Genetic Analysis of the First and Third Extracellular Loops of the C5a Receptor Reveals an Essential WXFG Motif in the First Loop[*]

Jeffery M. Klco[‡], **Gregory V. Nikiforovich**[§], **and Thomas J. Baranski**[‡1]

From the [‡]*Departments of Medicine and Molecular Biology and Pharmacology and the* [§]*Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, Missouri 63110*

The extracellular loops of G protein-coupled receptors (GPCRs) frequently contain binding sites for peptide ligands. However, the mechanism of receptor activation following ligand binding and the influence of the extracellular loops in other aspects of receptor function are poorly understood. Here we report a structure-function analysis of the first and third extracellular loops of the human C5a receptor, a GPCR that binds a 74-amino acid peptide ligand. Amino acid substitutions were randomly incorporated into each loop, and functional receptors were identified in yeast. The first extracellular loop contains a large number of positions that cannot tolerate amino acid substitutions, especially residues within the WXFG motif found in many rhodopsin-like GPCRs, yet disruption of these residues does not alter C5a binding affinity. These results demonstrate an unanticipated role for the first extracellular loop, and the WXFG motif in particular, in ligand-mediated activation of the C5a receptor. This motif likely serves a similar role in other GPCRs. The third extracellular loop, in contrast, contains far fewer preserved residues and appears to play a less essential role in receptor activation.

G protein-coupled receptors (GPCRs)[2] are the largest class of membrane-bound receptors, with more than 850 members in the human genome (1). The majority of GPCRs is grouped in the rhodopsin family by the presence of a small number of conserved amino acids in the transmembrane (TM) bundle, such as the DRY motif at the cytoplasmic end of TM3 and the NP*XX*Y motif in TM7 (2). These receptors share a similar molecular architecture within the seven-helix bundle, which was originally represented by the α-carbon template of the Baldwin model (3) and later confirmed by the high resolution structure of bovine rhodopsin (4). The extracellular loops, in contrast, are more divergent, both in length and function (5). Ligands frequently bind to the extracellular loops, and the variability of the loops is not surprising when considering the range of ligands known to interact with GPCRs. For many GPCRs, especially those that bind larger peptide ligands, the mechanisms by which ligand binding stimulates G protein activation and how the extracellular loops can influence receptor activation are poorly understood. Considering that more than half of all prescribed medications target GPCRs, mostly by disrupting or mimicking ligand binding (6), a better understanding of the function of the extracellular loops is extremely important.

To elucidate how GPCRs function as molecular switches to transduce signals, we study the complement factor 5a receptor (C5aR), a rhodopsin-like GPCR expressed primarily on the surface of neutrophils and other myeloid cells. C5a, a 74-amino acid peptide released during complement activation, binds to C5aR and directs neutrophil chemotaxis and the release of proteolytic enzymes and superoxide (7). C5a binds to both the transmembrane bundle and the amino terminus of C5aR (8, 9), and other binding sites in the extracellular loops have been proposed (8, 10). We previously performed structure-function analyses on the seven transmembrane helices of C5aR by randomly introducing amino acid substitutions into each helix at a high mutation rate (~40%) and selecting for functional receptors in *Saccharomyces cerevisiae*. This random saturation mutagenesis approach allowed for the identification of residues in the TM bundle that were important for C5aR function, such as in ligand binding, stability, or G protein activation (11, 12).

An elucidation of how the extracellular loops influence C5aR function, whether by ligand binding or other mechanisms, will greatly enhance our understanding of C5aR and potentially other GPCRs that bind peptide ligands. Recently, we applied the random saturation mutagenesis analysis to the second extracellular loop (EC2) and found that introducing amino acid substitutions into EC2 surprisingly generated a large number of constitutively active receptors (13). These results demonstrated that EC2 acts a negative regulator or damper of C5aR activity and that introducing mutations into EC2 can disrupt this function, thus shifting the receptor into an active conformation. In this study, we evaluated the contributions of the first and third extracellular loops to C5aR activation by subjecting the loops to random saturation mutagenesis. In contrast to EC2, substitutions in EC1 did not produce any activated receptors and only three of the 23 receptors containing substitutions in EC3 demonstrated ligand-independent activity. Our studies also found that EC1 and EC3 do not significantly contribute to C5a binding but that EC1 contains a number of residues that are essential for receptor activation following ligand binding. In particular, Trp-102, Phe-104, and Gly-105, within the WXFG motif found in many rhodopsin-like GPCRs, appear to be critical for C5a receptor function.

EXPERIMENTAL PROCEDURES

Library Construction—Silent restriction sites were engineered into the *C5aR* gene at the approximate boundaries of EC1 (BsrGI and HindIII) or EC3 (BstEII and HindIII). To eliminate the potential for wild-type contam-

[1] To whom correspondence should be addressed: Depts. of Medicine and Molecular Biology and Pharmacology, Washington University School of Medicine, Campus Box 8127, 660 S. Euclid Ave., St. Louis, MO 63110. Tel.: 314-747-3997; Fax: 314-362-7641; E-mail: baranski@wustl.edu.
[2] The abbreviations used are: GPCR, G protein-coupled receptor; AT, 3-aminotriazole; C5aR, complement factor 5a receptor; EC1, first extracellular loop; EC2, second extracellular loop; EC3, third extracellular loop; EndoH, endo-β-*N*-acetylglucosaminidase; TM, transmembrane helix; YFP, yellow fluorescent protein; MAPK, mitogen-activated protein kinase.

ination, EC1 and EC3 subcloning vectors were generated containing stuffer DNA between BsrGI and HindIII for EC1 and between BstEII and HindIII for EC3. The following oligonucleotides were used (Integrated DNA Technologies, Coralville, IA) with the underlined bases doped with 20% non-wild-type nucleotides: EC1, TATATTGTACA<u>ACATCATCATTGG</u>-<u>CCATTTGGTGGTGCTGCTTGTTCTTTTTGC</u>CAAGCTTG; EC3, ATATGGTTACC<u>ATATCAAGTTACTGGTATTATGATGTCTT</u>-<u>TTTTGGAACCATCTTCTCCAACTTTTTTGTTGTTGAATAAA</u>-<u>TTGGATTCTTTGTGTG</u>TAAGCTTAC. Double-stranded oligonucleotides were generated by mutual primed synthesis, as described previously (11). The products were digested with BsrGI and HindIII (EC1) or BstEII and HindIII (EC3) and subcloned into the *C5aR* gene in the pBS-SK Bluescript vector. 350,000 and 425,000 recombinants were recovered for the EC1 and EC3 libraries, respectively. Ten unselected recombinants were sequenced to determine the mutation rate of the libraries. The mutant receptors were then subcloned into an ADE2 yeast expression vector. 500,000 and 800,000 recombinants were recovered for the EC1 and EC3 libraries, respectively.

Yeast Transformation and Receptor Selection—The BY1143 yeast strain was made by transforming BY1142 (*MATα far1Δ1442 tbt1-1 P_FUS1-HIS3 can1 ste14::trp1::LYS2 ste3Δ1156 gpa1 (41)-Gα_{i3} lys2 ura3 leu2 trp1 his3 ade2*) with a URA3 plasmid expressing C5a and has been described previously (11). Activation of the MAPK-mating pathway in these cells leads to expression of the *P_FUS1-HIS3* reporter gene, allowing cells to grow in the absence of histidine. BY1143 cells were transformed with the EC1 or EC3 ADE2 libraries by lithium acetate according to standard protocols, and colonies were grown under adenine and uracil selection. Colonies were screened by replica plating on histidine-deficient medium containing 1 or 5 mM 3-amino-1,2,4-triazole (Sigma). The plasmids encoding functional C5aR were recovered, and the phenotype was reconfirmed by assessing growth in BY1143 on 0, 1, 2, 5, 10, 20, and 50 mM AT plates. cDNAs were sequenced at the Washington University Protein and Nucleic Acid Chemistry Laboratory. Constitutive activity was measured by the ability of the recovered C5a receptors to induce growth of BY1144 (BY1142 transformed with a URA3 empty vector). From a library of 500,000 receptors in the ADE2 yeast expression vector, 150,000 EC1 transformants were screened. For EC3, more than 500,000 transformants were screened from the library of 800,000. β-Galactosidase assays in BY1173 (*MATa ura3 leu2 trp1 his3 can1 gpa1Δ::ade2Δ::3XHA far1Δ::ura3Δ fus1::P_FUS1-HIS3 LEU2::PFUS1-lacZ sst2Δ::ura3Δ ste2Δ::G418R trp1::GPA1/Gα_{i1}*) were performed exactly as described previously (13). For Western blot analysis, an equivalent amount of yeast cells was lysed in 1× SDS sample buffer (50 mM Tris, pH 6.8, 2% SDS and 10% glycerol) supplemented with 2% β-mercaptoethanol, 1 μg/ml leupeptin, 1 μg/ml aprotinin, and 500 μM phenylmethylsulfonyl fluoride. The mixture was vortexed in the presence of 0.5-mm glass beads (Biospec Products, Bartlesville, OK) for 5 min and heated for 10 min at 50 °C.

