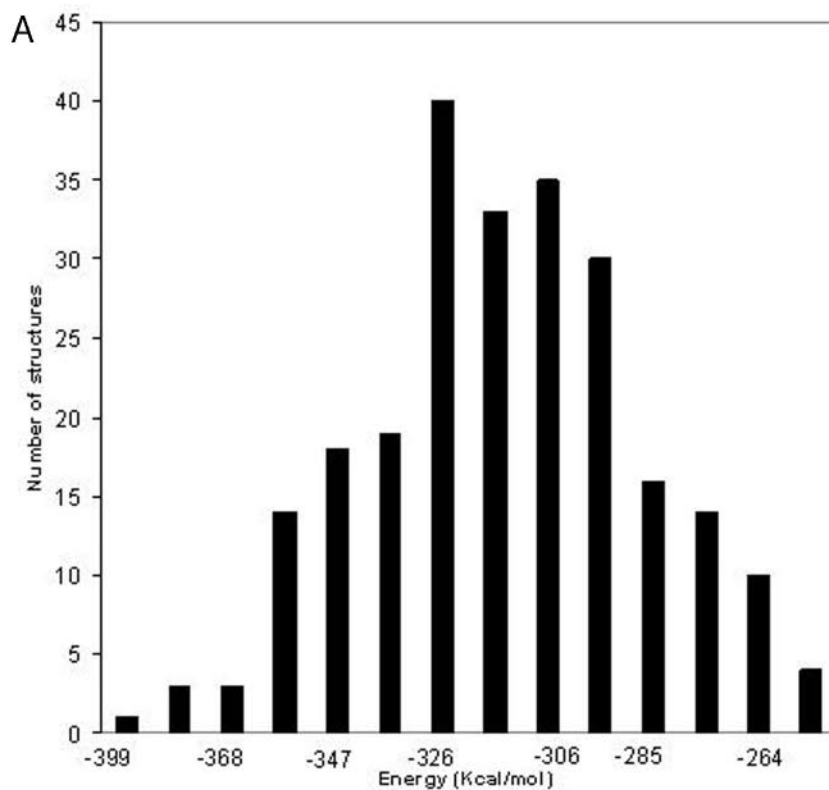
The 250 structures were also evaluated based on interaction of hotspot residues, defined here as residues of the loop regions whose mutation results in loss of binding for either chemokines or gp120. As shown in Table 5, hotspot residues consist of mostly acidic and aromatic amino acids in

the N-terminus, EL2, and EL3. The environment of these crucial residues was characterized by selection of the interacting residues from the set of the 250 structures (Fig. 8 B). The majority of the models showed interactions of EL1 with hotspot residues located in the N-terminus (D2-I12 and Y14-E18) and EL2 (Y176-T177). The second extracellular loop presented interactions with the D2-I12 fragment of the N-terminus. Close contacts between the Y14-E18 segment and EL3 originate from the disulfide bond between C20 and C269. Residues at either the C-end of the N-terminus (K26 and R31) or F263-N267 of EL3 did not show significant interaction with the other extracellular regions.

By comparison, simulations performed without the addition of tertiary distance constraints completed lacked interactions among the four domains, except in the immediate proximity of the two disulfide bonds. The four domains did not assemble into a compact fold, and several structures were found where the domains packed against the 7-TM helices. Such conformations are unrealistic in the membrane environment. They appear because the simulated annealing protocol has to be carried out in vacuo, thus ignoring the lipid environment of the TM helices.

## DISCUSSION

The x-ray structure of rhodopsin provides the template for models of the homologous GPCR. However, the specific sequence motifs of rhodopsin give rise to local distortions of the 7-TM helices that may not be present in other receptors. Deviations from the standard helical conformation result in a different orientation of the helices in the bundle and changes in the shape and amino acid composition of the GPCR binding pocket. Specific structural motifs may also provide a mechanism of receptor activation. Recently, Govaerts et al. (2001a) have investigated the role of the T<sub>82</sub>VP<sub>84</sub> motif in TM2 of CCR5 in receptor activation. A P84A mutation resulted in decreased affinity for chemokines, while mutations of T72 impaired receptor activation following binding of chemokines. In that work, the helix bent at P84 was correlated with the activation process. In this work, I have found that the helical distortion in TM2 of rhodopsin is characteristic of Gly-containing helices in interfacial regions (Javadpour et al., 1999), resulting in tight packing with TM3. The Gly motif of rhodopsin was then replaced in CCR5 with a motif derived from Pro-containing helices (Némety et al., 1992; Sankararamkrishnan and Vishveshwara, 1992) and residues at the C-terminal end of TM2 were oriented based on sequence conservation in a set of 50 sequences homologous to CCR5. The resulting helix kink orients the CC-chemokine conserved tryptophan (W86) in the binding pocket, in an orientation similar to that modeled by Govaerts et al. (2001a). Unlike the changes in TM2, helix 5 of CCR5 was modeled after rhodopsin with a bulge in the L203-V209 region. The TM5 and TM7 rhodopsin templates were maintained based on sequence com-



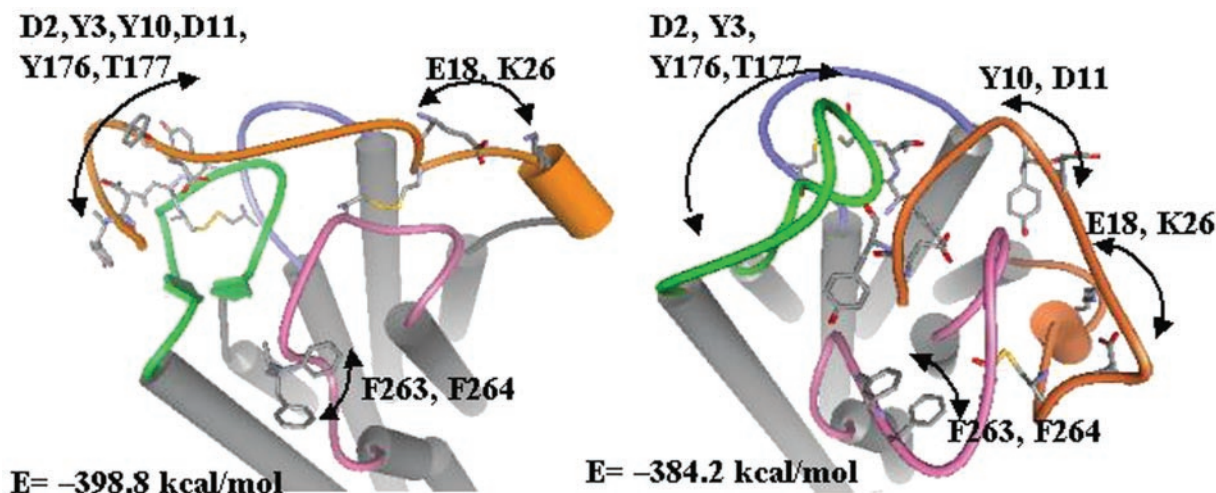


FIGURE 7 Schematic diagram of two representative low energy structures of the EC domains. Hotspot residues are shown in a stick representation. Orange: N-terminus; cyan: EL1; green: EL2; magenta: EL3.

dopsin and consideration of sequence homology and experimental data on closely related receptors.

