Evolutionary Trace of G Protein-coupled Receptors Reveals Clusters of Residues That Determine Global and Class-specific Functions*⑤

Srinivasan Madabushi‡§, Alecia K. Gross¶‖, Anne Philippi§, Elaine C. Meng**, Theodore G. Wensel‡¶, and Olivier Lichtarge‡§‡‡§§¶¶

*From the ‡Program in Structural and Computational Biology and Molecular Biophysics, ‡‡Program in Developmental Biology, and §§Program in Cellular and Molecular Biology, ¶Department of Biochemistry and the §Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas 77030 and the **Department of Cellular and Molecular Pharmacology, University of California, San Francisco, California 94143-0450*

G protein-coupled receptor (GPCR) activation mediated by ligand-induced structural reorganization of its helices is poorly understood. To determine the universal elements of this conformational switch, we used evolutionary tracing (ET) to identify residue positions commonly important in diverse GPCRs. When mapped onto the rhodopsin structure, these trace residues cluster into a network of contacts from the retinal binding site to the G protein-coupling loops. Their roles in a generic transduction mechanism were verified by 211 of 239 published mutations that caused functional defects. When grouped according to the nature of the defects, these residues subdivided into three striking sub-clusters: a trigger region, where mutations mostly affect ligand binding, a coupling region near the cytoplasmic interface to the G protein, where mutations affect G protein activation, and a linking core in between where mutations cause constitutive activity and other defects. Differential ET analysis of the opsin family revealed an additional set of opsin-specific residues, several of which form part of the retinal binding pocket, and are known to cause functional defects upon mutation. To test the predictive power of ET, we introduced novel mutations in bovine rhodopsin at a globally important position, Leu-79, and at an opsin-specific position, Trp-175. Both were functionally critical, causing constitutive G protein activation of the mutants and rapid loss of regeneration after photobleaching. These results define in GPCRs a canonical signal transduction mechanism where ligand binding induces conformational changes propagated through adjacent trigger, linking core, and coupling regions.

The profusion and diversity of G protein-coupled receptors (GPCRs)[1] give them a central role in health and disease. In humans, over 1000 genes encode these receptors (1), each of which responds to a single or few ligands by activating G proteins, which then modulate enzymes and channels to initiate highly amplified signaling cascades. Such cascades control sight, taste, smell, slow neurotransmission and the responses to most water-soluble hormones and chemokines. In fact, GPCRs are so ubiquitous that, although they are the targets of nearly 50% of current drugs (2), this is still a small fraction of their pharmacological potential (3).

Some of the major questions relevant to GPCR pharmacology include the following: What residues are critical for ligand binding and G protein activation? What do different receptor families have in common with regard to their activation mechanism? From a structural perspective, it is known that all GPCRs form a seven transmembrane (TM) α-helical bundle, connected by three intracellular and three extracellular loops, with an extracellular N terminus and an intracellular C terminus (4, 5). However, low overall sequence identity of 25% even within class A GPCRs (6–8) suggests that significant deviations can occur in ligand binding pockets and in interhelical contacts that stabilize or mediate the transition between active and inactive conformations (9).

From an experimental perspective, while there is a wealth of data on a handful of GPCRs, most are known only from translated DNA sequences, hence the need for computational methods to extract functional information from those sequences. For example, correlated mutational analysis (10) and sequence-based entropy (11) have been used to detect networks of functional residues. Other studies focused on receptor topology (12), functional fingerprints (13), dimerization interfaces (14), and modeling of receptor-ligand binding (15–17).

This study aims to test two specific and complementary hypotheses. First, that all GPCRs share a common mechanism of activation and G protein coupling. This hypothesis is consistent with common and often promiscuous activation of G proteins by GPCRs, even after whole TMs are swapped to form chimeric receptors (18). It also implies the presence of a ubiquitous GPCR activation switch made up of residues that are functionally important in all receptors. Our second hypothesis predicts that residues involved in ligand specificity are different in different receptors. This would be consistent with the enormous diversity of ligand sizes and types, ranging from ions and small molecules to peptides and large glycoproteins, and suggests that some residues will be important in some but not all GPCRs.

The search for these functionally important residues is based on the evolutionary trace method (hereafter ET, or tracing), which identifies sequence positions where variations among related proteins always correlate with evolutionary divergences (19). Control studies (20–22), genuine predictions fol-
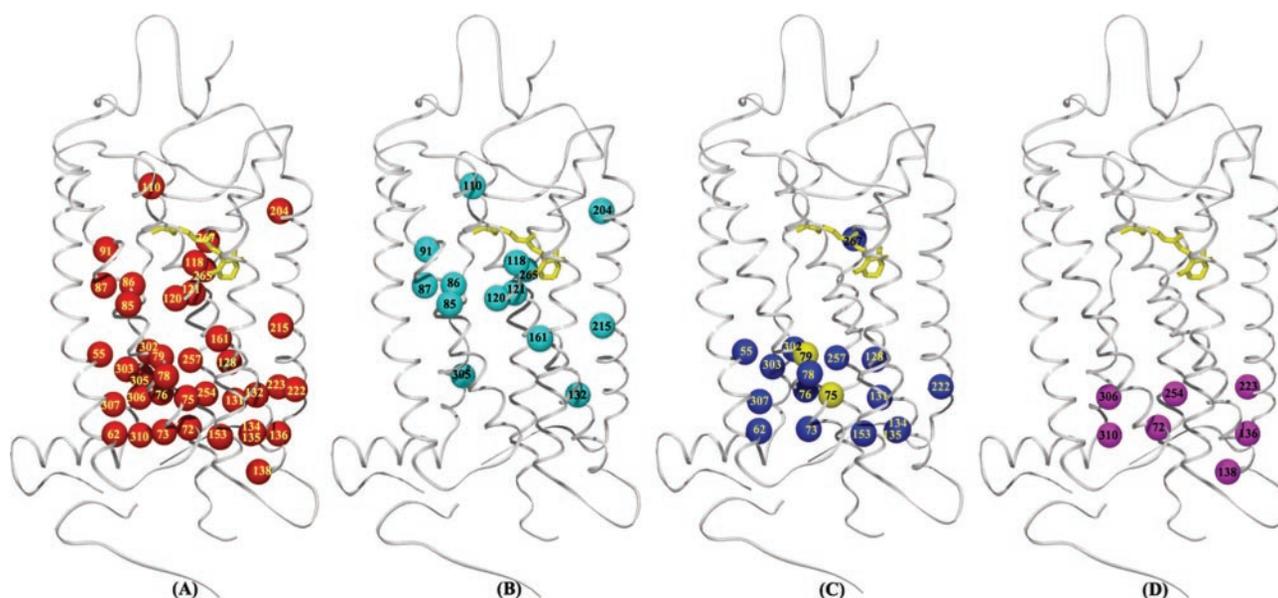
⑤ The on-line version of this article (available at http://www.jbc.org) contains Tables S1, S2 (*A* and *B*), and S3 and additional references.
[1] The abbreviations used are: GPCR, G protein-coupled receptor; TM, transmembrane; ET, evolutionary tracing.