Ligand Binding Studies—HEK293 cells were transfected with YFP-tagged C5a receptors using a standard calcium phosphate protocol. Two days after transfection, total membranes were prepared as described previously (14). 5 μg of protein from the membrane preparations was incubated with 40 pM ^{125}I-C5a in the presence of increasing amounts of unlabeled C5a for 45 min at room temperature. Binding was stopped by addition of cold Hanks' balanced salt solution with 25 mM Hepes, pH 7.3, and 0.1% bovine serum albumin followed by vacuum filtration. The radioactivity count was determined in Scintiverse scintillation fluid (Fisher), and the counts were recorded for 5 min. Binding data were analyzed by nonlinear regression analysis using Prism 4.0 (GraphPad Software, San Diego). No specific C5a binding was detected in untransfected HEK293 cells.

Endoglycosidase Treatment of C5aR—10 μg of total membranes from HEK293 cells were denatured in 1× SDS sample buffer with 2% β-mercaptoethanol and heated at 50 °C for 10 min. Samples were treated with 250 units of endo-β-N-acetylglucosaminidase (EndoH; New England Biolabs) at 37 °C for 3 h. Samples were resolved by SDS-PAGE and immunoblotted with rabbit polyclonal anti-C5aR, raised against residues 9–29 of the amino terminus (14).

Molecular Modeling—Extracellular loops of C5aR were modeled based on the x-ray structures of the extracellular loops of rhodopsin (Protein Data Bank entry 1F88) using sequence alignment by the ClustalW procedure. Small gaps in alignments as well as sterically hindered contacts in the Phe-182/Pro-183/Pro-184/Lys-185 region of EC2 were fixed by manual arrangement of corresponding peptide bonds. The complex containing extracellular loops and stems of TM helices (*i.e.* combining fragments 37–39, 95–110, 171–202, and 264–283) was then subjected to energy minimization employing the ECEPP/2 force field; spatial positions of side chains were optimized with energy minimization. Generally, the modeling procedure was as described previously (15).

Sequence Analysis of Rhodopsin Family GPCRs—All sequences were downloaded from the GPCR data base and aligned by ClustalW analysis (16). For Fig. 7B, the 100 GPCRs analyzed were identified by selecting all of the available unique human GPCRs in the peptide subfamily of the rhodopsin-like GPCRs.

RESULTS

Random Saturation Mutagenesis of EC1 and EC3—Libraries of C5a receptors containing substitutions within the coding region of 16 residues in EC1 or within 28 residues in EC3 were generated by random saturation mutagenesis (Fig. 1A). The boundaries for these studies were determined by the availability of restriction sites for subcloning, which lead to the inclusion of the extracellular ends of TM3 or TM6 and TM7 in the EC1 and EC3 libraries, respectively. The libraries were then screened in a modified *S. cerevisiae* strain, BY1142, that induces *HIS3* expression following activation of the endogenous MAPK-mating pathway, a G protein-mediated response. BY1142 has been engineered to lack the yeast GPCR normally responsible for activation of the mating pathway; thus, the recombinant C5aRs are the only GPCRs present to induce *HIS3* expression. In addition, the endogenous yeast Gα subunit, GPA1, was replaced by a chimeric GPA1/human Gα_{i3} to allow for more efficient coupling. Growth of yeast in histidine-deficient media while coexpressing the C5a ligand identified functional receptors, which were then sequenced. The signaling strength of each receptor was assayed by growth in the presence of 3-aminotriazole (AT), a competitive inhibitor of His3p.

Screening the libraries identified 29 functional receptors containing substitutions in EC1 and 23 receptors with substitutions in EC3 that grew on at least 5 mM AT when coexpressed with C5a (Figs. 2 and 3). This level of signaling is equal to or greater than that of the wild-type C5aR when stimulated with ligand. The selected receptors demonstrated an amino acid mutation rate of 24 and 27% for the EC1 and EC3 recombinants, respectively (Table 1). The number of mutations tolerated in a given receptor was distributed over a wide range as follows: between 1 and 10 mutations for the EC1-mutated receptors and between 2 and 14 mutations for the EC3 receptors. We did not identify any receptors that contained the wild-type sequence; we purposefully set the mutation rate at a high level to minimize the likelihood of selecting wild-type receptors. Sequencing 10 recombinants from each library prior to selection revealed an amino acid mutation rate of 38% in both the EC1 and EC3 libraries. The lower rate of mutation in the functional receptors is expected; more amino acid substitutions increase the

Table B — EC1 alignment (residues 98–113)
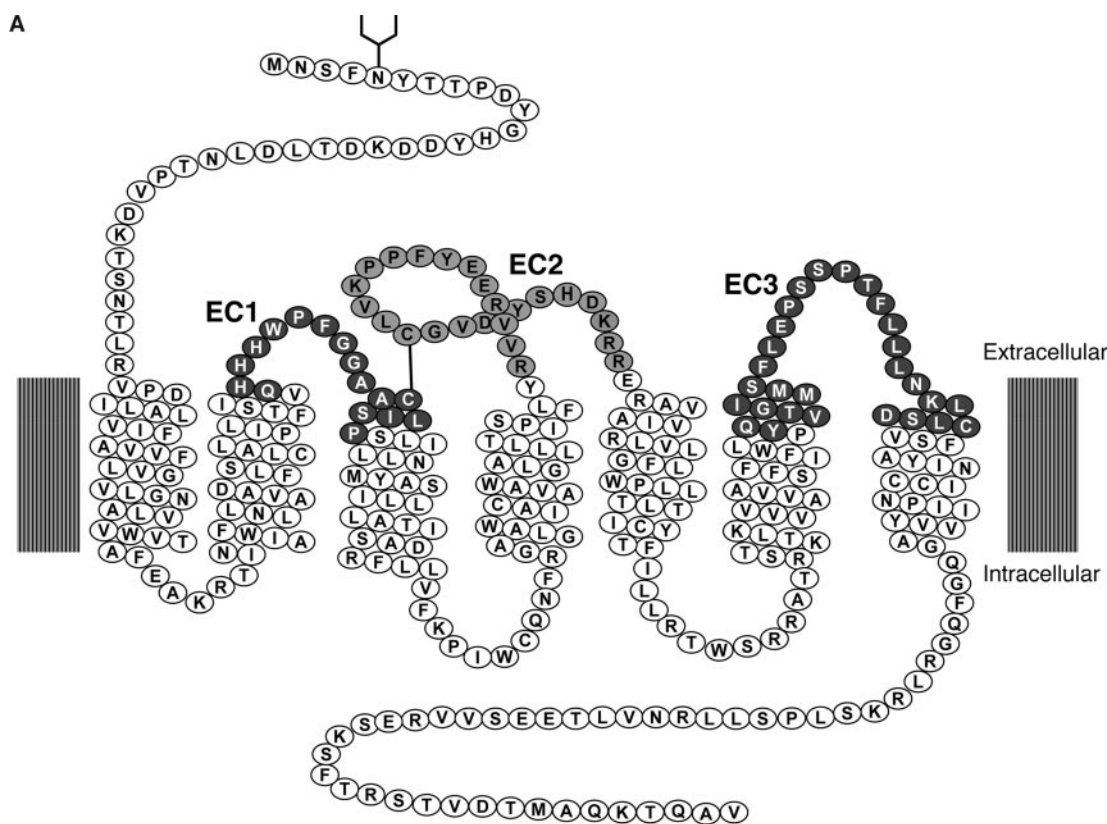
	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113
Human	Q	H	H	H	W	P	F	G	G	A	A	C	S	I	L	P
Chimpanzee	Q	H	H	H	W	P	F	G	G	A	A	C	S	I	L	P
Rhesus monkey	Q	H	H	H	W	P	F	G	G	A	A	C	R	I	L	P
Rabbit	Q	E	G	H	W	P	F	G	R	A	A	C	S	V	L	P
Guinea pig	H	L	N	H	W	Y	F	G	D	T	A	C	K	V	L	P
Mouse	N	H	N	Y	W	Y	F	D	A	T	A	C	I	V	L	P
Rat	K	H	N	H	W	P	F	G	D	Q	A	C	I	V	L	P
Dog	Q	Q	G	Y	W	P	F	G	N	A	A	C	R	I	L	P
Sheep	Q	Q	N	H	W	S	F	G	D	A	A	C	R	I	L	P
Pig	Q	H	N	H	W	S	F	G	D	A	A	C	R	I	L	P

Table C — EC3 alignment (residues 258–285)