Receptors in Table 1 share the highest sequence homology with CCR5. Highly conserved residues line the bottom of receptor cavity (Fig. 3 B), where they form a tight network of interacting side chains. These amino acids comprise the highly conserved “fingerprint” residues characteristic of the rhodopsin-GPCR family (Attwood and Findlay, 1994). It was surprising, however, to find few additional residues shared across receptors classes. Only the angiotensin receptors had residues in common with CC-chemokines (P34 (1.36), W86 (2.60), and C290 (7.46) of CCR5). CCR5 shares six residues with receptors CCR1 to CCR4 (Fig. 3 B). Two conserved aromatic clusters characterize the cavity in these receptors (Fig. 2). Four aromatic residues, TM1-Y37 (1.39), TM2-W86 (2.60), TM3-Y108 (3.32), and TM6-Y251 (6.51), are within an  $\sim 6$  Å radius from the acidic residue TM7-E283. The second aromatic cluster is adjacent to the first one and includes the highly conserved Y244 (6.44) and W248 (6.48) in TM6 and the CC-chemokine-specific residues TM2-F79 (2.53) and TM7-H289 (7.45). Therefore, the receptor cavity of receptors CCR1 through CCR5 utilizes class-specific aromatic residues for side chain interactions with the “fingerprint” residues of GPCR (Y244 and W248 in CCR5). Given the implication of these residues in receptor activation (Javitch et al., 1998), their interaction with the conserved aromatic cluster could serve as a common mechanism of activation in proteins CCR1 to CCR5.

Site-directed mutagenesis studies have characterized the binding mode of the CCR5 antagonist TAK-779 (Dragic et al., 2000). Single point mutations of residues L33, Y37, W86, Y108, and E283 to Ala decreased the efficacy of TAK-779 in antagonizing the binding of gp120 to CCR5. Our molecular docking simulations show that all of the above residues are part of TAK779 binding site (Dragic et al., 2000) (Fig. 5 and Table 3). In particular, the docking orientation is characterized by an electrostatic interaction between the ammonium group of TAK779 and E283. This glutamate is conserved in receptors CCR1 through CCR5, it is solvent-accessible (Fig. 3 B), and it is the only acidic residue in the extracellular end of the 7-TM bundle. Mutagenesis data have shown the importance of this glutamate not only for the binding of TAK779, but for antagonists of other CC-chemokine receptors as well. Recent mutagenesis data on CCR2 showed loss of binding of basic spiro-piperidine ligands upon mutation of Glu to either alanine or glutamine (Mirzadegan et al., 2000). Correspondingly, quaternization of the piperidine nitrogen has been shown to be essential for the high affinity of CCR1 antagonists (Liang et al., 2000; Naya et al., 2001). Given that small ligands of CC-chemokine receptors are characterized by a basic amino group (Liang et al., 2000; Ng et al., 1999; Sabroe et al., 2000), it is plausible that receptors CCR1 through CCR5 share a common binding mode characterized by an electrostatic interaction with the conserved TM7-glutamate. Additional interactions of TAK779 with residues in TM5 and TM6 find partial agreement with the mutagenesis data (Ta-

FIGURE 6 Analysis of the extracellular loop regions. (A) Histogram of the conformational energy distribution of the 250 structures generated by the simulated annealing procedure. Energies were binned at 10 kcal/mol. (B) Energy plot versus RMSD from the lowest energy structure ( $E = -398.9$  kcal/mol). The straight line is the least-square fit (Pearson coefficient = 0.42). Inset: the average rms deviation, of the nine low-energy structures. The RMSD is calculated from the average of  $(9 \times 8)/2$  pairwise RMSDs.