FIG. 1. **Global trace residues identify a canonical signal transduction pathway with three functional subdomains: a ligand trigger region, an allosteric linking core and a G protein-coupling region.** *A* shows the top 20% of class A determinants (Cα atoms) mapped onto the rhodopsin structure (1HZX) with retinal depicted as a *yellow stick model*. *B* shows exclusively the subset that affects ligand binding (*cyan spheres*) on mutation, forming the *trigger region*. *C* shows in *blue* the residues that cause constitutive activity or folding/expression effects on mutation. They cluster to form an intermediate *linking core* involved in conformational activation linking the *trigger region* to a *coupling region* shown in *D* (*magenta spheres*) consisting of residues that affect G protein coupling/activation. Ile-75 and Leu-79, which had not previously been assigned any function, are depicted as *yellow spheres*.

lowed by mutagenesis studies (23), and large scale testing by us (24, 25) or others (26–28) show that trace residues cluster in the three-dimensional structure of proteins and that these clusters predict binding or catalytic sites. Moreover, differential ET analysis, which subtracts residues traced on many evolutionary branches of a protein family from those traced on a subset of these branches, can also identify positions that are important to that subset but not to the entire protein family (21).

In keeping with the hypotheses, we find that a trace of diverse receptors identifies a common functional site shared by all GPCRs and that differential ET identifies ligand-specific functional sites. The published mutational record confirms the former is a generic activation switch. Moreover, novel mutations at two trace residues reveal that they are linked to constitutive activity and protein stability demonstrating the predictive value of the evolutionary trace model to guide rational mutagenesis of GPCRs.

EXPERIMENTAL PROCEDURES

GPCR Family Alignments—BLAST (29) retrieved 129 visual opsins, 69 olfactory, and 82 chemokine class A receptor sequences from the NCBI data base (see additional information available at imgen.bcm.tmc.edu/molgen/labs/lichtarge/gpcr/). Each family was aligned with ClustalW (30) separately, and their TM segments were excised based on exclusion of charged residues and poorly conserved residues near the TM boundaries. Using the bovine rhodopsin numbering, segment 36–64 is TM1, 72–97 is TM2, 108–141 is TM3, 153–175 is TM4, 202–230 is TM5, 250–277 is TM6, and 286–311 is TM7. The segments were concatenated, and alignment was pruned to produce a gapless alignment of 330 GPCRs containing 195 TM residues.

Global and Specific Determinants—Generically important residues were obtained by tracing this alignment (20, 24, 25), ranking the relative importance of each position, and selecting those in the top 5th, 10th, 15th, and 20th percentile ranks. The rank of a given residue is the number of branches of the phylogenetic tree at which it becomes invariant in each branch, starting from the root branch as 1. To identify trace residues uniquely important to each family, we subtracted the generic ones from those that were important in ET analysis of that family, at the same percentile rank (see Supplementary Material).

Mutational Data—Functional data on trace residues were gathered from the tGRAP mutant data base (32), the protein mutant database

(33), and the literature. For global trace residues we considered mutations at cognate residues across all class A receptors, but for opsin-specific trace residues we used opsin mutations only. A ligand binding effect means that binding assays showed altered affinity or specificity. A G protein-coupling effect signifies altered GTP binding or in inositol phosphate production under appropriate controls. Constitutive activity means signaling activity (inositol phosphate production/cAMP accumulation/G protein activation) in a non-activated (ligand-free) receptor. Folding or expression effects mean considerably less mutant receptor expression under identical conditions as wild-type. Subtle changes were not taken into account, and given multiple functional effects a residue was classified by the one most frequently observed (Fig. 1 and Table II). Two-tailed χ^2 tests were performed as described in a previous study (34).

DNA Mutagenesis and Expression of Opsins—All opsin variants were expressed using modified forms of a plasmid containing a synthetic opsin gene cDNA in a pMT3-based vector (35, 36). Mutations were introduced using the QuickChange® mutagenesis kit (Stratagene). Expression in COS-1 cells, membrane isolation, reconstitution with 11-*cis*-retinal, and protein purification using 1D4 antibody affinity chromatography were carried out as described previously (36–40). Absorbance spectrophotometry was carried out using an Olis/SLM-Aminco DW-2000 dual-beam instrument adapted for darkroom use. Photolysis of rhodopsin was carried out using a continuous wave argon ion laser directed into the sample compartment from above by a concave mirror that dispersed the beam evenly over the cuvette, with beam gating carried out with a manual shutter. Complete photoisomerization of rhodopsin and all mutants was achieved without significant heating in less than 1 s. Transducin activation assays were carried out as described previously (39).

