Genome Dynamics, Evolution, and Protein Modeling in the Olfactory Receptor Gene Superfamily[a]

DROR SHARON, GUSTAVO GLUSMAN, YITZHAK PILPEL, SHIRLEY HORN-SABAN, AND DORON LANCET[b]

Department of Molecular Genetics, The Weizmann Institute of Science, Rehovot 76100, Israel

ABSTRACT: The human olfactory subgenome represents several hundred olfactory receptor (OR) genes in a dozen or more clusters on several chromosomes. One OR gene cluster on human chromosome 17 has been characterized by us in detail. Based on a large-scale DNA sequence analysis, we have identified events of gene duplication and fusion as well as the generation of pseudogenes. The latter instances of 'gene death' could underlie the widespread phenomenon of human specific anosmias. Sixteen OR coding regions were found on this cluster, and six of them are pseudogenes. One of these pseudogenes, OR17-23, was found to be an intact open reading frame in an old world monkey. This may be a reflection of an OR repertoire diminution in man. A homology model of the OR protein was constructed by utilizing the rich information available on ~ 200 OR sequences. The putative odorant complementarity determining regions (CDR) was found to consist of 20 hypervariable residues facing an interior caving defined by transmembrane helices 3, 4 and 5. Such a model could be useful in analyzing additional OR gene sequences in the human genome in terms of odorant binding.

INTRODUCTION

Multigene families are divided into two main types: variant and invariant gene families.[1] Invariant multigene families constitute multiple identical copies of the same gene, and in some instances may be correlated with the need to synthesize increased amounts of a gene product. Representative examples include the genes for ribosomal RNA[2] and histones.[3] Variant multigene families consist of genes that are similar to each other, but differ in their sequence to some extent. A representative example is the immunoglobulin gene family.[4] In both types of families, gene conversion and unequal crossing-over are thought to be the two most important mechanisms that account for concerted evolution.[5,6]

The main force that accounts for gene family expansion is gene duplication, followed by a concerted evolution mechanism in which selection acts on each member of the gene family with regard to the other family members. Acceleration of nonsynonymous substitution is often found to be related to gene duplication, which causes one of the copies to diverge gradually and become a new functional gene or a functionless pseudogene.[7]

Olfactory receptor (OR) genes are an outstandingly variant multigene family, with 100–1000 genes per genome. While most biological receptors evolved to perform a specific recognition function, the OR gene product repertoire interacts with practically all

[a] Supported in part by a grant from the National Institutes of Health (NIDCD 00305).

[b] Corresponding author. Tel: (972)-8-9343683; fax: (972)-8-9344112; email: bmlancet@weizmann.weizmann.ac.il

volatile chemicals, constituting millions different odorous chemicals. The currently known collection of OR genes can be grouped into more than 20 gene families, whose members share 25–97 percent of amino acid identity, thus formally constituting a gene superfamily.[8] OR proteins are seven-transmembrane domain, G protein-coupled receptors (GPCRs),[9–11] 304–333 amino acids long. ORs are expressed mainly in the olfactory epithelium, where each cell expresses only one OR gene[12] and even just one allele at a given locus.[13] This expression pattern is believed to provide the molecular basis of odor sensitivity and discrimination.[14,11]

The mammalian genomes contain several hundred OR genes, which are organized in several clusters.[15,16] The collection of all genomic regions that code for ORs has been named the olfactory subgenome.[17] In the human genome, several OR clusters have been studied: in addition to the presently described cluster on chromosome 17,[15,18] OR clusters have been reported on chromosome 3,[19] chromosome 6,[20] chromosome 11,[21] chromosome 14 (L. Hood, personal communication), and chromosome 19[22] (Olsen, A., D. Lancet & S. Horn-Saban, in preparation).