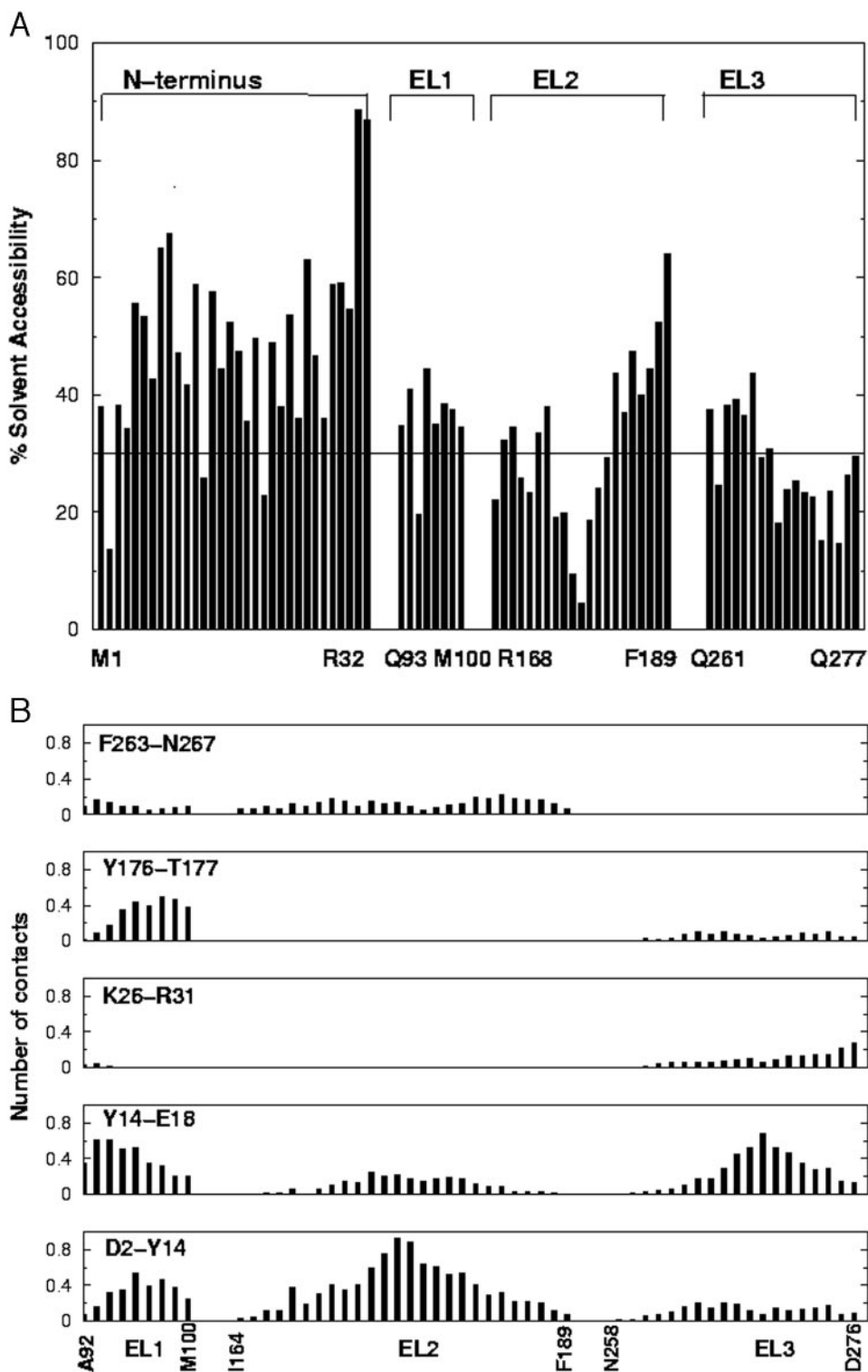


FIGURE 8 Analysis of the extracellular loop regions. (A) Average solvent accessibility surface areas of the extracellular regions of CCR5. Values represent the average over 250 structures generated by the simulated annealing procedure. (B) Interactions between hotspot residues, defined in Table 5, and the three extracellular loops. Normalized values indicate the probability that a given residue of the EC regions is within 5 Å of hotspot residues. Graphs were generated using the 250 structures obtained from the simulated annealing procedure.

ble 3). The residues involved are small (T195) or hydrophobic (L104, F112, and L255). It is plausible that their

substitution with Ala may perturb the binding energy less than the replacement of charged or polar aromatic residues.

**TABLE 5** Residues implicated in the interaction of CCR5 with either MIP-1 $\beta$  or dual-tropic HIV virus, as obtained from site-directed mutagenesis data

Extracellular Location	Residue Number	% Loss of MIP-1- $\beta$ Binding	% Loss of Dual-Tropic HIV Virus Binding	Reference	
N-terminus	D2	100	100	Blanpain et al., 1999a	
	Y3	100	60	Blanpain et al., 1999a	
	Y10	100	100	Blanpain et al., 1999a	
	D11	100	90	Blanpain et al., 1999a	
	I12	100	100	Blanpain et al., 1999a; Howard et al., 1999	
	Y14		100	Rabut et al., 1998	
	Y15		100	Rabut et al., 1998	
	S17			80	Rabut et al., 1998
	E18	100		80	Blanpain et al., 1999a
	K26	100			Zhou et al., 2000
	R31			70	Zhou et al., 2000
EL2	Y176		50	Genoud et al., 1999	
	T177		60	Genoud et al., 1999	
EI3	F263		50	Genoud et al., 1999	
	F264		50	Genoud et al., 1999	
	N267		50	Genoud et al., 1999	

Similarly, loss of opioid receptor affinities is greatest upon mutation of a critical Asp and nearby polar aromatic residues, while mutations of hydrophobic groups have smaller effects (Surratt et al., 1994; Befort et al., 1996; Metzger et al., 2001).

The docking mode of TAK779 (Fig. 5), while in agreement with the experimental data, differs from a previous orientation proposed from the analysis of these same data (Dragic et al., 2000). It has been previously suggested that the methyl-benzyl-heptadyl moiety binds among TM1, TM2, and TM7, while the positively charged ammonium group would orient toward the extracellular domain (Dragic et al., 2000). Automated docking simulations were not performed in that study. The proposed binding mode of TAK779 presented here is validated by the integration of systematic docking simulations together with mutagenesis, SAR, and sequence analysis.

The binding mode of TAK779 accounts for receptor affinity, but also shows that the above six residues are not unique to CCR5, as they are also found in CCR1 and CCR2, two receptors with low affinity for TAK-779 (Baba et al., 1999). Therefore, the selectivity of TAK-779 for CCR5 must derive from nonconserved side chains, such as those in Fig. 3 B. This hypothesis was tested by performing docking simulations of TAK779 to CCR1 and CCR2. Several docking orientations contained interaction of the ammonium nitrogen with the conserved glutamate in TM7, but the 4-methyl-benzocycloheptenyl group of TAK779 docked close to the extracellular end of the 7-TM bundle. We find that substitution of CCR5 amino acids with bulkier side chains in CCR2 and CCR1 is responsible for lack of binding deep in the TM pocket. Such residues are R206, F116, and Y124 in CCR2, and Y104 and Y118 in CCR1.

The docking results, when combined with current mutagenesis data and SAR on CCR5 and other CC-chemokine

receptors (Dragic et al., 2000; Liang et al., 2000; Mirzadegan et al., 2000; Naya et al., 2001) suggest a common binding mode for CC-chemokine antagonists. Conserved regions of CC-chemokine receptors (i.e., L33 (1.35), Y37 (1.39), W86 (2.60), Y108 (3.32), and E283 (7.39) in CCR5) bind the common chemical component of chemokines antagonists, i.e., the basic nitrogen of the piperidine ring, while residues in TM3, TM5, and TM6 interact with aromatic moieties of these molecules (the methyl-benzyl-heptadyl moiety in CCR5). The proposed binding mode, where conserved regions of receptors bind conserved chemical motifs of ligands, finds similarities with that of other GPCR, for example opioid and dopamine receptors (Simpson et al., 1999; Metzger et al., 2001; McFadyen et al., 2001). Extensive mutagenesis studies and docking studies of those two receptor classes has shown that a highly conserved Asp in TM3 (3.32) is responsible for the binding of the basic amino group of aminergic and opioid ligands, while divergent chemical moieties bind nonconserved amino acid residues. While the conserved Asp (3.32) is responsible for a large portion of the binding energy in the majority of aminergic and opioid receptors, the contribution of the nonconserved residues to binding was found only after extensive mutagenesis studies. The proposed binding mode for CC-chemokine antagonists can be used to guide further experimental studies to better define the interactions of TAK-779 with the nonconserved residues in TM5 and TM6.

Binding of TAK779 to CCR1 and CCR2 is likely to include residues of the EC regions, whose steric and electrostatic interactions give rise to poor affinity (CCR2) or lack of binding (CCR1). The likelihood that the EC region of CCR5 could interfere with the binding of the antagonist was investigated by pooling residues within a 5 Å radius of Y37, W86, and E278 from the set of 250 EC structures obtained by simulated annealing. Only 5% of the structures

had interactions between the binding pocket residues and either EL1 or EL2. Interactions with EL3 were found in 30% of the structures and centered on Q261 and the N273-Q277 region. These results are in agreement with the experimental findings (Dragic et al., 2000) that mutations of EL1 and EL2 residues have no effect on TAK779 binding, and Q261A, N273A-Q277A mutations mildly decreased the ability of TAK779 to interfere with gp120 binding. The results therefore exclude a contribution of the loops to the binding of TAK779 to CCR5.