RESULTS

Global Determinants—To find residues mediating a generic signal transduction mechanism, we traced jointly four evolutionarily distant families: the visual rhodopsin, bioamine, olfactory, and chemokine receptors. The 39 residues ranked in the top 20th percentile (Fig. 1 and Tables I and II) are predicted to be generically important. When they are mapped onto the rhodopsin structure, they form a three-dimensional structural cluster that is internal, mostly located in the cytoplasmic half of the membrane, and statistically significant ($p = 0.0002$, which is the probability that the same number of randomly selected residues would cluster as tightly in three dimensions

TABLE I
The top 20% of trace residues in global and opsin family analyses

Residues specific to the opsin family are *underlined*. Refer to the Supplementary Material for percentile ranks and functional classification of global trace residues. The top 20% of trace residues from global (class A) and opsin family analyses are shown.

TM	Global	Opsin
1	Asn-55, Thr-62	Asn-55, Gly-51, Thr-58
2	Leu-72, Asn-73, Ile-75, Leu-76, Asn-78, Leu-79, Phe-85, Met-86, Val-87, Phe-91	Leu-72, Asn-73, Leu-76, Asn-78, Gly-90, Phe-91
3	Cys-110, Thr-118, Gly-120, Gly-121, Leu-128, Leu-131, Ala-132, Glu-134, Arg-135, Tyr-136, Val-138	Cys-110, Glu-113, Gly-121, Leu-125, Leu-128, Leu-131, Glu-134, Arg-135
4	Ala-153, Trp-161	Ala-153, Gly-156, Trp-161, Cys-167, Pro-171, Trp-175
5	Val-204, Pro-215, Cys-222, Tyr-223	Met-207, Pro-215, Cys-222, Tyr-223, Val-230
6	Val-254, Met-257, Trp-265, Pro-267	Val-250, Met-253, Val-254, Pro-267
7	Asn-302, Pro-303, Ile-305, Tyr-306, Ile-307, Asn-310	Met-288, Phe-294, Ala-295, Lys-296, Asn-302, Pro-303, Tyr-306, Ile-307, Asn-310

(see Ref. 24)). The cluster includes critical GPCR motifs and residues, such as DRY motif in TM3, NP*XX*Y motif in TM7, Cys-110 in TM3 (disulfide-bonded), and Asn-55 in TM1. Most of its trace residues belong to TM2, TM3, TM7, and to a smaller extent TM6, consistent with mutagenesis, spectroscopy, and cross-linking evidence that the latter three helices undergo rigid-body motions upon activation (41–45).

To assess the functional importance of these 39 trace residues, we reviewed the published record of their mutations (Table II) and found 239 mutations (excluding structural studies) distributed among 37 trace residues. 211 (88%) of these mutations gave rise to functional differences, including at least one at each of the 37 trace positions. Residues mutated in 2 receptors affected both in 6 of 7 (86%) cases, those mutated in 3 receptors affected all three 7 out of 8 (87%) times and at least 2 in all 8 (100%) instances. Nineteen positions were mutated in 4 or more receptors. In all 19 cases function was disrupted in at least 3 different receptor types, and in 18 cases (95%) function was affected in 4 or more receptor types. This confirms the generic role of these residues in GPCR signaling.

To correlate the structural location of these trace residues with their function, we further sub-classified them according to their most frequently observed mutational effect and mapped this information onto the rhodopsin structure (see Fig. 1 and Supplementary Material). Fourteen residues, colored *cyan* in Fig. 1*B*, predominantly affect ligand binding (Table II) and segregate in the extracellular half of the cluster. Sixteen residues, colored *blue* in Fig. 1*C*, cause constitutive activity, folding, or expression defects, and these segregate roughly in the middle of the cluster (Table II). Seven residues, colored *magenta* in Fig. 1*D*, primarily affect G protein coupling and signaling, and those fill the cytoplasmic base of the transmembrane domain. Thus specific functional effects localize strikingly in three regions: a trigger region, composed of residues most immediately involved in ligand-induced conformational changes, a coupling region adjacent to the cytoplasmic loops where G proteins bind, and between them a linking core, composed of residues that stabilize TM interactions and folding. Two trace residues in this linking core, Ile-75 and Leu-79, have not been previously mutated, to our knowledge, and we mutated Leu-79 to test the functional importance of our prediction.

As a negative control, we also analyzed the mutational data on the bottom 20% of evolutionarily important positions (39 residues). 3 positions had no mutational data, 14 had mutations that had no effect in any receptors, and only 3 residues had mutations that affected a generic GPCR function, namely expression (1 case) and G protein interaction (2 cases). Mutations of the remaining 19 residues affected ligand interaction, which is a ligand-specific function. In terms of mutations, 60 out of 112 (54%) reported mutations that have no apparent

effect on function as opposed to only 12% at trace residues (Table III and Supplementary Material). This greater than 4-fold difference is significant with a *p* value of less than 0.0005 on a χ^2 test. Of the 52 mutations at low ranked positions that have functional effects, 27 mutations (52%) are in the top 20% of family-specific residues identified by differential trace analysis (data not shown). Finally, although 27 of 37 (73%) top ranked trace residues affect function in at least 3 receptors, only 7 of 36 (19%) low ranked positions do so (see Supplementary Material). Thus top ranked trace residues affect generic functions, and they do so in multiple receptor families, whereas bottom ranked ones often have no effect, and when they do this is mostly limited to ligand interactions in one or two receptors.