The rapid expansion of the number of known OR sequences, and accumulation of structural knowledge in many other GPCRs, has made it now possible to employ novel bioinformatics tools to the problem of OR structural modeling. This has led to our accurate identification of the variable site of ORs which may constitute the odorant complementarity determining regions (CDR), analogous to the antigen binding site in immunoglobulins.[23]

THE OR CLUSTER ON HUMAN CHROMOSOME 17

The OR cluster on human chromosome 17 (17p13)[15] includes at least 16 OR genes. Previously, a detailed study was reported by us, which included the sequence analysis of cosmid 39, located in the center of the cluster.[18] This analysis revealed two OR genes (OR17-40 and OR17-228) and two OR pseudogenes (OR17-24 and OR17-25), all belonging to the same subfamily (3A). The two genes were found to have evolved by ancient tandem duplication, about 100 million years old, of an 11-kb-long fragment, potentially mediated by recombination between mammalian-wide interspersed repeats (MIRs). Sequence analysis of the large duplication revealed evidence for the presence of a long upstream intron and a conserved, putative control region. The two pseudogenes were found to be fused by nonhomologous recombination. This study served as a template for the more extensive elucidation of the entire chromosome 17 OR cluster.

In the original report[15] OR genes were mapped according to their presence in, or absence from, the various cosmids, and the resolution level of their tentative locations was ± 5 kb. At present, of the 10 cosmids selected as a minimal set that cover the cluster, four are fully sequenced (cosmids 39, 46, 58 and 73), four are in finishing stages (cosmids 17, 26, 32 and D53), and one is in the shotgun stage (cosmid 68). Currently, about 300 kb of sequence are available.[24]

With the almost complete sequence data, the map is brought to its maximal resolution: the exact location and orientation for all 16 OR coding regions is now available (FIG. 1). No other non-OR genes were identified within the cluster. Sequence analysis revealed five previously undetected OR coding regions (OR17-1, OR17-7, OR17-25, OR17-30, and OR17-208). Of the 16 OR genes identified, six are pseudogenes. The open reading frames of these pseudogenes are interrupted by base deletions (OR17-1, OR17-23, and OR17-210), base substitution lead to a premature stop codon (OR17-208), gene fusion (OR17-24) or a combination of gene fusion and base deletions (OR17-25). Analysis of the sequences available from this cluster revealed a putative evolutionary dynamics, by which genes may have been born by duplication, and become inactive by gene fusion, base substitutions or base deletions (FIG. 2). In this proposed

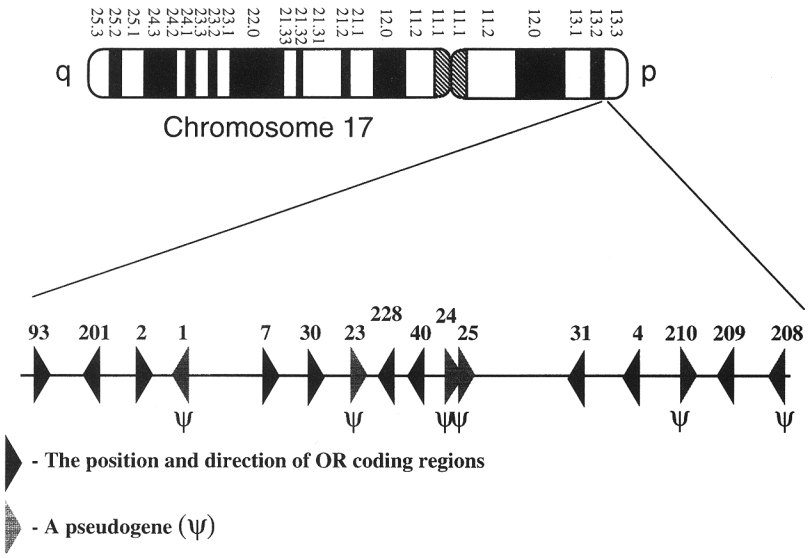


FIGURE 1. The olfactory receptor cluster located on human 17p13. *Upper part,* human chromosome 17 with *numbers* indicating the cytological bands. *Lower part,* cluster map. OR genes are denoted by trivial names and their orientations are represented by *arrow heads.* Pseudogenes are marked by Ψ.