I have used segment-matching (Sanchez and Sali, 1999) in combination with secondary prediction methods to generate structural templates and dihedral angle constraints of the four extracellular domains. The database searches provided segments that could be superimposed to generate an initial template. Typically, each loop was reconstructed by matching three to four overlapping segments. Given the geometrical variability of the CCR5 loops, it is unlikely that the matches found for CCR5 are a unique representation of the conformation of the EC domains. Therefore, I have chosen to use the database searches not for obtaining an actual structural template, but to uncover regions where regular secondary structure motifs may occur. Such regular structures occur in rhodopsin (Palczewski et al., 2000), where the N-terminus and EL2 assemble into  $\beta$ -sheet folds, and have been postulated to exist in other receptors as well (Paterlini et al., 1997; Moro et al., 1999; Zhang et al., 2002b). A previous modeling study of EL2 of the  $\kappa$ -opioid receptor based on database matches and secondary prediction (Paterlini et al., 1997) has been recently validated by NMR spectroscopy of this loop in a DPC micelle (Zhang et al., 2002b).

The conformational freedom of the individual loops was restricted by application of dihedral angle constraints based on both the predicted and observed secondary structure propensities when available. Tertiary interactions, in the form of distance constraints between different structural domains, were added from considerations of available epitope mapping results. A simulated annealing protocol was then used to generate loop models that were analyzed based on conformational energy criteria, conformational variability, and average physical properties. The set of  $\sim 150$  dihedral and geometrical constraints used in the simulations was not sufficient for complete structural determination of the 95 residues of EC domains. Loop conformations, when clustered according to their RMSD, showed structural variability from 3.5 Å for EL1 to 7 Å for the N-terminus. However, the distance constraints between the EC loops were necessary to maintain a compact fold of the four regions. Simulations carried out without the addition of tertiary distance constraints resulted in improbable structures void of interactions among the EC domains where the longer loops and the N-terminus packed against the hydrophobic TM helices.

Criteria used to assess the quality of modeled loops have limited applicability to CCR5 because of the length of the loops and interactions among the domains. I have estimated the accuracy of the simulated EC domains from the correlation between the conformational energy and RMSD and from the conformational variability of the low-energy conformations. This approach has been previously tested on single loops up to 12 residues in length (Fiser et al., 2000), and it is applied here to a considerably more complex system. I find a low-to-moderate correlation and low variability of some of the structures, those suggesting that a dominant low energy conformation has been found (Fig. 7). The quality of the correlation and variability is likely to improve by either increasing the number of constraints or, alternatively, by applying stricter distance constraints. However, available biological data suggested only a range of interacting residues (as between D95 and the N-terminus), thus making it necessary to adopt distance cutoffs that are not biased toward specific residue-residue interactions.

The binding of chemokine and gp120 has been extensively probed by site-directed mutagenesis of residues in the EC regions (Table 5). Single-point mutations can either eliminate essential side chain interactions with ligands, or perturb the tertiary structure of the EC domains. If the models can distinguish between the two cases we can utilize them to guide experimental design.

Average properties of the modeled EC regions were used to obtain common folding characteristics and pattern of interacting residues. The average solvent accessibility area (Fig. 8 A) reflects, in part, the geometrical constraints imposed by the disulfide bonds and between the N-terminus and EL2. The lower accessibility of EL3 is likely due to the disulfide bond between C20 and C269, which causes the N-terminus to lie directly above EL3 (Fig. 7). The observation that it has not been possible to raise monoclonal antibodies against EL3 (Lee et al., 1999) finds support in low solvent accessibility of EL3 in the current model. On the contrary, the C-terminal region of EL2 has clearly high solvent accessibility (Fig. 8 A), in agreement with the finding that several mAbs can recognize epitopes in this region (Lee et al., 1999).

Structures characteristic of major structural clusters show commonalities, despite the great variability among them. As illustrated in the case of low-energy structures (Fig. 7), the N-terminus cuts across the EC domain to reach EL2 thus separating EL1 from EL3. The interactions between the N-terminus and the three loops give rise to clusters of hotspot aromatic and acidic residues such as with D95 in EL1, Y176-T177 in EL2, and C269 in EL3 (Figs. 8 B). Some of the hotspot residues were also part of the epitope, and as such subjected to distance constraints (Table 4). However, the cutoffs used in the simulations were three to fivefold greater than the one used to identify close contacts ( $d \leq 5$  Å) and individual residue contacts were not specified. Many of the hotspot residues are also characterized by



low average solvent accessibility because they are involved in side chain-side chain interactions, such as D2, D95, R168, Y176, and T177. I suggest that the loss of binding observed for mutations at these loci originate from a perturbation of the tertiary organization of the EC regions. In contrast, regions such as the C-terminal end of the N-terminus and residues F263-N267 in EL3 both lack specific interactions and are solvent-accessible (Fig. 8, A and B). I propose that the observed loss of binding upon mutagenesis in these regions originates from a direct loss of interactions with chemokines or gp120. Gain-of-function studies where the affinity of CCR5 is restored by complementary replacement or swapping of residue pairs (Zhou et al., 1994; Govaerts et al., 2001b) could be used to validate side chain interactions obtained in the study. For example, experiments could involve residue pairs such as N-term-D2 and EL2-R168, N-term-N13 and EL1-D95, or N-term-E18 and EL3-R274. Putative sites of chemokines interaction outlined here may also be used to map the molecular determinants of RANTES recognition (Nardese et al., 2001) onto the surface of the EC models. For example, the clusters of negatively charged residues can be used to orient RANTES on the EC surface by matching the complementary basic clusters of this chemokines in docking simulations.

## CONCLUSIONS

Presently, GPCRs can be modeled from rhodopsin, the only member of this family for which the x-ray structure has been resolved. The usefulness of these homology models greatly depends on their ability to explain and predict the binding of their endogenous ligands and to efficiently aid the discovery of new synthetic compounds. Sensitivity of the 7-TM template to local distortions, imparted by sequence-specific motifs, can affect the size and nature of the binding cavity. The difficulty in obtaining structural data of the loop regions based on sequence homology makes it arduous to structurally characterize these regions, despite their importance in binding and recognition, as exemplified by CCR5.