Retinal Binding Site in Opsin Family—We asked next whether differential trace analysis (21) could highlight functional determinants unique to the visual rhodopsin family. We traced 129 rhodopsin sequences and subtracted from the resulting trace residues the global ET residues (Fig. 1) at the same percentile rank (Fig. 2). At the top 20th percentile rank, 17 opsin-specific trace residues were found, of which 11 cluster near the retinal binding pocket (*p* value < 0.001) (Table I and Supplementary Material). The other opsin-specific residues appear at the cytoplasmic edge or near it. A retrospective analysis of the literature, summarized in Table IV, shows 32 known opsin family mutations at 12 opsin-specific trace residues. Of these, 28 mutations (87%) in 12 trace residues are linked to spectral shifts, constitutive activity, night blindness, or retinitis pigmentosa. Interestingly, Trp-175 is the single apparent false positive. This is surprising given that it is conserved across all opsin sequences, yet past mutations to other aromatic residues caused no spectral shifts and no loss of G protein activation (46). We therefore focused on this amino acid, Trp-175, for further study.

Trace Residue Mutations in Rhodopsin—As a *bona fide* test of the ability of ET to guide function-altering mutations in GPCRs, we prepared mutants at one globally important residue Leu-79 (L79A and L79S) and at one opsin-specific residue Trp-175 (W175A, W175C, and W175H). All expressed at sufficient levels to allow isolation by immunoaffinity chromatography in detergent following reconstitution with 11-*cis*-retinal. The absorbance spectra in the rhodopsin form (with 11-*cis*-retinal bound) were essentially identical to wild-type (data not shown). All displayed wild-type ability to activate transducin upon exposure to light (Fig. 3).

There were, however, functional differences between the mutant proteins and wild-type opsin. As shown in Fig. 4, substitutions at both positions (W175A, W175C, W175H, and L79A) led to constitutive activity, *i.e.* G protein activation in the absence of ligand. Unlike wild-type opsin (47), none of the mutant opsins was sufficiently stable for isolation in detergent without added 11-*cis*-retinal, except L79S, which was obtained

TABLE II
Mutational effects of globally important (cognate) residues in different receptor families

Mutations of cognate trace residues (bovine rhodopsin numbering) in various receptors obtained from t-GRAP mutant database/alignments and protein mutant database and in some cases from literature search. 211 out of 239 (88%) mutations in 37 residues had functional effects. Ile-75 and Leu-79 had no mutational data. + refers to mutation causing a functional change; − indicates mutation that had no recorded effect. The numbers of receptor families in which trace residue position was mutated are listed. Refer to Supplementary Material for a detailed version of this table.

Residue (bovine rhodopsin)	No. of mutants +	No. of mutants −	No. of receptor families	Dominant mutational effect[a]
55	3		3	Mixed (CA, LB, Exp.)
62	2		2	Mixed/ G protein
72	1		1	G protein effects
73	3		2	Expression
75				
76	8		5	Mixed (CA, [↑ cAMP])
78	2	1	2	Mixed (G-protein, CA)
79				
85	3		3	Ligand binding
86	6	1	4	Ligand binding
87	5	2	6	Ligand binding
91	2		2	Ligand binding
110	15		15	Ligand binding
118	7	2	9	Ligand binding
120	11		9	Ligand binding
121	5	4	8	Ligand binding
128	1		1	CA
131	1	1	2	Mixed (CA, LB)
132	3	2	5	Ligand binding
134	19		9	CA
135	12		8	Mixed (CA, ME, G uncoupling)
136	6	1	4	G protein effects
138	3		3	G protein effects
153	2		2	Mixed (LB, expression)
161	10	2	10	Ligand binding
204	7	3	8	Ligand binding
215	3		3	Ligand binding
222	2		2	CA
223	3		3	G protein effects
254	10		4	G protein effects
257	3		3	CA
265	9	2	7	Ligand binding
267	6		4	Mixed (CA, folding)
302	8	2	10	Mixed (folding/expression, LB)
303	11		6	Folding/expression
305	5		3	Ligand binding
306	10	4	10	G protein effects
307	2	1	3	Mixed (LB, ME)
310	2		1	G protein effects

[a] CA, constitutive activity; LB, ligand binding effects; ME, mechanistic effects observed in structural studies (does not count as functional effect); Exp, decreased expression; G protein, G-protein coupling decreased/completely uncoupled/decreased G-protein stimulation.

in very low yield (data not shown).

To determine the ability of the mutants to regenerate the dark photoreceptive state following photoactivation, we used a laser flash to effect nearly instantaneous quantitative photoisomerization of bound 11-*cis*-retinal to the all-*trans* form in the presence of excess 11-*cis*-retinal and then measured the maximum extent of regeneration by absorbance at 500 nm (Fig. 4). Nearly 80% of wild-type rhodopsin was regenerated under these conditions, whereas the Leu-79 mutants ranged from 40% to 45%, and the Trp-175 mutants ranged from 8% to 14% maximal regeneration.

Thus both trace residues are functionally important. They contribute to keeping activity levels low in the absence of ligand. They also appear to have a role in protein stability as indicated by defects in regeneration of the functional dark state following photoactivation in detergent. Moreover, in this example, evolutionary trace rank is related to the severity of the

TABLE III
Distribution of mutations for the top and bottom 20% of ET residues

Mutational effectiveness at a trace *versus* non-trace residue is significantly different with a χ^2 test p value of 0.0005 or better. Refer to Supplementary Material for related tables. Number of mutations for top and bottom 20% of ET residues are shown.

Fraction of ET residues	No. of mutations with recorded effect	No. of mutations with no recorded effect
Top 20%	211	28
Bottom 20%	52	60

defect. Mutations at Leu-79, which ranks in the 20th percentile for the larger set of Class A sequences but only in the 30th percentile for the opsin family, had a less severe effect on rhodopsin function than mutations at position Trp-175, which ranks in the 5th percentile of importance for the opsin family. These results confirm that, as with soluble protein families examined previously, ET can be used to direct mutagenesis experiments efficiently to residues of functional importance in transmembrane domains of GPCR and likely of other transmembrane proteins as well. In practice, it is interesting that we found two new sites at which mutations induce constitutive activity. Previously, despite very extensive mutagenesis of rhodopsin, only seven such positions had been identified. This result points to the ability of the evolutionary trace as a method for finding novel sites of interest even in previously intensively studied proteins and protein families.