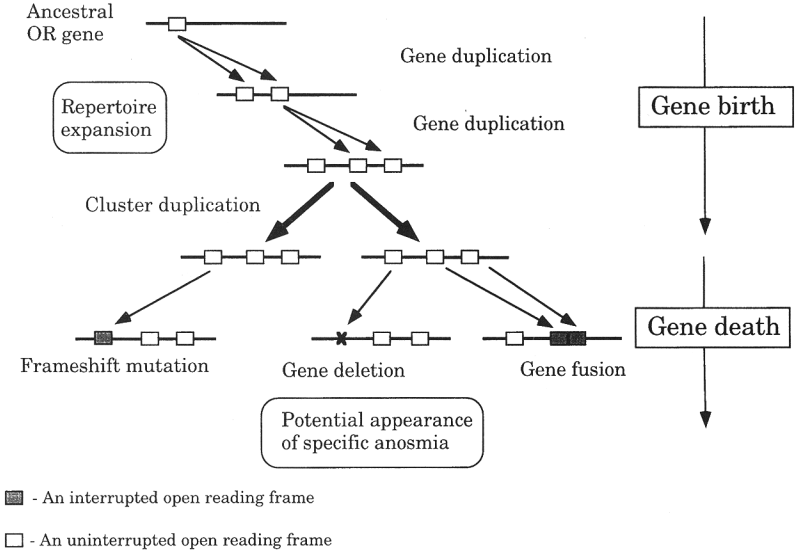


FIGURE 2. The evolutionary dynamics of the olfactory receptor subgenome.

model, the OR ancestral genes were duplicated to form clusters. Further expansion of the OR repertoire may have occurred by the generation of new clusters by large-scale duplications.

The OR coding regions located on the chromosome 17 OR cluster as well as identified ORs elsewhere in the human genome were multiply aligned, and the resulted evolutionary tree is presented in FIGURE 3. The chromosome 17 OR genes belong to four subfamilies of family 1 (1D, 1E, 1G, 1P) and one subfamily of family 3 (3A). Subfamily 3A has only five known members in the human genome, all of which are located on this cluster. Subfamilies 1G and 1P contain only one human OR member each (OR17-209 and OR17-208, respectively) and subfamily 1E contains three human genes, all located on this cluster. This evolutionary tree pattern shows that the OR genes located on chromosome 17 originated from at least two ancestral genes, and represent a relatively specialized segment of the human OR repertoire.

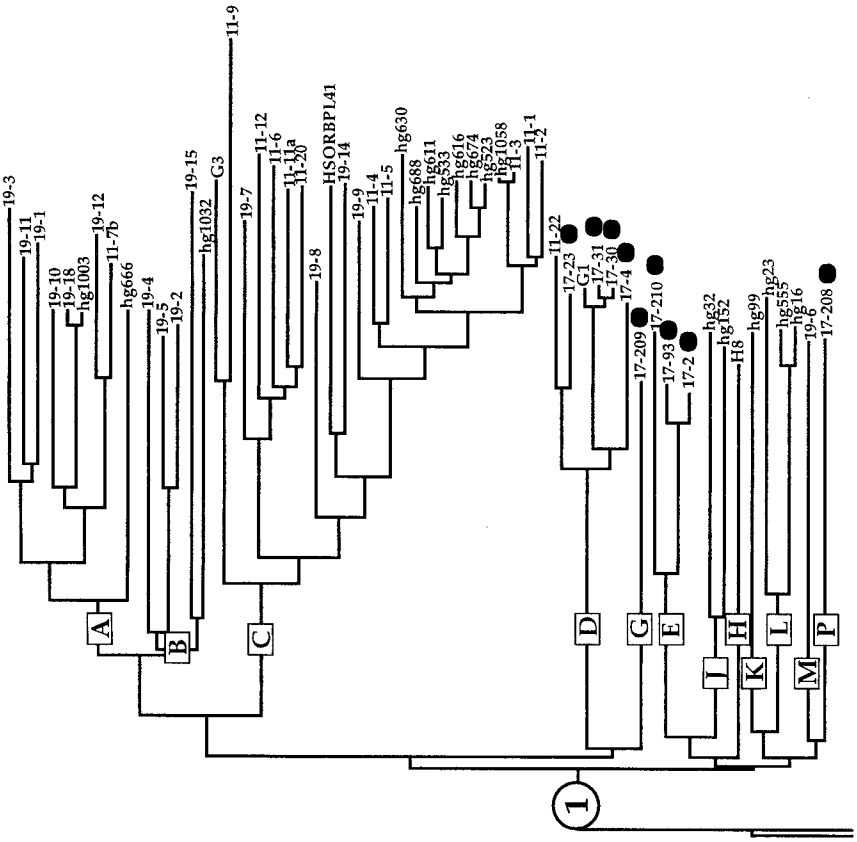
THE EVOLUTION OF THE OLFACTORY RECEPTOR PSEUDOGENE OR17-23

Pseudogenes have been found in almost every gene family that has been examined in detail.[5,25,26] Nonfunctional pseudogenes may be involved in gene conversion events and thus serve as variability donors to the functional gene pool.[4] On the other hand, pseudogenes can be corrected by gene conversion events and return to be functional.[27]