I have presented a comprehensive model of CCR5 that elucidates the binding of both small ligands and its sensitivity to mutations for binding of chemokines and the coat protein gp120 of HIV-1. The computational approach has sought to enhance homology-modeling techniques with *ab initio* simulations and knowledge-based information to generate structural models that are then corroborated by comparison with additional data. The model of the transmembrane region was validated by probing the binding of TAK779 using automated docking simulations. The docking mode finds support in available experimental data and specific hypotheses have been formulated to explain both affinity and selectivity of TAK-779 for CCR5. Application of evolutionary tracing to the 7-TM region readily identifies CC-conserved residues, such as E283 and aromatic residues

in TM1, TM2, and TM3, that create a CC-chemokine class-specific receptor pocket. Correlation of the conserved glutamate with the presence of a quaternary ammonium group in current CC-chemokine antagonists further suggests commonalities in the binding of CCR1-CCR5 that can be exploited when designing small ligands for these receptors.

The size and complexity of the EC regions examined here required that supplementary constraints be added during modeling, in addition to knowledge-base information on the secondary structure of the individual loops. These were obtained in the form of tertiary constraints based on the epitope mapping studies. Analysis of the average properties of the simulated structures suggests different roles for the functionally important residues, as either maintaining the tertiary structure of the EC domain, or as being accessible to binding by chemokines or gp120. The constrained simulated annealing protocol, in combination with conformational clustering, provides a systematic approach for generating low-resolution structures of EC domains for further experimental validation and design.

I thank Loren Gragert for assistance with the computations, and am grateful to Philip Portoghesi and Andrew Shenker for helpful discussions.

This work was supported by National Institutes of Health Grant 5K01DA0073 and by the Minnesota Supercomputer Institute.

## REFERENCES

- Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25:3389–3402.
- Atwood, T. K., and J. B. C. Findlay. 1994. Fingerprinting G-protein coupled receptors. *Protein Eng.* 7:195–203.
- Baba, M., O. Nishimura, N. Kanzaki, M. Okamoto, H. Sawada, Y. Iizawa, M. Shiraiishi, Y. Aramaki, K. Okonogi, Y. Ogawa, K. Meguro, and M. Fujino. 1999. A small-molecule, nonpeptide CCR5 antagonist with highly potent and selective anti-HIV-1 activity. *Proc. Natl. Acad. Sci. USA.* 96:5698–5703.
- Bairoch, A., and R. Apweiler. 2000. The SWISS-PROT protein sequence database and its supplement TrEMBL. *Nucleic Acids Res.* 28:45–48.
- Baldwin, J. M., G. F. Schertler, and V. M. Unger. 1997. An alpha-carbon template for the transmembrane helices in the rhodopsin family of G-protein-coupled receptors. *J. Mol. Biol.* 272:144–164.
- Ballesteros, J., S. Kitanovic, F. Guarnieri, P. Davies, B. J. Fromme, K. Konvicka, L. Chi, R. P. Millar, J. S. Davidson, H. Weinstein, and S. C. Sealfon. 1998. Functional microdomains in G-protein-coupled receptors. The conserved arginine-cage motif in the gonadotropin-releasing hormone receptor. *J. Biol. Chem.* 273:10445–10453.
- Ballesteros, J. A., and H. Weinstein. 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.
- Bayly, C. I., P. Cieplak, W. D. Cornell, and P. A. Kollman. 1993. A well-behaved electrostatic potential based method using charge restraints for determining atom-centered charges: The RESP model. *J. Phys. Chem.* 97:10269.
- Befort, K., L. Tabarra, D. Kling, B. Maigret, and B. L. Kieffer. 1996. Role of aromatic transmembrane residues of the delta-opioid receptor in ligand recognition. *J. Biol. Chem.* 271:10161–10168.