DISCUSSION

The amino acid determinants of GPCR function should fall in two categories: those that mediate ligand binding, expected to be highly specific because of the chemical and structural diversity of ligands and sequence divergence among receptors, and those that mediate the subsequent conformational change and receptor activation of G proteins, expected to be similar across receptors because of the promiscuity of G protein coupling (48), the activity of chimeric constructs (18), and the evidence for similar conformational changes (20, 28–30). Both parts of this hypothesis are validated by this study.

Global Trace Residues Reveal a Canonical GPCR Conformational Switch—First, a trace of four GPCR families with low sequence identity extracted common functional determinants. The structural clustering of these trace residues, and their contact network that is especially dense at the cytoplasmic end near the G protein-coupling interface, suggest that they participate in a canonical GPCR conformational switch. This notion is supported by existing mutations that show, first, that ET is highly specific, because mutations at every top ranked residue affect function in at least one receptor and only 12% of mutations at these residues have no effect, and second that trace residues are critical to many receptors, because mutations of 73% of them affect at least three receptors. In addition, correlation of functional effects of mutations with location revealed that the switch involves three functionally distinct but structurally connected sub-clusters: a *trigger region* (Fig. 1*B*), near the retinal binding pocket of rhodopsin, a *coupling region* (Fig. 1*D*) near the G protein-coupling site, and a *linking core* (Fig. 1*C*) between the first two.

The seven coupling region residues fill the cytoplasmic base of the transmembrane domain, presumably forming a platform for G protein interactions, and six residues border the key G protein coupling loops 2 and 3 (Fig. 1*D*). Mutations at Leu-72 and Val-254 decrease G protein activation in rhodopsin, vasoactive intestinal peptide receptor, and muscarinic receptors. Others in this region affect G protein coupling in rhodopsin, angiotensin-2, thyrotropin, and α,β-adrenergic receptors (Table II). Residue Tyr-136 of the (E/D)RY motif is especially
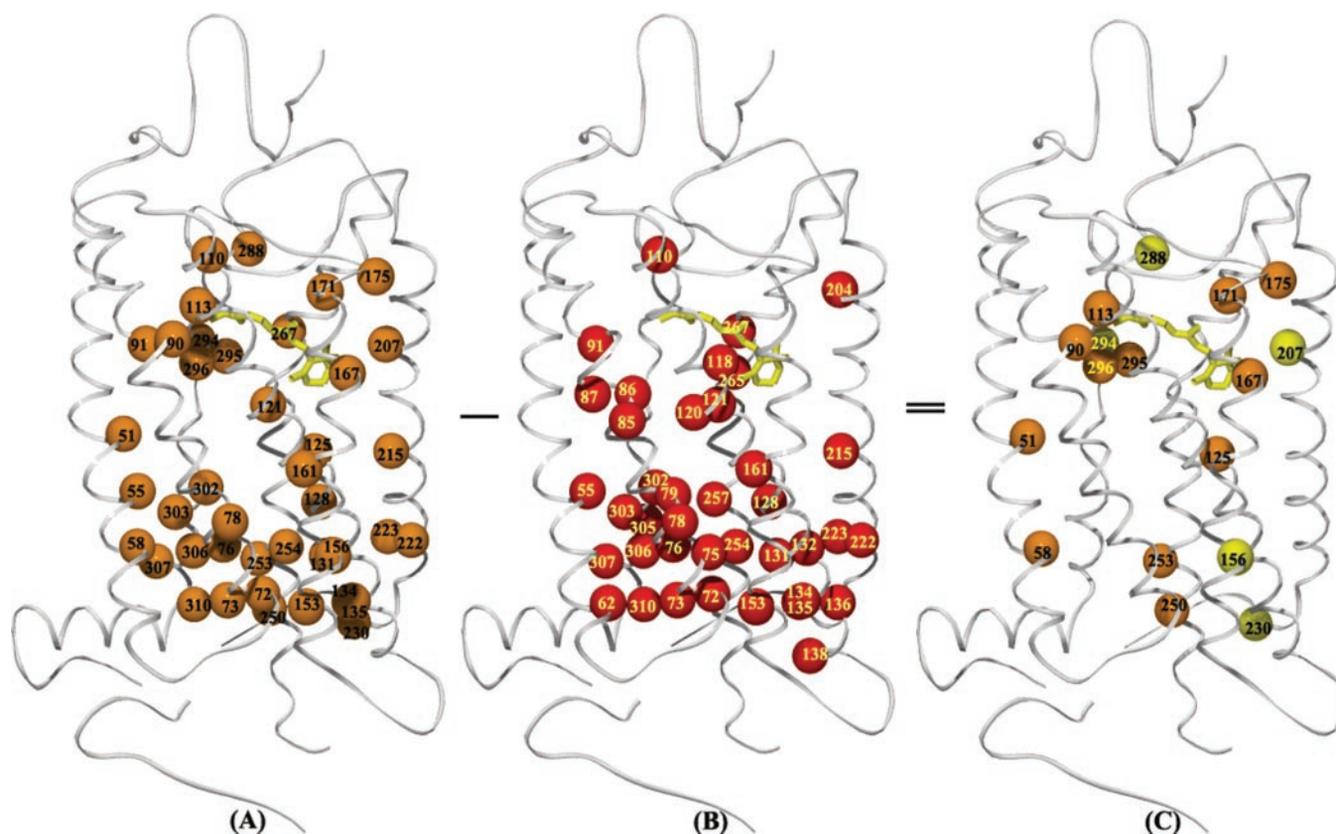
FIG. 2. **Differential analysis identifies retinal binding site.** The top 20% of trace residues (Cα atoms) are mapped onto the rhodopsin structure (1HZX) with retinal depicted as *yellow sticks*. *A* shows a trace of the rhodopsin family with trace residues in *orange*. *B* shows the trace of class A sequences from Fig. 1*A*. *C* shows opsin family-specific residues obtained by subtracting *B* from *A*. They mostly cluster around the retinal binding site with a few residues found near the cytoplasmic half of the receptor. Met-207, Met-288, Phe-294, Gly-156, and Val-230 have no mutational data and are depicted as *yellow spheres*.