At least 35% of all known human OR coding regions are pseudogenes, mainly due to base deletions or insertions, which cause frameshifts and premature stop codons. One specific human OR pseudogene in the chromosome 17 cluster, OR17-23, is of particular interest. OR17-23 belongs to the subfamily 1D (FIG. 3), which contains five other known members, three of them located on the same cluster (17-4, 17-30, and 17-31). Sequence analysis of the members of this subfamily revealed that OR17-23 sustained two single base deletions: 496delG and 846delT, which cause a frameshift and therefore a premature stop codon. An intriguing question is how long ago did this variation arise? Our working hypothesis is that the human olfactory subgenome has undergone a relatively recent contraction, following a relaxation of the selective advantage of a keen sense of smell. In other words, many pseudogenes in humans may be active functional genes in lower mammals.

In order to shed light on this question, we have launched a study of OR genes in different primates. The coding regions of the orthologous genes (same gene loci in different species) for OR17-23 were sequenced from one ape (gorilla) and one old world monkey (macaque). Two potential alleles were cloned and sequenced from one gorilla individual. Both alleles share the same base deletions as the human ortholog; however, they differ from one another by six base substitutions and an internal duplication of 17 bp present only in allele #2 (FIG. 4A,B). Sequencing the PCR product from one macaque individual revealed two potential alleles, both of which have an intact open reading frames and differ in three base substitutions.

According to the above results, it is reasonable to assume that the human and gorilla orthologs have a common ancestor that contains the two base deletions, while the old world monkeys appear to have diverged from these species before the occurrence of the two deletions. Thus, the two base deletions, Gdel496 and Tdel846, are estimated to have happened between 8 to 31 million years ago, after the divergence time between old world monkeys and apes but before the divergence time between gorilla and human. This finding may represent one instance of the diminution of the OR repertoire in higher primates.


19-3
19-11
19-1
19-10
19-18
hg1003
19-12
11-7b
hg666
19-4
19-5
19-2
19-15
hg1032
G3
11-9
19-7
11-12
11-6
11-7a
11-20
19-8
HSORBP1A1
19-14
19-9
11-4
11-5
hg630
hg688
hg611
hg533
hg616
hg674
hg523
hg1058
11-3
11-1
11-2
11-22
17-23
G1
17-31
17-30
17-4
17-209
17-210
17-93
17-2
hg32
hg152
H8
hg99
hg23
hg555
hg16
19-6
17-208
A
B
C
D
G
E
I
H
K
L
M
P
1