- Berger, E. A., R. W. Doms, E. M. Fenyo, B. T. Korber, D. R. Littman, J. P. Moore, Q. J. Sattentau, H. Schuitemaker, J. Sodroski, and R. A. Weiss. 1998. A new classification for HIV-1. *Nature*. 391:240.
- Berman, H. M., J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. 2000. The Protein Data Bank. *Nucleic Acids Res.* 28:235–242.
- Blanpain, C., B. J. Doranz, J. Vakili, J. Rucker, C. Govaerts, S. S. Baik, O. Lorthioir, I. Migeotte, F. Libert, F. Baleux, G. Vassart, R. W. Doms, and M. Parmentier. 1999a. Multiple charged and aromatic residues in CCR5 amino-terminal domain are involved in high affinity binding of both chemokines and HIV-1 Env protein. *J. Biol. Chem.* 274:34719–34727.
- Blanpain, C., B. Lee, J. Vakili, B. J. Doranz, C. Govaerts, I. Migeotte, M. Sharron, V. Dupriez, G. Vassart, R. W. Doms, and M. Parmentier. 1999b. Extracellular cysteines of CCR5 are required for chemokine binding, but dispensable for HIV-1 coreceptor activity. *J. Biol. Chem.* 274:18902–18908.
- Burritt, J. B., S. C. Busse, D. Gizachew, D. W. Siemsen, M. T. Quinn, C. W. Bond, E. A. Dratz, and A. J. Jesaitis. 1998. Antibody imprint of a membrane protein surface. Phagocyte flavocytochrome b. *J. Biol. Chem.* 273:24847–24852.
- Case, D. A., D. A. Pearlman, T. E. Caldwell, T. E. Cheatham III., W. S. Ross, C. L. Simmerling, T. A. Darden, K. M. Merz, R. V. Stanton, A. L. Cheng, J. J. Vincent, M. Crowley, D. M. Ferguson, R. J. Radmer, G. L. Deibel, U. C. Singh, P. K. Weiner, and P. A. Kollman. 1995. AMBER 5. University of California, San Francisco.
- Cornell, W. D., P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, Jr., D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. 1995. A second generation force field for the simulation of proteins and nucleic acids. *J. Am. Chem. Soc.* 117:5179–5197.
- Cotecchia, S., O. Rossier, F. Fanelli, A. Leonardi, and P. G. De Benedetti. 2000. The alpha 1a and alpha 1b-adrenergic receptor subtypes: molecular mechanisms of receptor activation and of drug action. *Pharm. Acta Helv.* 74:173–179.
- Donnelly, D., J. P. Overington, and T. L. Blundell. 1994. The prediction and orientation of alpha-helices from sequence alignments: the combined use of environment-dependent substitution tables, Fourier transform methods and helix capping rules. *Protein Eng.* 7:645–653.
- Dragic, T., 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, and P. J. Maddon. 1998. Amino-terminal substitutions in the CCR5 coreceptor impair gp120 binding and human immunodeficiency virus type 1 entry. *J. Virol.* 72:279–285.
- Dragic, T., A. Trkola, D. A. Thompson, E. G. Cormier, F. A. Kajumo, E. Maxwell, S. W. Lin, W. Ying, S. O. Smith, T. P. Sakmar and J. P. Moore, J. P. 2000. A binding pocket for a small molecule inhibitor of HIV-1 entry within the transmembrane helices of CCR5. *Proc. Natl. Acad. Sci. USA.* 97:5639–5644.
- Dunbrack, R. L., Jr., and M. Karplus. 1993. Backbone-dependent rotamer library for proteins. Application to side-chain prediction. *J. Mol. Biol.* 230:543–574.
- Elling, C. E., U. Raffetseder, S. M. Nielsen, and T. W. Schwartz. 2000. Disulfide bridge engineering in the tachykinin NK1 receptor. *Biochemistry.* 39:667–675.
- Elling, C. E., K. Thirstrup, S. M. Nielsen, S. A. Hjorth, and T. W. Schwartz. 1997. Metal-ion sites as structural and functional probes of helix-helix interactions in 7TM receptors. *Ann. N.Y. Acad. Sci.* 814:142–151.
- Farzan, M., H. Choe, L. Vaca, K. Martin, Y. Sun, E. Desjardins, N. Ruffing, L. Wu, R. Wyatt, N. Gerard, C. Gerard, and J. Sodroski. 1998. A tyrosine-rich region in the N terminus of CCR5 is important for human immunodeficiency virus type 1 entry and mediates an association between gp120 and CCR5. *J. Virol.* 72:1160–1164.
- Fiser, A., R. K. Gian Do, and A. Sali. 2000. Modeling of loops in protein structures. *Protein Sci.* 9:1753–1773.
- Flanagan, C. A., W. Zhou, L. Chi, T. Yuen, V. Rodic, D. Robertson, M. Johnson, P. Holland, R. P. Millar, H. Weinstein, R. Mitchell, and S. C. Sealfon. 1999. The functional microdomain in transmembrane helices 2 and 7 regulates expression, activation, and coupling pathways of the gonadotropin-releasing hormone receptor. *J. Biol. Chem.* 274:28880–28886.
- Frisch, M. J., G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery, Jr., R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, P. Salvador, J. J. Dannenberg, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, A. G. Baboul, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, J. L. Andres, C. Gonzalez, M. Head-Gordon, E. S. Replogle, and J. A. Pople. 2001. Gaussian 98, Gaussian, Inc., Pittsburgh, PA.
- Genoud, S., F. Kajumo, F., Y. Guo, D. Thompson, and T. Dragic. 1999. CCR5-mediated human immunodeficiency virus entry depends on an amino-terminal gp120-binding site and on the conformational integrity of all four extracellular domains. *J. Virol.* 73:1645–1648.
- Govaerts, C., C. Blanpain, X. Deupi, S. Ballet, J. A. Ballesteros, S. J. Wodak, G. Vassart, L. Pardo, and M. Parmentier. 2001a. The TXP motif in the second transmembrane helix of CCR5. *J. Biol. Chem.* 276:13217–13225.
- Govaerts, C., A. Lefort, S. Costagliola, S. J. Wodak, J. A. Ballesteros, J. Van Sande, S. J. Wodak, L. Pardo, and G. Vassart. 2001b. A conserved Asn in transmembrane helix 7 is an on/off switch in the activation of the thyrotropin receptor. *J. Biol. Chem.* 276:22991–22999.
- Harpaz, Y., M. Gerstein, and C. Chothia. 1994. Volume changes on protein folding. *Structure.* 2:641–649.
- Hill, C. M., D. Kwon, M. Jones, C. B. Davis, S. Marmon, B. L. Daugherty, J. A. DeMartino, M. S. Springer, D. Unutmaz, and D. R. Littman. 1998. The amino terminus of human CCR5 is required for its function as a receptor for diverse human and simian immunodeficiency virus envelope glycoproteins. *Virology.* 248:357–371.
- Hoffman, T. L., C. C. LaBranche, W. Zhang, G. Canziani, J. Robinson, I. Chaiken, J. A. Hoxie, and R. W. Doms. 1999. Stable exposure of the coreceptor-binding site in a CD4-independent HIV-1 envelope protein. *Proc. Natl. Acad. Sci. USA.* 96:6359–6364.
- Howard, O. M., K. Shirakawa, J. A. Turpin, A. Maynard, G. J. Tobin, M. Carrington, J. Oppenheim, and M. Dean. 1999. Naturally occurring CCR5 extracellular and transmembrane domain variants affect HIV-1 co-receptor and ligand binding function. *J. Biol. Chem.* 274:16228–16234.
- Hubbard, S. J., and J. M. Thornton. 1993. NACCESS. Computer program, Department of Biochemistry and Molecular Biology, University College, London.
- Hubbell, W. L., D. S. Cafiso, and C. Altenbach. 2000. Identifying conformational changes with site-directed spin labeling. *Nat. Struct. Biol.* 7:735–739.
- Javadpour, M. M., M. Eilers, M. Groesbeek, and S. O. Smith. 1999. Helix packing in polytopic membrane proteins: role of glycine in transmembrane helix association. *Biophys. J.* 77:1609–1618.
- Javitch, J. A., J. A. Ballesteros, H. Weinstein, and J. Chen. 1998. A cluster of aromatic residues in the sixth transmembrane-spanning segment of the dopamine D2 receptor is accessible in the binding-site crevice. *Biochemistry.* 37:998–1006.
- Javitch, J. A., D. Fu, G. Liapakis, and J. Chen. 1997. Constitutive activation of the beta2 adrenergic receptor alters the orientation of its sixth membrane-spanning segment. *J. Biol. Chem.* 272:18546–18549.
- Johnson, J. M., and G. M. Church. 2000. Predicting ligand-binding function in families of bacterial receptors. *Proc. Natl. Acad. Sci. USA.* 97:3965–3970.
- Kolakowski, Jr., L. F. 1994. GCRDb: a G-protein-coupled receptor database. *Receptors, Channels.* 2:1–7.
- Kolchinsky, P., T. Mirzabekov, M. Farzan, E. Kiprilov, M. Cayabyab, L. J. Mooney, H. Choe, and J. Sodroski. 1999. Adaptation of a CCR5-using, primary human immunodeficiency virus type 1 isolate for CD4-independent replication. *J. Virol.* 73:8120–8126.