TABLE IV
Mutational effects of opsin family specific trace residues

Recorded functional effects of mutations in opsins are listed. Gly-156, Met-207, Val-230, Met-288, and Phe-294 had no available mutational data. Overall, 28 of 32 (87%) mutations had functional effects. Refer to Supplementary Material for references of mutations listed. Mutational effects of opsin family specific trace residues are shown.

| TM	Residue (bovine rhodopsin no.)	Mutational effect
1	51	ADRP[a]
	58	ADRP, ↓ transducin activity, expression
2	90	Congenital stationary night blindness, blue shift, CA
3	113	Blue shift, CA, spectral shift
	125	ADRP, ↓ retinal binding, misfolding, expression, no spectral activity ↓ transducin activation, ↓ time in M-II
4	156	
	167	ADRP, low expression levels, misfolding
	171	ADRP, low expression levels, misfolding, localization
	175	Normal spectrum and G protein activation
5	207	
5	230	
6	250	↓ G-protein coupling, switched G-protein specificity
	253	Spin labeling: rhodopsin
7	288	
	294	
	295	Spectral shift
	296	CA, ↓ expression levels, arrestin and rhodopsin kinase mediated effects affected

[a] ADRP, autosomal dominant retinitis pigmentosa.

important in triggering GDP exchange in G proteins (49, 50).

The generic trigger region extends into the transmembrane area nearly up to the retinal binding pocket in rhodopsin (Fig. 1*B*), suggesting an intimate role in ligand sensing, which is supported by the mutagenesis data (Table II). The few residues far from the retinal binding pocket that affect ligand binding are likely coupled allosterically.

Most trace residues, however, lie in the intermediate linking core (Fig. 1*C*) drawn heavily from residues forming key hydro-

gen bond networks such as TM1-TM2-TM7 (critical role for Asn-55), TM2-TM3-TM4 (Asn-78), and TM6-TM7 (Met-257, Asn-302, and Pro-303) (50–52). Past mutations at these positions cause constitutive activity, misfolding, or expression defects and demonstrate a key role of this residue in maintaining structural integrity and dynamics. This is consistent with this linking core being an allosteric pathway coupling ligand binding to G protein activation by means of a conformational change initiated at the trigger region and which extends to the
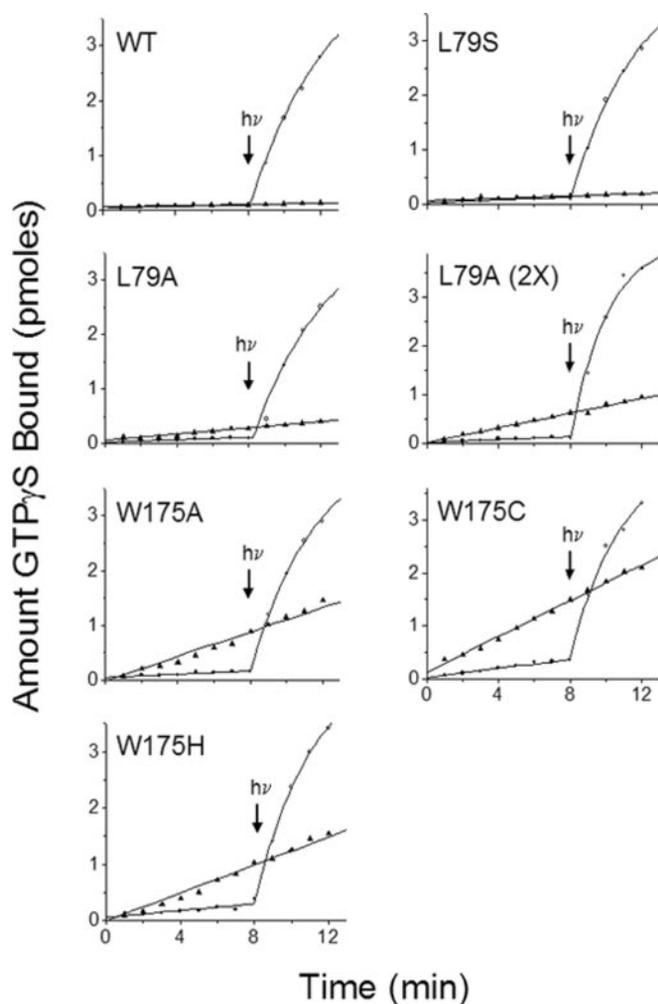
FIG. 3. **Constitutive activation of transducin by mutant opsins.** Transducin activation by wild-type (*WT*) and mutant opsins is shown in the presence and absence of 11-*cis*-retinal and light. Transducin activity was assayed using membranes from transfected COS cells as previously described (39): ▲, in the absence of retinal; ●, in the presence of retinal; and ○, time course for the reaction in the presence of retinal after exposure to light (*hν*) for each mutant listed. Membrane amounts were selected to give the same light-dependent transducin activation kinetics; experiments with detergent-purified rhodopsins confirmed that the specific activities in the light were not measurably different for the mutants and wild-type (data not shown). For *L79A (2X)*, twice the amount of membranes was assayed.
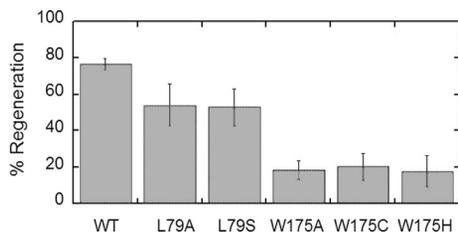
FIG. 4. **Efficiencies of rhodopsin regeneration following photolysis.** The 500-nm absorbance values of purified rhodopsins were measured in the dark. Subsequently, in the presence of a 2-fold molar excess of 11-*cis*-retinal, the samples were exposed to a 1-s flash of 514-nm laser light at intensity sufficient to effect complete photoisomerization. The increase in 500-nm absorbance due to regeneration was recorded over 35 min to determine the maximal percent regeneration for each protein. The values represent the averages for each mutant repeated three times, and *error bars* indicate the standard deviations.

coupling region (Table II). There is an overlap between all three of these regions, and these roles are also supported by our findings that mutations at Leu-79 caused constitutive activity

(L79A) and decreased stability (both L79A and L79S).