	258		260					265					270					275					280					285
Human	Y	Q	V	T	G	I	M	M	S	F	L	E	P	S	S	P	T	F	L	L	L	N	K	L	D	S	L	C
Chimpanzee	Y	Q	V	T	G	I	M	M	S	F	L	E	P	S	S	P	T	F	R	L	L	N	K	L	D	S	L	C
Rhesus monkey	Y	Q	V	T	G	M	M	M	S	F	L	E	P	S	S	P	T	F	L	L	L	K	K	L	D	S	L	C
Rabbit	Y	Q	V	M	G	M	I	L	A	L	L	H	P	S	S	A	T	F	R	W	A	I	R	L	D	P	L	C
Guinea pig	Y	Q	V	T	G	I	L	L	A	W	H	S	P	N	S	A	T	Y	R	N	T	K	A	L	D	A	V	C
Mouse	Y	Q	V	T	G	V	M	I	A	W	L	P	P	S	S	P	T	L	K	R	V	E	K	L	N	S	L	C
Rat	Y	Q	V	T	G	V	I	L	A	W	L	P	R	S	S	S	T	F	Q	S	V	E	R	L	N	S	L	C
Dog	Y	Q	V	T	G	M	M	M	A	L	F	Y	K	H	S	E	S	F	R	R	V	S	R	L	D	S	L	C
Sheep	Y	Q	V	T	G	M	M	L	A	F	F	Q	P	Q	W	D	S	F	K	L	V	S	S	L	D	A	L	C
Pig	Y	Q	V	T	G	V	M	L	A	F	F	H	R	Q	L	D	I	F	K	F	V	S	S	L	D	A	L	C

FIGURE 1. **Schematic diagram of the human C5aR.** *A,* the 16 amino acids (98–113) within EC1 and the 28 residues (258–285) in the EC3 region subjected to random saturation mutagenesis are shaded in *dark gray* with *white lettering*. The 24 residues in EC2 targeted previously (13) are shown in *light gray*. An *N*-linked glycosylation site is represented by a *stick figure,* and the lipid bilayer is depicted as *gray rectangles*. Residues in EC1 (*B*) and EC3 (*C*) of the human C5aR were aligned with the C5aR from different species by ClustalW analysis. Identical residues are highlighted in *black*; homologous residues, as determined by the PAM250 scoring matrix, are highlighted in *gray*. The following GenBank™ accession numbers were used: P21730 (human); P79240 (chimpanzee); P79188 (rhesus monkey); AAF13030 (rabbit); O70129 (domestic guinea pig); P30993 (mouse); P97520 (rat); P30992 (dog); AAG12475 (sheep); and AAG12474 (pig).

chance of deleterious mutations. The functional receptors with mutations in EC1 or EC3 contain fewer amino acid substitutions than the receptors identified in our previous studies on EC2 (13), suggesting a functional role for EC1 and EC3 in receptor activation. However, a direct comparison between the mutation rates of the EC loops is complicated by the fact that the higher mutation rate observed in EC2 may have resulted from the large number of constitutively active receptors obtained in the selection.

FIGURE 2 — Analysis of the first extracellular loop of the C5aR.

EC1	J13	J3	J11	J17	J4	J5	J12	J14	J15	J16	J18	J24	J26	J27	J28	J29	J30	J31	J32	J33	J34	J35	J36	J37	J38	J39	J40	J41	J42	Genetic code	Preserved	Species cons.	EC1
98 Q	H	.	.	.	.	H	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	H	.	.	.	.	.	H	X	X	Q 98
H	.	.	.	.	.	.	Q	Q	R	.	.	.	R	.	Q	R	Q	Q	.	R	Q	.	.	.	Y	.	.	.	R	Y L P R Q N D			H
100 H	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	Q	.	.	.	.	.	.	.	.	.	.	Y L P R Q N D	X		H 100
H	.	.	.	.	.	.	N	.	.	.	.	.	.	.	L	.	.	R	.	.	R	D	Y	.	Q	R	N	Q D	. Q	Y L P R Q N D			H
W	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	@ C G L S R	X	X	W
P	.	H	.	.	.	L	.	.	.	.	.	.	.	.	T	.	.	T	.	.	.	.	.	.	.	.	.	.	.	L Q R T A S			P
F	L	Y	Y	.	Y	Y	.	Y	Y	Y	Y	Y	.	Y	Y	Y	.	.	.	.	.	Y	.	Y	Y	.	Y	Y	Y	C S V Y L I	X	X	F
105 G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	D	.	.	.	.	.	.	.	.	C R V A D S	X	X	G 105
G	.	.	D	.	.	.	A	V	R	D	.	.	R	P	.	D	A	C	P	A	S	D	.	E	D	.	.	.	D	C R V A D S			G
A	T	.	P	.	.	R	.	T	.	T	.	.	R	.	T	.	V	T	.	.	.	.	N	V	T	.	.	.	.	D V G P S T			A
A	.	G	.	.	.	.	G	.	G	.	.	.	.	.	.	.	.	.	.	G	G	.	.	.	G	G	.	.	.	D V G P S T	X	X	A
C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	@ W F R G S Y	X	X	C
110 S	T	.	.	A	.	.	T	.	.	.	T	.	P	T	.	T	.	T	P	T	.	T	.	T	.	T	.	A	C	F Y C A P T			S 110
I	.	.	M	N	.	.	.	.	.	M	.	.	.	.	.	V	V	.	.	.	.	.	.	.	.	.	.	.	.	N S T V L M F		X	I
L	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	V	.	.	.	.	.	.	.	.	.	.	.	.	.	S @ W F M V	X	X	L
113 P	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	S	.	.	.	.	.	.	.	.	.	.	.	T A S	X	X	P 113

Receptor signaling strength (+C5a / −C5a) shown at the bottom of each sequence column; −C5a row all "o" (no constitutive activity).

FIGURE 2. **Analysis of the first extracellular loop of the C5aR.** The sequences of the 29 functional receptors identified are shown, in comparison to the wild-type sequence (*far left* and *right* in *boldface*). The sequences are shown from the extracellular end of TM2 (*top*) to the start of TM3 (*bottom*). The individual mutant receptors are shown in columns with their sequences (*dot* means unchanged from wild-type). Receptor signaling strength is shown at the bottom of each sequence column as follows: ++++++, growth on 50 mM AT; +++++, growth on 20 mM AT; ++++, growth on 10 mM AT; +++, growth on 5 mM AT; ++, growth on 2 mM AT; +, growth on 1 mM AT; 0, no growth on 1 mM AT. Genetic code refers to amino acid substitutions made possible by a single nucleotide change. *Boldface letters* indicate amino acids that are not conserved with respect to the wild-type C5aR sequence. Characteristics of residues in functional receptors (indicated by × in the *boxed columns*) are presented in the following two classes: *Preserved*, amino acid positions where side chain character is preserved (see text); *Species cons.*, residues conserved in at least 10 different mammalian species (see Fig. 1B). @ indicates the stop codon.

Patterns of Preserved Residues—Trp-102 and Cys-109 did not tolerate any amino acid substitutions in the functional receptors identified from the EC1 library screen (Fig. 2). Seven other positions in EC1 were also identified as preserved. We define a residue as being preserved if no amino acid substitutions are tolerated in the mutant receptors or if only conservative substitutions occur, as determined by a score of one or greater in the PAM250 log-odds matrix (17). His-100, Gly-105, and Leu-112 only accepted a single conservative substitution, and although Phe-104 and Ala-108 tolerated multiple changes, all were to conservative amino acids. Phe-104, for example, was mutated in 20 of the 29 receptors, 19 of which were to tyrosine and a single change to leucine. Not all of the nucleotides in codons 98 and 113 were included in the random mutagenesis, and substitutions at those locations were also very uncommon in the library prior to selection, so it is unclear whether Gln-98 and Pro-113 are truly preserved residues. Pro-113 was, however, identified as a preserved residue in our previous analysis of TM3 (11). In contrast, numerous substitutions occurred at the other preserved positions in the unselected library, demonstrating that there was little or no bias in the randomization of the library.

In comparison to EC1, EC3 demonstrated a greater ability to tolerate mutations. Unlike Trp-102 and Cys-109 in EC1, no position in EC3 was completely unchanged in the functional receptors. However, some of the positions, Gln-259, Pro-270, Phe-275, Asn-279, Lys-280, and Asp-282, were preserved because they only accepted conservative substitutions (Fig. 3). Like the preserved locations in EC1, these residues were also frequently altered in the library prior to screening. Regions 258–267 and 282–285 were previously included in analyses of TMs 6 and 7, respectively. Asp-282 in TM7 was also identified as a preserved residue, whereas Gln-259 in TM6 was not preserved in the TM6 scan (11).

As shown in Table 1, the number of preserved residues in both EC1 and EC3 is significantly higher than what we observed in our previous analysis of EC2 (13). In fact, EC1 contains the highest percentage of preserved residues in any domain of the C5a receptor that we have evaluated by random mutagenesis, including the seven transmembrane helices (11–13). The high degree of amino acid preservation, especially in EC1, suggests that the secondary structure of these loops may be important for C5aR function. As discussed below, EC1 and EC3 also differ from EC2 in their ability to regulate the activity state of C5aR, as evidenced by the relative lack of constitutively active receptors resulting from mutations in EC1 or EC3.