- Laskowski, R. A., M. W. MacArthur, D. S. Moss, and J. M. Thornton. 1993. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* 26:283–291.
- Lee, B., 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, and R. W. Doms. 1999. Epitope mapping of CCR5 reveals multiple conformational states and distinct but overlapping structures involved in chemokine and coreceptor function. *J. Biol. Chem.* 274:9617–9626.
- Li, W., S. Liang, R. Wang, L. Lai, and Y. Han. 1999. Exploring the conformational diversity of loops on conserved frameworks. *Protein Eng.* 12:1075–1086.
- Liang, M., C. Mallari, M. Rosser, H. P. Ng, K. May, S. Monahan, J. G. Bauman, I. Islam, A. Ghannam, B. Buckman, K. Shaw, G. P. Wei, W. Xu, Z. Zhao, E. Ho, J. Shen, H. Oanh, B. Subramanyam, R. Vergona, D. Taub, L. Dunning, S. Harvey, R. M. Snider, J. Hesselgesser, M. M. Morrissey, and H. D. Perez. 2000. Identification and characterization of a potent, selective, and orally active antagonist of the CC chemokine receptor-1. *J. Biol. Chem.* 275:19000–19008.
- Lichtarge, O., H. R. Bourne, and F. E. Cohen. 1996. An evolutionary trace method defines binding surfaces common to protein families. *J. Mol. Biol.* 257:342–358.
- Lu, Z. L., and E. C. Hulme. 2000. A network of conserved intramolecular contacts defines the off-state of the transmembrane switch mechanism in a seven-transmembrane receptor. *J. Biol. Chem.* 275:5682–5686.
- Luo, Z., D. J. Butcher, and Z. Huang. 1997. Molecular modeling of interleukin-8 receptor beta and analysis of the receptor-ligand interaction. *Protein Eng.* 10:1039–1045.
- McFadyen, I. J., T. G. Metzger, M. G. Paterlini, and D. M. Ferguson. 2001. Exploring the unique pharmacology of a novel opioid receptor, ZFOR1, using molecular modeling and the “message-address” concept. *Protein Eng.* 14:953–960.
- Meng, E. C., B. K. Shoichet, and I. D. Kuntz. 1992. Molecular docking using shape descriptors. *J. Comp. Chem.* 13:266–278.
- Metzger, T. G., M. G. Paterlini, D. M. Ferguson, and P. S. Portoghese. 2001. Investigation of the selectivity of oxymorphone- and naltrexone-derived ligands via site-directed mutagenesis of opioid receptors: exploring the “address” recognition locus. *J. Med. Chem.* 44:857–862.
- Metzger, T. G., M. G. Paterlini, P. S. Portoghese, and D. M. Ferguson. 1996. Application of the message-address concept to the docking of naltrexone and selective naltrexone-derived opioid antagonists into opioid receptor models. *Neurochem. Res.* 21:1287–1294.
- Mirzadegan, T., F. Diehl, B. Ebi, S. Bhakta, I. Polsky, D. McCarley, M. Mulkins, G. S. Weatherhead, J. M. Lapierre, J. Dankwardt, D. Morgans, R. Wilhelm, and K. Jarnagin. 2000. Identification of the binding site for a novel class of CCR2b chemokine receptor antagonists: binding to a common chemokine receptor motif within the helical bundle. *J. Biol. Chem.* 275:25562–25571.
- Mizobe, T., M. Maze, V. Lam, S. Suryanarayana, and B. K. Kobilka. 1996. Arrangement of transmembrane domains in adrenergic receptors. Similarity to bacteriorhodopsin. *J. Biol. Chem.* 271:2387–2389.
- Moro, S., C. Hoffmann, and K. A. Jacobson. 1999. Role of the extracellular loops of G protein-coupled receptors in ligand recognition: a molecular modeling study of the human P2Y1 receptor. *Biochemistry.* 38:3498–3507.
- Nardese, V., R. Longhi, S. Polo, F. Sironi, C. Arcelloni, R. Paroni, C. DeSantis, P. Sarmientos, M. Rizzi, M. Bolognesi, V. Pavone, and P. Lusso. 2001. Structural determinants of CCR5 recognition and HIV-1 blockade in RANTES. *Nat. Struct. Biol.* 8:611–615.
- Naya, A., Y. Sagara, K. Ohwaki, T. Saeki, D. Ichikawa, Y. Iwasawa, K. Noguchi, and N. Ohtake. 2001. Design, synthesis, and discovery of a novel CCR1 antagonist. *J. Med. Chem.* 44:1429–1435.
- Némety, G., K. D. Gibson, K. A. Palmer, C. N. Yoon, M. G. Paterlini, A. Zagari, S. Rumsey, and H. A. Scheraga. 1992. Energy parameters in polypeptides. 10. Improved geometrical parameters and nonbonded interactions for use in the ECEPP/3 algorithm, with application to proline-containing peptides. *J. Phys. Chem.* 96:6472–6484.
- Ng, H. P., K. May, J. G. Bauman, A. Ghannam, I. Islam, M. Liang, R. Horuk, J. Hesselgesser, R. M. Snider, H. D. Perez, and M. M. Morrissey. 1999. Discovery of novel non-peptide CCR1 receptor antagonists. *J. Med. Chem.* 42:4680–4694.
- Odile-Colson, A., J. H. Perlman, A. Smolyar, M. C. Gershengorn, and R. Osman. 1998. Static and dynamic roles of extracellular loops in G-protein-coupled receptors: a mechanism for sequential binding of thyrotropin-releasing hormone to its receptor. *Biophys. J.* 74:1087–1100.
- Olson, W. C., E. E. R. Gwénaél, K. A. Nagashima, D. N. H. Tran, D. J. Anselma, S. P. Monard, J. P. Segal, D. A. D. Thompson, F. Kajumo, Y. Guo, J. P. Moore, P. J. Maddon, and T. Dragic. 1999. Differential inhibition of human immunodeficiency virus type 1 fusion, gp120 binding, and CC-chemokine activity by monoclonal antibodies to CCR5. *J. Virol.* 73:4145–4155.
- Overington, J., D. Donnelly, M. S. Johnson, A. Sali, and T. L. Blundell. 1992. Environment-specific amino acid substitution tables: tertiary templates and prediction of protein folds. *Protein Sci.* 1:216–226.
- Palczewski, K., T. Kumasaka, T. Hori, C. A. Behnke, H. Motoshima, B. A. Fox, T. Le, D. C. Teller, T. Okada, R. E. Stenkamp, M. Yamamoto, and M. Miyano. 2000. Crystal structure of rhodopsin: a G protein-coupled receptor. *Science.* 289:739–745.
- Paterlini, G., P. S. Portoghese, and D. M. Ferguson. 1997. Molecular simulation of dynorphin A-(1–10) binding to extracellular loop 2 of the kappa-opioid receptor. A model for receptor activation. *J. Med. Chem.* 40:3254–3262.
- Pogozheva, I. D., A. L. Lomize, and H. I. Mosberg. 1998. Opioid receptor three-dimensional structures from distance geometry calculations with hydrogen bonding constraints. *Biophys. J.* 75:612–634.
- Prieto, J., and L. Serrano. 1997. C-capping and helix stability: the Pro C-capping motif. *J. Mol. Biol.* 274:276–288.
- Rabut, G. E., J. A. Konner, F. Kajumo, J. P. Moore, and T. Dragic. 1998. Alanine substitutions of polar and nonpolar residues in the amino-terminal domain of CCR5 differently impair entry of macrophage and dualtropic isolates of human immunodeficiency virus type 1. *J. Virol.* 72:3464–3468.
- Rizzuto, C. D., R. Wyatt, N. Hernandez-Ramos, Y. Sun, P. D. Kwong, W. A. Hendrickson, and J. Sodroski. 1998. A conserved HIV gp120 glycoprotein structure involved in chemokine receptor binding. *Science.* 280:1949–1953.
- Robertson, D. L., J. P. Anderson, J. A. Bradac, J. K. Carr, B. Foley, R. K. Funkhouser, F. Gao, B. H. Hahn, M. L. Kalish, C. Kuiken, G. H. Learn, T. Leitner, F. McCutchan, S. Osmanov, M. Peeters, D. Pieniazek, M. Salminen, P. M. Sharp, S. Wolinsky, and B. Korber. 2000. HIV-1 nomenclature proposal. *Science.* 288:55–56.
- Sabroe, I., M. J. Peck, B. J. Van Keulen, A. Jorritsma, G. Simmons, P. R. Clapham, and J. E. Williams Pease. 2000. A small molecule antagonist of chemokine receptors CCR1 and CCR3. Potent inhibition of eosinophil function and CCR3-mediated HIV-1 entry. *J. Biol. Chem.* 275:25985–25992.
- Sali, A., and T. L. Blundell. 1993. Comparative protein modeling by satisfaction of spatial restraints. *J. Mol. Biol.* 234:779–815.
- Samson, M., G. LaRosa, F. Libert, P. Paindavoine, M. Detheux, G. Vassart, and M. Parmentier. 1997. The second extracellular loop of CCR5 is the major determinant of ligand specificity. *J. Biol. Chem.* 272:24934–24941.
- Sanchez, R., and A. Sali. 1999. Comparative protein structure modeling in genomics. *J. Comp. Phys.* 151:388–401.
- Sankaramakrishnan, R., and S. Vishveshwara. 1992. Geometry of proline-containing alpha-helices in proteins. *Int. J. Pept. Protein Res.* 39:356–363.
- Simpson, M. M., J. A. Ballesteros, V. Chiappa, J. Chen, M. Suehiro, D. S. Hartman, T. Godel, L. A. Snyder, T. P. Sakmar, and J. A. Javitch. 1999. Dopamine D4/D2 receptor selectivity is determined by a divergent aromatic microdomain contained within the second, third, and seventh membrane-spanning segments. *Mol. Pharmacol.* 56:1116–1126.
- Sowa, M. E., W. He, T. G. Wensel, and O. Lichtarge. 2000. A regulator of G protein signaling interaction surface linked to effector specificity. *Proc. Natl. Acad. Sci. USA.* 97:1483–1488.
- Strader, C. D., T. M. Fong, M. R. Tota, D. Underwood, and R. A. Dixon. 1994. Structure and function of G protein-coupled receptors. *Annu. Rev. Biochem.* 63:101–132.