ET Residues Cluster at Functional Sites in Opsins—Through differential analysis we also identified rhodopsin-specific trace residues, and they were indeed directly related to light sensing. Most cluster around the retinal binding site (Fig. 2*C*), and many influence interactions with the chromophore (Table IV). For example, Lys-296 is linked to retinal via a Schiff base, and introduction of a negative charge at Gly-90 (G90D) provides an alternative counterion to the positive charge of the protonated Schiff base and displaces the salt bridge between Lys-296 and Glu-113. Mutants E113Q and G90D cause a blue shift in the retinal absorption spectrum and constitutive activation of transducin, whereas mutations at Leu-125, Cys-167, and Pro-171 cause retinitis pigmentosa, poor retinal binding, misfolding, and reduced expression. Rhodopsin-specific residues in the cytoplasmic half of rhodopsin appear to affect G protein activation/coupling and TM-TM interactions (Fig. 2*C*), as revealed by mutations at Gly-51, Thr-58, and Val-250 (refer to Supplementary Material). Note that a family-specific residue could be important in some other families as well but is distinguished by not being important in all families.

We found no mutational data for Met-207, Met-288, and Phe-294 (near retinal) or for Gly-156 and Val-230 (near the cytoplasmic side). Their proximity to the retinal or to other key residues suggests that they play a role in light sensing. For example, Phe-294 is within ~8 Å of retinal and adjacent to Lys-296, whereas Met-207 and Met-288 are also within 4 Å and 7.5 Å, respectively, of retinal.

Trp-175 is the one trace residue for which mutational data refute functional importance; mutation of this invariant residue to phenylalanine or tyrosine did not give rise to a spectral shift or a defect in light-triggered G protein activation (46). We hypothesized that the functional importance of this residue could be revealed by additional mutations or assays. Indeed, mutations W175A, W175C, and W175H all have severe functional effects *in vitro*, although all display normal absorbance spectra and light-triggered G protein activation. These results suggest that ET can not only be used to identify novel sites where mutations are likely to provide insight into functional importance (*e.g.* L79A), but also to suggest positions at which apparent lack of functional effect by mutations may need to be investigated in more detail (*e.g.* Trp-175).

Applied Phylogenomics—The present evolutionary study of GPCRs suggests some answers to the questions raised at the outset about the mechanisms for signal transduction and ligand sensing. First, distantly related Class A receptors share common determinants for ligand sensing, G protein coupling and allosteric relay/control of activation state. In our experience, whenever divergent members of a protein family share such extensive functional sites, their structures are also highly related. This mutual consistency of sequence, structure, and function during evolution is pervasive (24, 25, 31), and here it strongly suggests that GPCR transmembrane structures are similar at least in the vicinity of the G protein interface.

In practice, this study provides a systematic and rational guide for engineering GPCR mutations. First, as shown here in rhodopsin, differential ET opens the door for a strategy to identify and test likely ligand binding sites in other receptor families. Second, studies of the conformational switch targeted to the trigger, linking, and coupling regions may help control receptor sensitivity, constitutive activity, and coupling specificity, which is of broad pharmaceutical interest. More generally, ET analyses should similarly inform studies in other large families of membrane proteins, such as ion channels and ABC transporters. In summary, our results suggest that ET is a reliable and highly efficient method for

identifying multiple functional sites and residues for further experimental study.