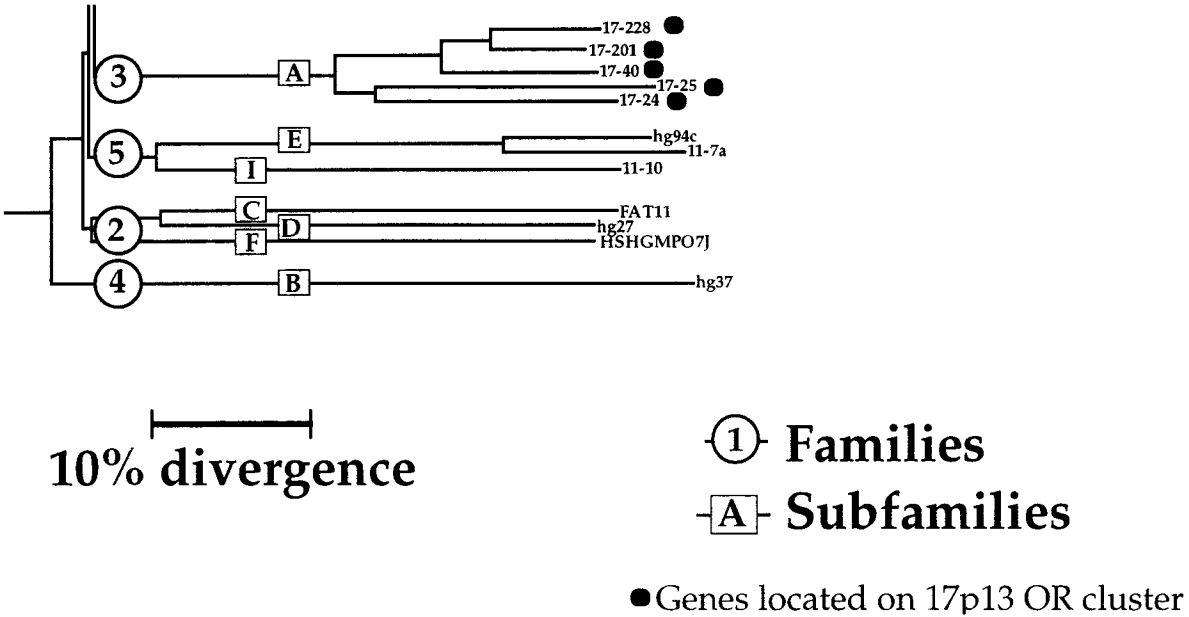


FIGURE 3. Phylogram of 69 human olfactory receptor protein sequences. The genes located in the 17p13 OR cluster are marked with *black circles.* Sequences were aligned using the ClustalW program.[29]

A (part 1)

```
Human       ATGGATGGAG GCAACCAGAG TGAAGGTTCA GAGTTCCTTC TCCTGGGGAT CTCAGAGAGT   60
Gorilla #1  .......... .......... .......... .......... .......... G.........
Gorilla #2  .......... .......... .......... .......... .......... G.........
Macaque     .......... .......... .A........ .......... .......... G.........
OR17-30     .......... AT........ ...GAAC... C......... .......... ..........

Human       CCTGAGCAGC AGCAGATGCT GTTTTGGATG TTCCTGGTCA GGTACCTGGT CACGGTGCTG  120
Gorilla #1  .......... .......C.. .......... .......... T......... .......G..
Gorilla #2  .......... .......C.. .......... .......... T......... .......G..
Macaque     .......... ...G...C.. .......... ......TC.. T......... .......G..
OR17-30     .......... .......C.. .......... ......TC.. T......... ..........

Human       GGAAATGTGC TCATCATCCT GGCCATCAGC TCTGATTCCC GCCTGCACAC CCCCATGTAC  180
Gorilla #1  .......... .......... .......... .......... .......... ..........
Gorilla #2  .......... .......... .......... .......... .......... ..........
Macaque     .......C.. .......... ........C. ........T. .......... ..........
OR17-30     .......... .......... .......... .......... A......... ..........

Human       TTCTTCCTGG CCAACCTCTC CTTCACTGAC CTCTTCTTTG TCACCAACAC AATCCCCAAG  240
Gorilla #1  .......... .......... .......... .......... .......... ..........
Gorilla #2  .......... ...G...... .......... .......... .......... ..........
Macaque     .......... T......... .......... .......... .......... ..........
OR17-30     .......... .......... .......... .......... .......... ..........

Human       ATGCTGGTGA ACCTCCAGTC CCAGAACAAA GCCATCTCCT ACACAGGGTG TCTGACACAG  300
Gorilla #1  .......... .......... .......... .......... ..G....... .....TG...
Gorilla #2  .......... .......... .......... .......... ..G....... .....TG...
Macaque     .......... .......... .......... ......G... .TG....... ......G...
OR17-30     .......... ..T....... .......... .......... .TG....... ..........

Human       CTCTACTTCC TGGTCTCCTT GGTGGCCCTG GACAACCTCA ACCTGGCCGT GATGGCGTAT  360
Gorilla #1  .......... .......... .......... .......... T......T.. ..........
Gorilla #2  .......... .......... .......... .......... T......... ..........
Macaque     .......... .A........ .......... .......... TT........ ..........
OR17-30     .......... .......... ....A..... .......... T......... ..........

Human       GATCGCTATG TGGCCATCTG CCGTCCCCTC CACTATGTCA CAGCCATGAT CCCTGGGCTC  420
Gorilla #1  .......... .......... .......... .......... .........G ..........
Gorilla #2  .......... .......... .......... .......... .........G ..........
Macaque     ..C....... .......... ..AC..T... .......... .........G ......A...
OR17-30     .......... .......... .T.C...... .......... .........G ..........

Human       TGTATCTTGC TCCTCTCCTT GTGTTGGGTG TTCTCTGCCC TCTATGGCCT CATCCATATC  480
Gorilla #1  .......... .......... ........G. .......... .......... ..........
Gorilla #2  .......... .......... ........G. .......... .......... ..........
Macaque     .......... .......... .......... .......T.. .......... ........C.
OR17-30     ...G...... .......... ........G. C.G....TT. .......... .C...TC.C.
```