- Struthers, M., H. Yu, M. Kono, and D. D. Oprian. 1999. Tertiary interactions between the fifth and sixth transmembrane segments of rhodopsin. *Biochemistry*. 38:6597–6603.
- Subramaniam, S. 1998. The Biology Workbench—a seamless database and analysis environment for the biologist [editorial]. *Proteins*. 32:1–2.
- Subramaniam, G., M. G. Paterlini, D. L. Larson, P. S. Portoghese, and D. M. Ferguson. 1998. Conformational analysis and automated receptor docking of selective arylacetamide-based kappa-opioid agonists. *J. Med. Chem.* 41:4777–4789.
- Subramaniam, G., M. G. Paterlini, P. S. Portoghese, and D. M. Ferguson. 2000. Molecular docking reveals a novel binding site model for fentanyl at the mu-opioid receptor. *J. Med. Chem.* 43:381–391.
- Surratt, C. K., P. S. Johnson, A. Moriwaki, B. K. Seidleck, C. J. Claschak, J. B. Wang, and G. R. Uhl. 1994. Mu opiate receptor. Charged transmembrane domain amino acids are critical for agonist recognition and intrinsic activity. *J. Biol. Chem.* 269:20548–20553.
- Thirstrup, K., C. E. Elling, S. A. Hjorth, and T. W. Schwartz. 1996. Construction of a high affinity zinc switch in the kappa-opioid receptor. *J. Biol. Chem.* 271:7875–7878.
- Thompson, J. D., D. G. Higgins, and T. J. Gibson. 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.
- Unger, V. M., P. A. Hargrave, J. M. Baldwin, and G. F. Schertler. 1997. Arrangement of rhodopsin transmembrane alpha-helices. *Nature*. 389:203–206.
- Van Vlijmen, H. W. T., and M. Karplus. 1997. PDB-based protein loop prediction: parameters for selection and methods for optimization. *J. Mol. Biol.* 267:975–1001.
- Viguera, A. R., and L. Serrano. 1999. Stable proline box motif at the N-terminal end of alpha-helices. *Protein Sci.* 8:1733–1742.
- White, S. H. 1994. Hydropathy plots and prediction of membrane protein topology. In *Membrane Protein Structure: Experimental Approaches*. S. H. White, editor. Oxford University Press, New York. 97–126.
- Wu, L., N. P. Gerard, R. Wyatt, H. Choe, C. Parolin, N. Ruffing, A. Borsetti, A. A. Cardoso, E. Desjardin, W. Newman, C. Gerard, and J. Sodroski. 1996. CD4-induced interaction of primary HIV-1 gp120 glycoproteins with the chemokine receptor CCR-5 [see comments]. *Nature*. 384:179–183.
- Zhang, L., R. N. DeHaven, and M. Goodman. 2002b. NMR and modeling studies of a synthetic extracellular loop II of the  $\kappa$  opioid receptor in DPC micelle. *Biochemistry*. 41:61–68.
- Zhang, C., J. Hou, and S.-H. Kim. 2002a. Fold prediction of helical proteins using torsion angle dynamics and predicted restraints. *Proc. Natl. Acad. Sci. USA*. 99:3581–3585.
- Zhou, W., C. Flanagan, J. A. Ballesteros, K. Konvicka, J. S. Davidson, H. Weinstein, R. P. Millar, and S. C. Sealfon. 1994. A reciprocal mutation supports helix 2 and helix 7 proximity in the gonadotropin-releasing hormone receptor. *Mol. Pharmacol.* 45:165–170.
- Zhou, N., Z. Luo, J. W. Hall, J. Luo, X. Han, and Z. Huang. 2000. Molecular modeling and site-directed mutagenesis of CCR5 reveal residues critical for chemokine binding and signal transduction. *Eur. J. Immunol.* 30:164–173.