The importance of residues in EC1 and EC3 can also be evaluated by their patterns of conservation across evolution (Fig. 1, *B* and *C*). In the EC1 region, 9 of the 16 amino acids are conserved in C5aRs from nine other mammalian species. Strikingly, 8 of the 9 residues were identified as preserved amino acids in our genetic analysis of EC1. For the 28 amino acids scanned in the EC3 region, 15 are highly conserved in the mammalian C5aRs; however, all but one of these map to the transmembrane helices, which typically demonstrate significantly higher levels of amino acid conservation. The only conserved residue within the loop region of EC3 is Phe-275, which is only changed to a leucine or tyrosine in the guinea pig or mouse C5aR, respectively. Interestingly, in our random saturation mutagenesis study, Phe-275 was identified as a preserved residue that only tolerated changes to leucine or tyrosine. The high degree of divergence in the loops, especially in EC3, when compared with the TM helices may reflect a requirement to bind to regions of the C5a ligand that diverge between species or the lack of conservation might reflect a less essential role for EC3 in the C5a receptor function.
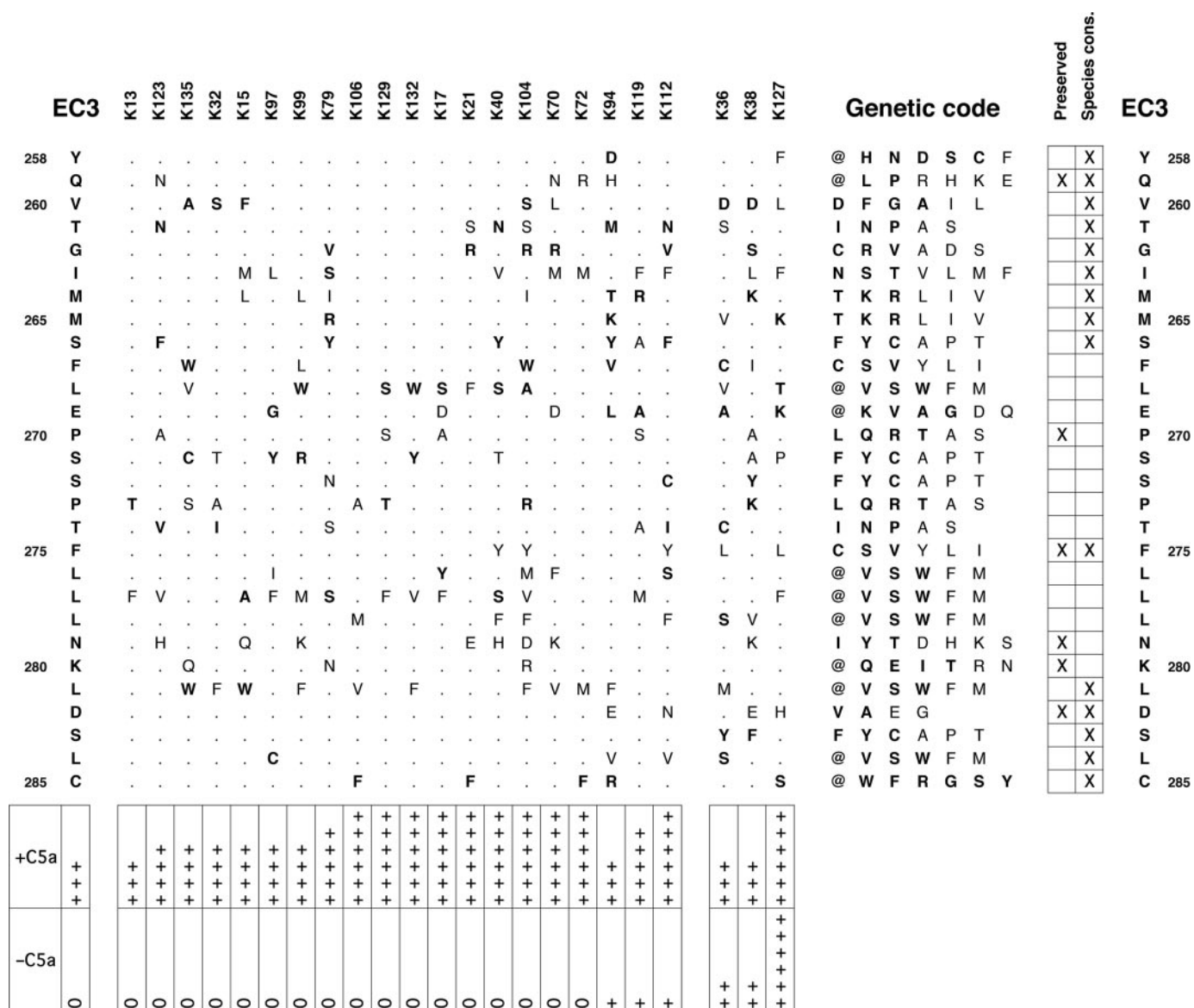
FIGURE 3. Analysis of the third extracellular loop of the C5aR. The sequences of the 23 functional receptors identified are shown. The three constitutively active receptors are shown on the *right*. The sequences are displayed from the end of TM6 (*top*) to the start of TM7 (*bottom*).

Pos	WT	K13	K123	K135	K32	K15	K97	K99	K79	K106	K129	K132	K17	K21	K40	K104	K70	K72	K94	K119	K112	K36	K38	K127	Genetic code	Preserved	Species cons.	EC3
258	Y	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	D	.	.	.	.	F	@ H N D S C F		X	Y 258
	Q	.	N	.	.	.	.	.	.	.	.	.	.	.	N	R	H	.	.	.	.	.	.	.	@ L P R H K E	X	X	Q
260	V	.	.	A	S	F	.	.	.	.	.	.	.	.	.	.	S	L	.	.	.	D	D	L	D F G A I L		X	V 260
	T	.	N	.	.	.	.	.	.	.	.	S	N	S	.	.	M	.	N	.	.	S	.	.	I N P A S		X	T
	G	.	.	.	.	.	.	V	.	.	.	.	R	.	R	R	.	.	.	.	V	.	S	.	C R V A D S		X	G
	I	.	.	.	M	L	.	S	.	.	.	.	.	.	.	V	.	M	M	.	F	.	L	F	N S T V L M F		X	I
	M	.	.	.	.	L	.	L	I	.	.	.	.	.	.	.	I	.	.	T	R	.	.	K	T K R L I V		X	M
265	M	.	.	.	.	.	.	R	.	.	.	.	.	.	.	K	.	.	V	.	K	.	.	K	T K R L I V		X	M 265
	S	.	F	.	.	.	.	Y	.	.	.	.	.	Y	.	.	.	Y	A	F	.	.	.	.	F Y C A P T		X	S
	F	.	.	W	.	.	L	.	.	.	.	.	.	.	.	W	.	V	.	.	C	I	.	.	C S V Y L I			F
	L	.	.	V	.	.	W	.	S	W	S	F	S	A	.	.	V	.	T	.	.	V	.	T	@ V S W F M			L
	E	.	.	.	G	.	.	.	.	.	.	D	.	.	.	D	.	L	A	.	K	A	.	K	@ K V A G D Q			E
270	P	.	A	.	.	.	.	.	.	.	S	.	A	.	.	.	.	S	.	.	A	.	.	.	L Q R T A S	X		P 270
	S	.	C	T	.	Y	R	.	.	.	Y	.	.	T	.	.	.	.	.	.	.	A	.	P	F Y C A P T			S
	S	.	.	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	Y	.	F Y C A P T			S
	P	T	.	S	A	.	.	.	A	T	.	.	R	.	.	.	.	.	.	.	.	.	K	.	L Q R T A S			P
	T	.	V	.	I	.	.	S	.	.	.	.	.	.	.	.	A	I	.	.	.	C	.	.	I N P A S			T
275	F	.	.	.	.	.	.	.	.	.	.	Y	Y	.	.	.	.	.	Y	.	.	L	.	L	C S V Y L I	X	X	F 275
	L	.	.	.	.	I	.	.	.	.	Y	.	.	M	F	.	.	.	S	.	.	.	.	.	@ V S W F M			L
	L	F	V	.	A	F	M	S	.	F	V	F	.	S	V	.	.	M	.	.	S	V	.	.	@ V S W F M			L
	L	.	.	.	.	.	.	M	.	.	F	F	.	.	.	F	.	.	.	.	F	.	.	.	@ V S W F M			L
	N	.	H	.	Q	.	K	.	.	.	E	H	D	K	.	.	.	.	.	.	.	.	K	.	I Y T D H K S	X		N
280	K	.	.	Q	.	.	.	N	.	.	.	.	.	R	.	.	.	.	.	.	.	.	.	.	@ Q E I T R N	X		K 280
	L	.	.	W	F	W	.	F	.	V	.	F	.	.	F	V	M	F	.	M	.	.	.	.	@ V S W F M		X	L
	D	.	.	.	.	.	.	.	.	.	.	.	.	.	E	.	N	.	E	H	V A E G	X	X	D
	S	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	Y	F	.	F Y C A P T		X	S
	L	.	.	.	C	.	.	.	.	.	.	.	.	.	V	.	V	S	.	.	@ V S W F M		X	L
285	C	.	.	.	.	.	.	.	F	.	.	.	F	.	.	F	R	.	.	.	.	S	@ W F R G S Y		X	C 285

Bottom activity rows (+C5a / −C5a): growth scored with "+" and "o" symbols beneath each receptor column; the three constitutively active receptors (K36, K38, K127) show growth in the absence of C5a.