FIGURE 4. **(A)** Multiple alignment of human OR17-23, its orthologs from gorilla and macaque, and its paralog (human OR17-30). For clarity, the only complete sequence shown is the human OR17-23; nucleotides are shown for other genes at positions where the sequence differs from that of human OR17-23. Gaps, represented by '-', were introduced to minimize nucleotide differences among aligned sequences.

A STRUCTURAL MODEL OF THE OLFACTORY RECEPTOR PROTEIN

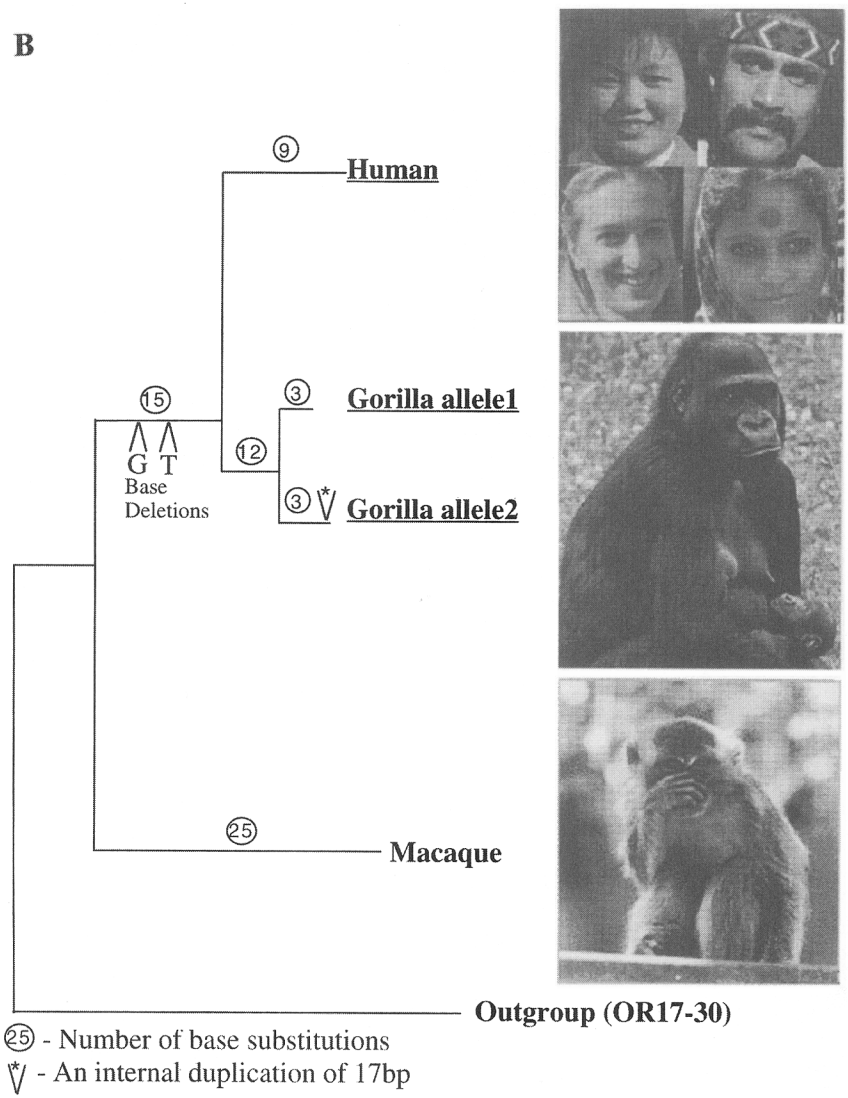
A recent effort in our laboratory has utilized the large amount of available information on OR protein sequences to shed new light on the potential structure of odor-