REFERENCES

1. Nestler, E. J., and Landsman, D. (2001) *Nature* **409,** 834–835
2. Drews, J. (1996) *Nat. Biotechnol.* **14,** 1516–1518
3. Flower, D. R. (1999) *Biochim. Biophys. Acta* **1422,** 207–234
4. Baldwin, J. M. (1994) *Curr. Opin. Cell Biol.* **6,** 180–190
5. Kolakowski, L. F. J. (1994) *Receptors Channels* **2,** 1–7
6. 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., Miyano, M. (2000) *Science* **289,** 739–745
7. Horn, F., Bywater, R., Krause, G., Kuipers, W., Oliveira, L., Paiva, A. C., Sander, C., and Vriend, G. (1998) *Receptors Channels* **5,** 305–314
8. Donnelly, D. (1997) *FEBS Lett.* **409,** 431–436
9. Chothia, C., and Lesk, A. M. (1986) *EMBO J.* **5,** 823–826
10. Oliveira, L., Paiva, A. C., and Vriend, G. (2002) *Chembiochem* **3,** 1010–1017
11. Suel, G. M., Lockless, S. W., Wall, M. A., and Ranganathan, R. (2003) *Nat. Struct. Biol.* **10,** 59–69
12. Baldwin, J. M. (1993) *EMBO J.* **12,** 1693–1703
13. Attwood, T. K., and Findlay, J. B. (1993) *Protein Eng.* **6,** 167–176
14. Gouldson, P. R., Higgs, C., Smith, R. E., Dean, M. K., Gkoutos, G. V., and Reynolds, C. A. (2000) *Neuropsychopharmacology* **23,** S60–S77
15. Vaidehi, N., Floriano, W. B., Trabanino, R., Hall, S. E., Freddolino, P., Choi, E. J., Zamanakos, G., and Goddard, W. A., 3rd (2002) *Proc. Natl. Acad. Sci. U. S. A.* **99,** 12622–12627
16. Bissantz, C., Bernard, P., Hibert, M., and Rognan, D. (2003) *Proteins* **50,** 5–25
17. Attwood, T. K., Croning, M. D., and Gaulton, A. (2002) *Protein Eng.* **15,** 7–12
18. Kudo, M., Osuga, Y., Kobilka, B. K., and Hsueh, A. J. (1996) *J. Biol. Chem.* **271,** 22470–22478
19. Lichtarge, O., Bourne, H. R., and Cohen, F. E. (1996) *J. Mol. Biol.* **257,** 342–358
20. Lichtarge, O., Bourne, H. R., and Cohen, F. E. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93,** 7507–7511
21. Lichtarge, O., Yamamoto, K. R., and Cohen, F. E.(1997) *J. Mol. Biol.* **274,** 325–337
22. Sowa, M. E., He, W., Wensel, T. G., and Lichtarge, O. (2000) *Proc. Natl. Acad. Sci. U. S. A.* **97,** 1483–1488
23. Sowa, M. E., He, W., Slep, K. C., Kercher, M. A., Lichtarge, O., and Wensel, T. G. (2001) *Nat. Struct. Biol.* **8,** 234–237
24. Madabushi, S., Yao, H., Marsh, M., Kristensen, D. M., Philippi, A., Sowa, M. E., and Lichtarge, O. (2002) *J. Mol. Biol.* **316,** 139–154
25. Yao, H., Kristensen. D. M., Mihalek, I., Sowa, M. E., Shaw, C., Kimmel, M., Kavraki, L., and Lichtarge, O. (2002) *J. Mol. Biol.* **326,** 255–261
26. Pritchard, L., and Dufton, M. J. (1999) *J. Mol. Biol.* **285,** 1589–1607
27. Landgraf, R., Fischer, D., and Eisenberg, D. (1999) *Protein Eng.* **12,** 943–951
28. Innis, C. A., Shi, J., and Blundell, T. L. (2000) *Protein Eng.* **13,** 839–847
29. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) *J. Mol. Biol.* **215,** 403–410
30. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) *Nucleic Acids Res.* **22,** 4673–4680
31. Mihalek, I., Yao, H., Res, I., and Lichtarge, O. (2003) *J. Mol. Biol.* **331,** 263–279
32. Kristiansen, K., Dahl, S. G., and Edvardsen, O. (1996) *Proteins* **26,** 81–94
33. Nishikawa, K., Ishino, S., Takenaka, H., Norioka, N., Hirai, T., Yao, T., and Seto, Y. (1994) *Protein Eng.* **7,** 773
34. Brown, B. W., and Hollander, M. (1977) *Statistics: A Biomedical Introduction, Wiley Series in Probability and Mathematical Statistics*, pp. 196–201, John Wiley & Sons, New York
35. Ferretti, L., Karnik, S. S., Khorana, H. G., Nassal, M., and Oprian, D. D. (1986) *Proc. Natl. Acad. Sci. U. S. A.* **83,** 599–603
36. Franke, R. R., Sakmar, T. P., Oprian, D. D., and Khorana, H. G. (1988) *J. Biol. Chem.* **263,** 2119–2122
37. Oprian, D. D., Molday, R. S., Kaufman, R. J., and Khorana, H. G. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84,** 8874–8878
38. Robinson, P. R., Buczylko, J., Ohguro, H., and Palczewski, K. (1994) *Proc. Natl. Acad. Sci. U. S. A.* **91,** 5411–5415
39. Gross, A. K., Rao, V. R., and Oprian, D. D. (2003) *Biochemistry* **42,** 2009–2015
40. Zhukovsky, E. A., Robinson, P. R., and Oprian, D. D. (1991) *Science* **251,** 558–560
41. Sheikh, S. P., Zvyaga, T. A., Lichtarge, O., Sakmar, T. P., and Bourne, H. R. (1996) *Nature* **383,** 347–350
42. Sheikh, S. P., Vilardarga, J. P., Baranski, T. J., Lichtarge, O., Iiri, T., Meng, E. C., Nissenson, R. A., and Bourne, H. R. (1999) *J. Biol. Chem.* **274,** 17033–17041
43. Farrens, D. L., Altenbach, C., Yang, K., Hubbell, W. L., and Khorana, H. G. (1996) *Science* **274,** 768–770
44. Gether, U., Ballesteros, J. A., Seifert, R., Sanders-Bush, E., Weinstein, H., and Kobilka, B. K. (1997) *J. Biol. Chem.* **272,** 2587–2590
45. Fanelli, F., Barbier, P., Zanchetta, D., De Benedetti, P. G., and Chini, C. (1999) *Mol. Pharm.* **56,** 214–225
46. Lin, S. W., and Sakmar, T. P. (1996) *Biochemistry* **35,** 11149–11159
47. Gross, A. K., Xie, G., and Oprian, D. D. (2003) *Biochemistry* **42,** 2002–2008
48. Neves, S. R., Ram, P. T., and Iyengar, R. (2002) *Science* **296,** 1636–1639
49. Acharya, S., and Karnik, S. S. (1996) *J. Biol. Chem.* **271,** 25406–25411
50. Sakmar, T. P., Menon, S. T., Marin, E. P., and Awad, E. S. (2002) *Annu. Rev. Biophys. Biomol. Struct.* **31,** 443–484
51. Le Gouill, C., Parent, J. L., Caron, C. A., Gaudreau, R., Volkov, L., Rola-Pleszczynski, M., and Stankova, J. (1999) *J. Biol. Chem.* **274,** 12548–12554
52. Han, M., Smith, S. O., and Sakmar, T. P. (1998) *Biochemistry* **37,** 8253–8261