TABLE 1

Characteristics of mutated extracellular loop libraries

AAΔavg indicates the average number of amino acid substitutions per recombinant and the rate of substitution as shown in parentheses. Constitutively active indicates receptors that allow for yeast growth in the presence of 2 mM AT or greater in the absence of C5a. Characteristics of the EC2 library were obtained from a previous study (13).

Library				Selected receptors			
EC loop	Size	Amino acids scanned	AAΔavg	Total	AAΔavg	Preserved residues	Constitutively active
EC1	350,000	16	6.1 (38%)	29	3.9 (24%)	9	0
EC2	40,000	24	10.5 (44%)	29	8.6 (36%)	1	23
EC3	425,000	28	10.5 (38%)	23	7.5 (27%)	6	3

Constitutive Activity—In our previous study we found that substitutions in EC2 of the C5a receptor induce constitutive activity (13). To determine whether this is a common result of altering amino acids in the extracellular loops of C5aR, we tested the 52 functional receptors containing substitutions in EC1 or EC3 for ligand-independent activity (Figs. 2 and 3). None of the receptors with mutations in EC1 demonstrated constitutive activity. Of the EC3-mutated receptors, only one receptor, K127, demonstrated strong constitutive activity, and two other receptors (K36 and K38) were moderately active in the absence of

ligand. K127, containing 11 substitutions within the EC3 region, demonstrated ligand-independent growth at 50 mM AT, the highest concentration tested. An explanation for the constitutive activity of K127 is not obvious. However, K127 does have a substitution at position 258 in TM6, which was frequently mutated in constitutively active receptors with mutations in TM6 from previous studies (11). It is therefore possible that the Y258F substitution in K127, although a conservative change, may be responsible for constitutive activity rather than the substitutions in EC3. There are, however, other substitutions in K127 that

TABLE 2

Effect of single point mutations in EC1

Yeast strain BY1173 was cotransformed with C5aR-YFP, either wild-type or mutant, together with a plasmid encoding C5a (+C5a) or empty vector (−C5a). Receptor signaling was assayed by growth on histidine-deficient media in the presence of aminotriazole: +++++, growth in 50 mM AT; −, no growth in 1 mM AT. Data are representative of three separate experiments.

	+C5a	−C5a
Wild type	+++++	−
H100A	−	−
W102A	−	−
F104A	−	−
G105L	−	−
A108L	+++++	−

may account for the activated phenotype. For example, E269K and S271P mutations are only found in K127. Nonetheless, an E269K/S271P double mutation was not sufficient for ligand-independent activity nor was it necessary for constitutive activity in the K127 receptor.[3] The relative lack of constitutively active receptors resulting from mutations in EC1 and EC3 is in stark contrast to our previous analysis of EC2, in which 23 of the 29 functional receptors identified were ligand-independent, despite all three libraries sharing similar overall characteristics (Table 1).

Essential Residues in EC1—To validate the findings of the saturation mutagenesis study on EC1, preserved residues were targeted by site-directed mutagenesis of the wild-type C5aR containing a carboxyl-terminal YFP. Receptors with H100A, W102A, F104A, or G105L substitutions failed to induce yeast growth in histidine-deficient media when coexpressed with C5a (Table 2), consistent with their preservation in the functional receptors identified in the library screen. Surprisingly, despite tolerating only changes to glycine in the receptors identified from the EC1 library screen, introduction of a leucine for Ala-108 did not affect signaling.

In these studies, receptor activation was monitored in the yeast strain BY1173, which lacks the yeast RGS protein, SstII, resulting in higher levels of MAPK activation. This yeast strain also contains a β-galactosidase reporter that responds to activation of the MAPK pathway, thus providing a growth-independent assay of signaling. To induce β-galactosidase, cells are treated with the hexapeptide agonist W5Cha, which in contrast to C5a can cross the yeast cell wall. Thus, another advantage of using the BY1173 strain is that C5aR activity can be monitored in response to two different ligands as follows: the 74-amino acid C5a, which interacts with multiple regions of the receptor, and W5Cha, a synthetic hexapeptide that binds to a pocket within the transmembrane bundle (18). In contrast to the results from the growth assay, the mutations in EC1, with the exception of a W102A mutation, did not completely disrupt receptor function when treated with W5Cha (Fig. 4*A*). However, only receptors with H100A or A108L mutations demonstrated an EC_{50} value similar to that of the wild-type receptor. Despite displaying virtually no activity, the W102A receptor was expressed at similar levels to the other receptors (Fig. 4*B*). Therefore, Trp-102, and to a lesser extent Phe-104 and Gly-105, is essential for activation of the C5aR in response to both C5a and W5Cha. His-100, in contrast, appears to only be important in receptor activation following C5a stimulation. These studies did not address Cys-109 and Leu-112 because those positions were evaluated previously. Cys-109 forms a disulfide bond with Cys-188 in EC2, and disruption of the disulfide bond blocks ligand-dependent activation (13). This disulfide bond is found in over 91% of all human GPCRs and is important in many other receptors (19). Leu-112
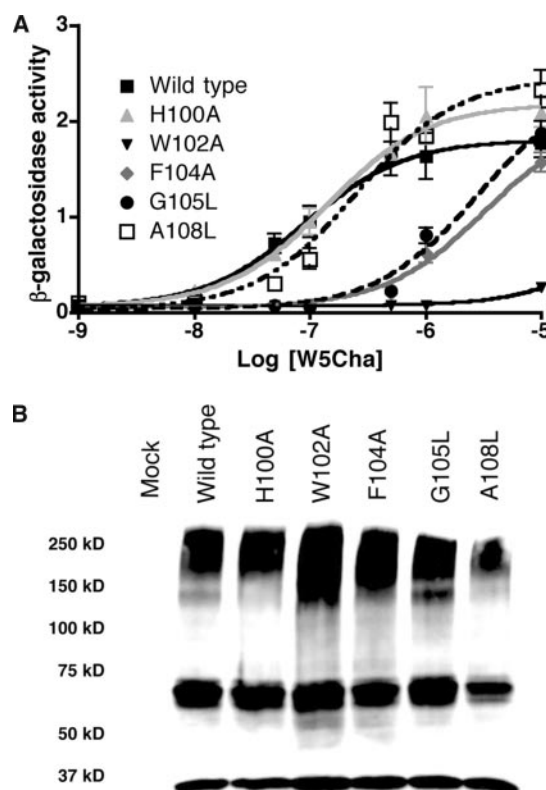
FIGURE 4. **Disruption of preserved residues in EC1.** *A*, BY1173 cells were transfected with YFP-tagged C5aRs containing H100A, W102A, F104A, G105L, and A108L mutations or the wild-type sequence. Receptor activity was measured after a 4-h treatment with W5Cha. The mean of each experiment performed in triplicate is shown ± S.E. and is representative of two separate transformations. *B*, expression of EC1 mutants in BY1173. Total cell lysates were resolved by SDS-PAGE and immunoblotted with anti-C5aR antibody. Monomeric C5aR-YFP migrates at ∼70 kDa, and the higher molecular species are SDS-resistant aggregates.

TABLE 3

Effects of Trp-102 mutations on C5aR activity

Yeast strain BY1173 was cotransformed with C5aR-YFP, either wild type or mutant, together with a plasmid encoding C5a (+C5a) or empty vector (−C5a). Receptor signaling was assayed by growth on histidine-deficient media in the presence of aminotriazole: +++++, growth in 50 mM AT; −, no growth in 1 mM AT. NQ indicates I124N/L127Q; R17 indicates constitutively active receptor described previously (13). Data are representative of three separate experiments.

	+C5a	−C5a
Wild type	+++++	−
W102A	−	−
W102V	−	−
W102F	+++++	−
NQ	+++++	+++++
NQ, W102A	+++++	+++++
R17	+++++	+++++
R17, W102A	+++++	+++++

was identified as an essential residue in TM3 (11), and a receptor with a L112A mutation failed to respond to C5a (18).

Many peptide GPCRs also have a tryptophan at a similar location in EC1, suggesting that this residue may be important for GPCR function. To further investigate the role of Trp-102 in C5aR activation, we also generated receptors containing W102V and W102F mutations. A W102V mutation, similar to a W102A mutation, resulted in a nonfunctional receptor, whereas the activity of a C5a receptor with a W102F substitution was indistinguishable from the wild-type receptor (Table 3). The ability of a phenylalanine to functionally replace tryptophan demonstrates the importance of a large hydrophobic surface at position 102. One possibility is that Trp-102 is necessary for C5aR to achieve an active conformation. To address this, a W102A substitution was intro-

---

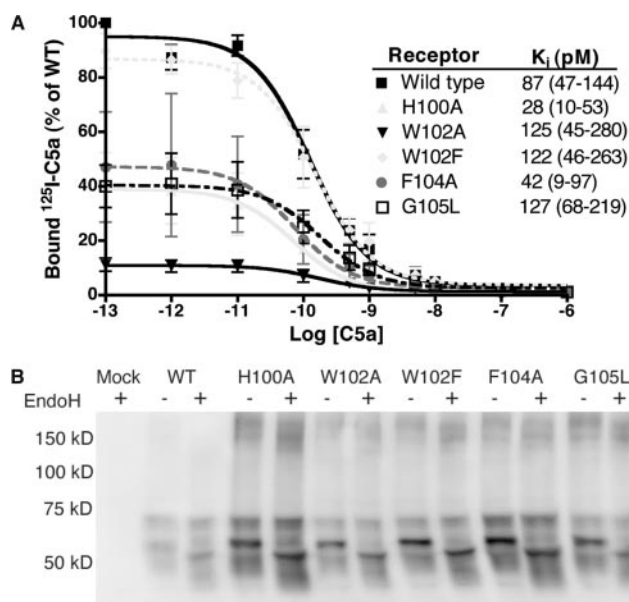[3] J. M. Klco, unpublished results.