A (part 2)

```
Human       CTCCTCATGA CCAGG-TGAC CTTCTGTGGG TCTCAAAAGA TCCACTACCT CTTCTGTGAG   540
Gorilla #1  .......... .......... .......... .......... .......... ..........
Gorilla #2  .......... .......... .......... .......... .......... ..........
Macaque     .......... .....G.... .......... ..C.G..... .......... ..........
OR17-30     T.....C... .....G.... .......... C...G.G... .......... .........C

Human       ATGTACTTCC TGCTAAGGCT GGCATGTTCC AACATCCACG TCAACCACAC AGTACTGGTT   600
Gorilla #1  .......... ....GC.... .......... .......... .......... ..........
Gorilla #2  .......... ....GC.... .......... .......... .......... ..........
Macaque     ......G... ....GCR... .......... .......... .......... ....T.....
OR17-30     ......A... ....GT.... .......... ....C....A ...TT..... ..C.T..A..

Human       GCCACGGGCT GCTTCATCTT CCTCATCCCC TTAGGTTTCA TGATCACATC CTACGCCCGC   660
Gorilla #1  ..T....... .......... .......... .......... .......... ...T..T.A.
Gorilla #2  ..T....... .......... .......... .......... .......... ...T..T...
Macaque     ....S..... .......... .......... .......... .......G.. ..........
OR17-30     .....T.... .G........ .....C..T. .....G.... ...C...... ...T.TA..T

Human       ATTGTCAGAG CCATCCTCCA AATACCCTCA GCCACTGGGA AGTACAAAGC CTTCTCCACC   720
Gorilla #1  .......... .......... .......... .......... .......... ..........
Gorilla #2  .......... .......... .......... .......... .......... ..........
Macaque     .......... .........G .......... .......... .......... ..........
OR17-30     .........A .......T.. ...G.....G ...T..AA.. .A......A. T.....T...

Human       TGTGCTTCCC ATTTGGCTGT GGTCTCCCTC TTCTATGGGA CTCTGGGTAT GGTGTACCTG   780
Gorilla #1  .....C.... .......... .....T.... .....C.... ..G....... ..........
Gorilla #2  .C...C.... .......... .....T.... .....C.... .......... ..........
Macaque     .....C.... .......... .........T .......... .......... ...A......
OR17-30     .....C..G. ......G... .......... ..T....... .G..T.C... ..........

Human       CAGCCCCTCC AAACCT---- ---------- ---ACTCCAT GAAGGACTCA GTAGCCACAG   840
Gorilla #1  .......... .......... .......... .......... .......... ..........
Gorilla #2  .......... ......GCAG CCCCTCCAAA CCT....... .......... ..........
Macaque     .......... .......... .......... .......... .......... ..........
OR17-30     .......... .T........ .......... .......... .......... ..........

Human       TGATGTATGC GGTGGTGACG CC-ATGATTA ACCCTTTCAT CTACAGCCTG AGGAACAAGG   900
Gorilla #1  .......... .......... .......... .......T.. .......... ..........
Gorilla #2  .......... .......... .......... .......T.. T......... ..........
Macaque     .......... .........A ..C.....G. ....C..... .......... .R........
OR17-30     .......... T...C....A ..T.....G. .......... T......... ........A.

Human       ACATGCATGG GGCTCTGGGA AGACTTCGCC AAGGAAAAGC CTTCCAGAAG TTGACATGA    959
Gorilla #1  .......... .......... ..G....... .......... .......... .........
Gorilla #2  .......... .......... ..G....... .......... .......... .........
Macaque     .......... .......... .......T.. ....------ ---------- ---------
OR17-30     .......... .....C.... ...G.---.. T.T.G.G.C. ...T....G. CCT.A....
```

FIGURE 4. (A) *(continued)*

ants and G protein binding sites.[23] An amino acid variability profile was generated based on the multiple sequence alignment of 197 ORs from human, rat, mouse, dog and fish. This has lead to a definition of the relatively conserved sites of ORs, a subset of which is common also to other GPCRs. Among the OR-unique sites are the conserved motifs in the second and third intracellular loops that are implicated in G protein binding and specificity in other GPCRs (FIG. 5A). In addition, four cysteines are implicated in two OR-unique S-S bridges, one within extracellular loop 2, and the other between intracellular loops 2 and 3 (FIG. 5A). Such additional S-S bridges could be important for an enhanced structural stability of the OR proteins, known to be located only a few