FIGURE 5. **Ligand binding and expression of EC1 mutants in mammalian cells.** *A,* membranes were prepared from HEK293 cells transfected with YFP-tagged wild-type, H100A, W102A, W102F, F104A, and G105L C5a receptors. Binding of 40 pm ^{125}I-C5a was competed with varying amounts of unlabeled C5a. Binding is normalized to the maximal binding to the wild-type receptor. The mean binding from three separate experiments ± S.E. is shown. K_i values were determined following nonlinear regression; 95% confidence intervals are shown in *parentheses*. *B,* solubilized membranes were treated with (+) and without (−) EndoH and then analyzed by immunoblotting for C5aR. EndoH-resistant C5a receptors containing complex *N*-linked oligosaccharides (*) and EndoH-sensitive high mannose oligosaccharides (+) are shown. *WT,* wild type.

duced into two different classes of constitutively active receptors. Constitutive C5aR activity can be induced by mutations in the second extracellular loop, such as those in the R17 C5a receptor described previously (13), or by disrupting a hydrophobic cluster near the cytoplasmic ends of TMs 3, 6, and 7, as in the NQ C5a receptor (11). If a W102A substitution prevented these activated receptors from stimulating the MAPK pathway, it would imply that Trp-102 was important for the overall three-dimensional structure of C5aR. A W102A mutation, however, did not alter constitutive activity of the R17 receptor or the NQ receptor (Table 3); therefore, Trp-102 does not impact receptor stability or the ability of the receptor switch mechanism in the TM bundle to activate G proteins. These "intramolecular epistasis" results also suggest that Trp-102 acts upstream of both EC2 and the TM bundle in C5aR activation. Another explanation for the essential role of Trp-102 and other residues in EC1 is that these residues participate directly or indirectly in binding C5a. However, receptors containing H100A, W102A, F104A, or G105L substitutions demonstrated binding affinities for C5a similar to that of the wild-type receptor (Fig. 5*A*). These findings are consistent with other studies on the first extracellular loop of C5aR (20). Despite not altering the binding affinity for C5a, we observed less overall binding to receptors with mutations in EC1. All the receptors were expressed at a level at or above that of the wild-type C5aR, and the amounts of complex *N*-linked oligosaccharides were also similar, suggesting that comparable levels of receptor were processed in the secretory pathway and reached the plasma membrane (Fig. 5*B*). Therefore, despite an unaltered number of receptors, binding sites were less abundant. The C5a receptor is found in both a high and low affinity state for C5a (21), and a likely interpretation of the binding data is that mutations in EC1, especially a W102A mutation, decrease the fraction of receptors in the high affinity state. A similar effect was observed with an R206Q mutation near the extracellular side of TM5 in C5aR (22). Unfortunately, binding of C5a to the low affinity receptors, which have a K_d ~100 nM in eosinophils (21),

TABLE 4
Effect of single point mutations in EC3 in BY1173

Yeast strain BY1173 was cotransformed with the wild-type C5aR or receptors containing mutations in EC3, together with C5a (+C5a) or empty vector (−C5a). All receptors contain YFP attached to their C terminus. Receptor signaling was assayed by growth on histidine-deficient media in the presence of aminotriazole: +++++, growth in 50 mM AT; −, no growth in 1 mM AT. Data are representative of three separate experiments.

	+C5a	−C5a
Wild type	+++++	−
P270A	+++++	−
F275A	+++++	−
N279A	+++++	−
K280A	+++++	−
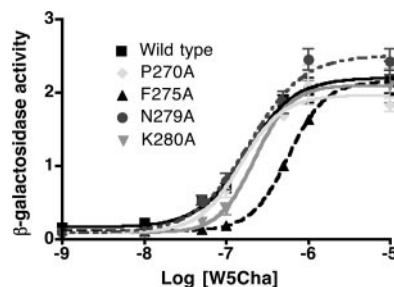
FIGURE 6. **Disruption of preserved residues in EC3.** BY1173 cells were transfected with YFP-tagged wild-type C5aR or receptors containing P270A, F275A, N279A, or K280A mutations. Receptor activity was measured after a 4-h treatment with W5Cha. The mean of each experiment performed in triplicate is shown ± S.E. and is representative of two separate transformations.

would be too weak to be observed in our bindings studies. As expected from the functional data in yeast, a W102F substitution had no effect on the K_i value or on the number of high affinity binding sites.

Essential Residues in EC3—In contrast to Trp-102 and Cys-109 in EC1, no residue in EC3 was found unchanged in all of the mutant, yet functional, receptors. However, Gln-259, Pro-270, Phe-275, Asn-279, Lys-280, and Asp-282 tolerated only a few conservative substitutions, suggesting that these residues might be important for C5aR function. As with our evaluation of the preserved residues in EC1, alanine was introduced for Pro-270, Phe-275, Asn-279, and Lys-280. Gln-259 in TM6 and Asp-282 in TM7 were not evaluated further because both locations were studied previously in our analyses of TMs 6 and 7 (11). Surprisingly, receptors containing P270A, F275A, N279A, and K280A signaled as well as the wild-type receptor, as assayed by growth on aminotriazole (Table 4). The mutant receptors did not demonstrate any ligand-independent activity; therefore, the observed activity results from C5a binding. When signaling was monitored by β-galactosidase induction, an F275A mutation produced a small but significant increase in EC_{50} when stimulated with W5Cha, suggesting that an F275A mutation may have a minor deleterious effect on receptor function (Fig. 6). The failure of alanine substitutions at these four locations to disrupt receptor activation suggests that these residues are not essential for C5a-dependent activation despite their preservation in the functional receptors identified in the library screen. It is possible that these amino acids play a minor role in the wild-type receptor that can be compensated for by other neighboring wild-type residues, and only when EC3 incurs numerous mutations do these residues play a more significant role.

DISCUSSION

In this study, the first and third extracellular loops of the C5a receptor were evaluated by randomly introducing amino acid substitutions into each loop and screening for functional receptors in yeast. We found that EC1 is extremely resistant to amino acid substitutions and that it is essential in receptor activation but not ligand binding, with Trp-102,

Phe-104, and Gly-105 playing the most significant roles. EC3, in contrast, contained significantly fewer residues that were essential for C5aR function. In combination with our previous studies on EC2, we now have a comprehensive functional map of the three extracellular loops of the C5aR.

Many rhodopsin family GPCRs, especially receptors for peptides and biogenic amines, contain a W*X*FG motif in EC1 (Fig. 7*A*). To appreciate the conservation of this motif, we evaluated 100 human rhodopsin-like GPCRs in the peptide subfamily made available at the GPCR data base. We found that 81 of these receptors have a tryptophan in the first position and that 8 receptors contain a phenylalanine (Fig. 7*B*). Thus, almost 90% of GPCRs in the peptide subfamily have the aromatic residues tryptophan or phenylalanine in the first position. Only a single

receptor contained the aromatic tyrosine side chain; the reduction in hydrophobicity of tyrosine relative to phenylalanine, because of its polar hydroxyl group, might account for its relative absence from this motif. There is no consensus residue at the second position in this motif, although 24% of the receptors have a proline, as in the C5aR. At the third position, phenylalanine is the most common side chain but it is frequently replaced by another hydrophobic residue. Glycine is found at the fourth position 84 times, yet there is no general pattern in the remaining receptors. Considering these patterns, we found that 78 of the 100 human peptide-binding GPCRs evaluated have a (W/F)*X*ΦG sequence, in which Φ represents the hydrophobic amino acids valine, isoleucine, leucine, phenylalanine, tryptophan, tyrosine, or methionine. Our results on the C5aR presented here suggest that the conservation of this motif is important for receptor activation. We demonstrated that a tryptophan or phenylalanine, but not alanine or valine, is necessary in the first position and that the third position was either the wild-type phenylalanine or the hydrophobic residues tyrosine or leucine in the functional receptors identified by random mutagenesis.

Within the peptide subfamily of the rhodopsin-like GPCRs, the melanocortin receptors are the major example of receptors that do not contain the W*X*FG motif. However, these receptors are unique in that they virtually lack an EC2 and do not share the TM3-EC2 disulfide bond present in most rhodopsin-like GPCRs. Furthermore, these receptors demonstrate high basal activity (23). In the setting of constitutively active GPCRs, this motif might be dispensable, as we observed here in C5aRs that contain activating mutations. The W*X*FG motif is also not found in the olfactory receptors, as well as the more divergent secretin and glutamate families of GPCRs and the yeast GPCR Ste2. Although Ste2 lacks the W*X*FG motif, it has similarly been suggested that its EC1 is important in initiating receptor activation following pheromone binding (24).

To gain a better understanding of the requirement for the aromatic and hydrophobic residues in W*X*FG motif in EC1, we generated a three-dimensional model of the extracellular loops of the C5aR. The resulting three-dimensional model displayed conformations of EC1 and EC3 that are very similar to those in the x-ray structure of rhodopsin; however, the spatial position of EC2 is slightly different, showing a small movement away from the TM bundle. Most likely, this movement resulted from replacement of the Pro-170/Pro-171 fragment in rhodopsin by a more flexible Pro-170/Ser-171 in C5aR (Fig. 8). The molecular model predicts that the Trp-102 side chain of the W*X*FG motif resides within a hydrophobic pocket primarily formed by Val-186 in EC2 and the Cys-

Rhodopsin	99	L	H	G	Y	F	V	F	G	P	T	G	C
Peptide	C5aR	98	Q	H	H	H	W	P	F	G	G	A	A	C
	NK1R	94	V	H	N	E	W	Y	Y	G	L	F	Y	C
	NPY2R	112	L	M	G	E	W	K	M	G	P	V	L	C
	CCR5	92	A	Q	–	–	W	D	F	G	N	T	M	C
	SS1R	120	L	R	H	–	W	P	F	G	A	L	L	C
	PAR1	164	S	G	S	D	W	Q	F	G	S	E	L	C
	OT-R	101	I	T	F	R	F	Y	G	P	D	L	L	C
	CXCR4	100	A	N	–	–	W	Y	F	G	F	L	L	C
	CCK-AR	103	L	L	K	D	F	I	F	G	S	A	V	C
	AT1AR	90	M	E	Y	R	W	P	F	G	N	Y	L	C
Biogenic Amine	5-HT1A	98	V	L	N	K	W	T	L	G	Q	V	T	C
	D2R	96	V	V	G	E	W	K	F	S	R	I	H	C
	β2AR	95	L	M	K	M	W	T	F	G	N	F	W	C
	Consensus				L		W		F	G			L	C

Motif	Residue	Frequency
W	W	81%
	F	8%
x		
F	F	70%
	L	14%
	Φ	90%
G	G	84%

FIGURE 7. **W*X*FG motif in EC1.** *A,* alignment of bovine rhodopsin and representative human GPCRs for peptide ligands and biogenic amines in the rhodopsin family. Sequences span from the approximate beginning of EC1 to the conserved cysteine at the top of TM3. The following abbreviations are used: substance P receptor, *NK1R*; neuropeptide Y receptor type 2, *NPY2R*; C-C chemokine receptor type 5, *CCR5*; proteinase-activated receptor 1 [precursor], *PAR1*; oxytocin receptor, *OT-R*; C-X-C chemokine receptor type 4, *CXCR4*; cholecystokinin type A receptor, *CCK-AR*; type 1 angiotensin II receptor, *AT1AR*; 5-hydroxytryptamine 1A receptor, *5-HT1A*; dopamine D2 receptor, *D2R*; β₂-adrenergic receptor, *β2AR*. *B,* frequency of side chains in the W*X*FG motif in 100 human peptide-binding GPCRs in the rhodopsin family obtained from the GPCR data base. Φ represents valine, isoleucine, leucine, phenylalanine, tryptophan, tyrosine, or methionine (35).

FIGURE 8. **Three-dimensional model of the extracellular loops of C5aR.** Loops are shown as *tubes* in *red* (EC1), *green* (EC2), and *magenta* (EC3). Not all side chains are shown. Atoms are shown using CPK coloring (carbon, *gray*; nitrogen, *blue*; oxygen, *red*; and sulfur, *yellow*), unless otherwise noted. Hydrogens are not shown. *A,* hydrophobic pocket involving EC1 and EC2. Trp-102 and Phe-104 are shown in *cyan*. *B,* the 11 preserved residues in the extracellular loops are shown. Preserved residues in TM3 (Leu-112), TM6 (Gln-259), and TM7 (Asp-282) are not shown. Trp-102, Phe-104, and Gly-105 of the conserved W*X*FG motif are shown in *cyan*. Stems of TM helices adjacent to loops are shown by semi-transparent atomic volume contours in *gray*.
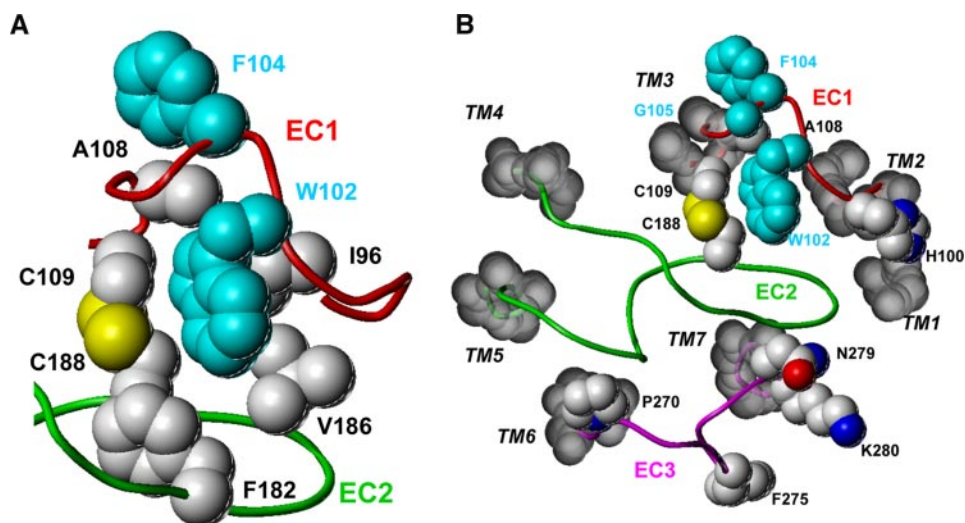
109/Cys-188 disulfide bridge (Fig. 8*A*). Interestingly, our previous work on EC2 demonstrated that position 186 tolerated only hydrophobic residues (13). The ability of phenylalanine, but not valine or alanine, to functionally replace a tryptophan at position 102 might reflect an essential packing requirement of a bulky hydrophobic residue within the pocket. In contrast, Phe-104 points away from the hydrophobic pocket (Fig. 8), and the requirement for a hydrophobic side chain at this position is not apparent. It is possible that Phe-104 can indirectly influence the interaction of Trp-102 with the hydrophobic pocket by contacting Ala-108, which resides in the hydrophobic pocket, thus preventing its solvation. On the other hand, our homology model is based on the single snapshot of the x-ray structure of rhodopsin in its inactive state, and the possibility exists that movements of EC1 during receptor activation may bring Phe-104 closer to the helix bundle. The presence of a glycine residue, which can impart flexibility to the peptide backbone, in the W*X*FG motif may allow for this movement.

What is the role of the W*X*FG motif in the activation mechanism of GPCRs? Our findings that mutations in the W*X*FG motif disrupt receptor activation but not the binding affinity for C5a argue that the W*X*FG motif operates downstream of ligand binding but upstream of the TM bundle switch mechanism. We propose that this motif may be important in translating the ligand-binding signal directly to movements within the TM bundle. This action would require structural rigidity in EC1 so that movements in the loop could be directly transferred to TMs 2 and 3, instead of the loop accommodating part of the movement. Indeed, biophysical analysis of an EC1 peptide from the AT$_{1A}$ receptor, which like C5aR contains a WPFG sequence in EC1, showed that the fragment in solution maintains a distinct type II β-turn, indicating that possible conformations of the EC1 peptide are limited (25). Also, a molecular modeling study of possible low energy conformers of the extracellular and intracellular loops in rhodopsin both in the dark-adapted and activated states predicted that EC1 was significantly less flexible than EC3 and EC2 (15).

The W*X*FG motif likely serves an additional role in regulating the percentage of receptors that are in the high affinity ligand binding state. GPCRs exist in low affinity and high affinity receptor populations, the latter of which can be favored by coupling to G proteins (26, 27). In other receptors, such as the tachykinin NK2 receptor (28), M$_1$ muscarinic receptor (29), AT$_1$ receptor (30), and cholecystokinin receptors (31), disruption of the residues in the W*X*FG motif resulted in an apparent failure to bind ligand. The loss of binding reported for these receptors, however, could have been due to a decrease in the number of high affinity binding sites to levels that could not be detected. In the C5aR, mutating Trp-102, Phe-104, or Gly-105 decreased the total number of high affinity sites but did not affect the binding affinity of those receptors that remained in the high affinity state. Furthermore, the preservation of the W*X*FG motif and the fact that GPCRs respond to ligands of different size and character make it doubtful that the W*X*FG motif directly participates in ligand binding. It is also unlikely that mutations in the W*X*FG motif destabilize the receptor because receptors containing substitutions at Trp-102, Phe-104, and Gly-105 were expressed at levels similar to the wild-type receptor and had similar C5a binding affinities.

Two distinct binding sites on C5aR for C5a have been identified, the amino terminus and the transmembrane bundle, yet other studies proposed that the extracellular loops may make additional contacts (8–10, 32). Although not specifically tested, the failure of single point mutations at preserved residues in EC3 to abolish signaling in response to C5a suggests that EC3, like EC1, is also not a major binding surface for C5a. A previous study on C5aR did find that a P270L mutation in EC3

disrupted ligand binding and inositol phosphate liberation in COS cells (33). As shown in Fig. 6 and Table 4, a receptor with a P270A mutation demonstrated ligand-dependent activation similar to the wild-type receptor. The reason for this discrepancy is unclear; however, alanine is a more conservative substitution for proline than leucine. Furthermore, a limitation of our analysis in yeast is that the concentration of C5a ligand in the periplasmic space is unknown; residues that make minor contributions to ligand binding will become unnecessary if the ligand concentration is especially high. Although no major roles in C5a binding were observed for EC1 and EC3, it is still reasonable to assume that the extracellular loops can influence binding but do not substantially impact the overall binding energy.

Our studies on the extracellular loops of C5aR generated 81 different receptors with 545 random substitutions at 68 locations in EC1, EC2, and EC3. The pattern of amino acid conservation in these receptors provides a comprehensive functional map of the extracellular loops of the C5a receptor (Fig. 8*B*). Moreover, our studies along with previous work by other groups on the amino terminus of the C5aR provide evidence that each extracellular domain has a unique role in C5a receptor function and supports the two-site model of C5a binding. The amino terminus is required for high affinity binding of C5a (8, 9), and our results on EC1, EC2, and EC3 demonstrate that these regions appear to play less significant roles in binding to C5a. After binding, EC1 and its W*X*FG motif are necessary for receptor activation, most likely by inducing movement of the transmembrane helices. EC2 is a negative regulator of receptor activation, a function that can be disrupted by amino acid substitutions to generate ligand-independent activity. The possibility exists that one of the consequences of ligand binding is to induce movement of EC2, thus allowing receptors to assume their active conformation. Finally, EC3 does not appear to have a major role in receptor function, yet it is not completely dispensable considering the small number of functional receptors isolated from a library of over 400,000 recombinants. It is likely that the structural integrity of EC3 is important for proper spacing and movement of TMs 6 and 7 during receptor activation, yet no single amino acid is essential for this function. In contrast, residues in EC3 in other receptors have a well defined role in ligand binding and receptor activation (34).

These studies, in combination with our previous work on EC2, represent the most comprehensive evaluation of the extracellular loops of a GPCR. Our findings have begun to uncover how the extracellular loops collaborate with the activation switch in the transmembrane bundle and will deepen our understanding of not only the C5a receptor but also other GPCRs that bind peptide ligands. Specifically, the W*X*FG motif in EC1 described here may be a key component in the activation mechanism in many GPCRs in the rhodopsin family.

REFERENCES

1. Fredriksson, R., and Schioth, H. B. (2005) *Mol. Pharmacol.* **67,** 1414–1425
2. Gether, U. (2000) *Endocr. Rev.* **21,** 90–113
3. Baldwin, J. M., Schertler, G. F., and Unger, V. M. (1997) *J. Mol. Biol.* **272,** 144–164
4. Palczewski, K., Kumasaka, T., Hori, T., Behnke, C. A., Motoshima, H., Fox, B. A., Le Trong, I., Teller, D. C., Okada, T., Stenkamp, R. E., Yamamoto, M., and Miyano, M. (2000) *Science* **289,** 739–745
5. Mirzadegan, T., Benko, G., Filipek, S., and Palczewski, K. (2003) *Biochemistry* **42,** 2759–2767
6. Leurs, R., Smit, M. J., Alewijnse, A. E., and Timmerman, H. (1998) *Trends Biochem. Sci.* **23,** 418–422
7. Gerard, C., and Gerard, N. P. (1994) *Annu. Rev. Immunol.* **12,** 775–808

8. Chen, Z., Zhang, X., Gonnella, N. C., Pellas, T. C., Boyar, W. C., and Ni, F. (1998) *J. Biol. Chem.* **273,** 10411–10419

9. DeMartino, J. A., Van Riper, G., Siciliano, S. J., Molineaux, C. J., Konteatis, Z. D., Rosen, H., and Springer, M. S. (1994) *J. Biol. Chem.* **269,** 14446–14450

10. Huber-Lang, M. S., Sarma, J. V., McGuire, S. R., Lu, K. T., Padgaonkar, V. A., Younkin, E. M., Guo, R. F., Weber, C. H., Zuiderweg, E. R., Zetoune, F. S., and Ward, P. A. (2003) *J. Immunol.* **170,** 6115–6124

11. Baranski, T. J., Herzmark, P., Lichtarge, O., Gerber, B. O., Trueheart, J., Meng, E. C., Iiri, T., Sheikh, S. P., and Bourne, H. R. (1999) *J. Biol. Chem.* **274,** 15757–15765

12. Geva, A., Lassere, T. B., Lichtarge, O., Pollitt, S. K., and Baranski, T. J. (2000) *J. Biol. Chem.* **275,** 35393–35401

13. Klco, J. M., Wiegand, C. B., Narzinski, K., and Baranski, T. J. (2005) *Nat. Struct. Mol. Biol.* **12,** 320–326

14. Klco, J. M., Lassere, T. B., and Baranski, T. J. (2003) *J. Biol. Chem.* **278,** 35345–35353

15. Nikiforovich, G. V., and Marshall, G. R. (2005) *Biophys. J.* **89,** 3780–3789

16. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) *Nucleic Acids Res.* **22,** 4673–4680

17. Dayhoff, M. O., Schwartz, R. M., and Orcutt, B. C. (1978) in *Atlas of Protein Sequence and Structure* (Dayhoff, M. O., ed) Vol. 5, pp. 345–353, National Biomedical Research Foundation, Washington, D. C.

18. Gerber, B. O., Meng, E. C., Dotsch, V., Baranski, T. J., and Bourne, H. R. (2001) *J. Biol. Chem.* **276,** 3394–3400

19. Karnik, S. S., Gogonea, C., Patil, S., Saad, Y., and Takezako, T. (2003) *Trends Endocrinol. Metab.* **14,** 431–437

20. Cain, S. A., Woodruff, T. M., Taylor, S. M., Fairlie, D. P., Sanderson, S. D., and Monk, P. N. (2001) *Biochem. Pharmacol.* **61,** 1571–1579

21. Gerard, N. P., Hodges, M. K., Drazen, J. M., Weller, P. F., and Gerard, C. (1989) *J. Biol. Chem.* **264,** 1760–1766

22. Raffetseder, U., Roper, D., Mery, L., Gietz, C., Klos, A., Grotzinger, J., Wollmer, A., Boulay, F., Kohl, J., and Bautsch, W. (1996) *Eur. J. Biochem.* **235,** 82–90

23. Holst, B., and Schwartz, T. W. (2003) *Ann. N. Y. Acad. Sci.* **994,** 1–11

24. Akal-Strader, A., Khare, S., Xu, D., Naider, F., and Becker, J. M. (2002) *J. Biol. Chem.* **277,** 30581–30590

25. Nicastro, G., Peri, F., Franzoni, L., de Chiara, C., Sartor, G., and Spisni, A. (2003) *J. Pept. Sci.* **9,** 229–243

26. De Lean, A., Stadel, J. M., and Lefkowitz, R. J. (1980) *J. Biol. Chem.* **255,** 7108–7117

27. Unson, C. G., Wu, C. R., Sakmar, T. P., and Merrifield, R. B. (2000) *J. Biol. Chem.* **275,** 21631–21638

28. Labrou, N. E., Bhogal, N., Hurrell, C. R., and Findlay, J. B. (2001) *J. Biol. Chem.* **276,** 37944–37949

29. Matsui, H., Lazareno, S., and Birdsall, N. J. (1995) *Mol. Pharmacol.* **47,** 88–98

30. Hjorth, S. A., Schambye, H. T., Greenlee, W. J., and Schwartz, T. W. (1994) *J. Biol. Chem.* **269,** 30953–30959

31. Silvente-Poirot, S., Escrieut, C., and Wank, S. A. (1998) *Mol. Pharmacol.* **54,** 364–371

32. Pease, J. E., Burton, D. R., and Barker, M. D. (1994) *Eur. J. Immunol.* **24,** 211–215

33. Kolakowski, L. F., Jr., Lu, B., Gerard, C., and Gerard, N. P. (1995) *J. Biol. Chem.* **270,** 18077–18082

34. Lawson, Z., and Wheatley, M. (2004) *Biochem. Soc. Trans.* **32,** 1048–1050

35. Aasland, R., Abrams, C., Ampe, C., Ball, L. J., Bedford, M. T., Cesareni, G., Gimona, M., Hurley, J. H., Jarchau, T., Lehto, V. P., Lemmon, M. A., Linding, R., Mayer, B. J., Nagai, M., Sudol, M., Walter, U., and Winder, S. J. (2002) *FEBS Lett.* **513,** 141–144