(B) The evolution of pseudogene OR17-23 in primates. The evolutionary tree is based on the nucleotide sequences of OR17-23 in human, gorilla, and macaque with OR17-30 as outgroup. Pseudogenes are *underlined.* Genomic DNA was isolated from blood of two monkeys provided by the Israeli Safari Zoo: a gorilla *(Gorilla gorilla)* individual and a macaque *(Macaca fascicularis)* individual. The polymerase chain reaction (PCR) primers used were: upper—ATTTGTTGGTGTTAATGTTGC and lower—CACTTCAATGCCCAAAATTAC. PCR products were cloned and sequenced using the standard procedures. Sequences were aligned using the ClustalW program.[29] Accession numbers: OR17-23: U04679; OR17-30: U04681.

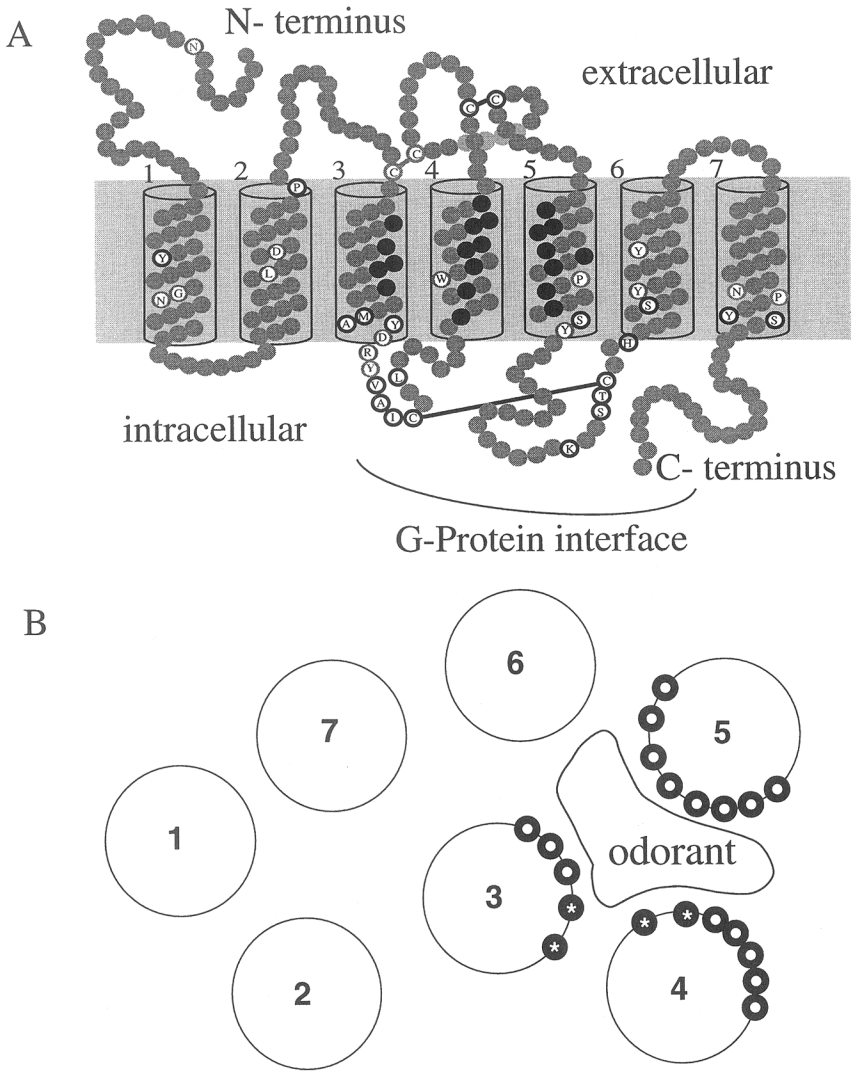


FIGURE 5. **(A)** Schematic flattened diagram of the OR protein. The set of the 20 variable positions that constitute the putative CDR on transmembrane segments 3, 4, and 5 are shown in *dark.* These positions are mainly clustered along one side of each helix. Highly conserved positions are either unique to ORs *(encircled in black)* or conserved in all GPCRs *(encircled in gray).* Putative cycteins bridges are marked in *black line* (unique to ORs) or *gray line* (conserved in all GPCRs). **(B)** A two-dimensional schematic representation of the OR protein, an extracellular view. The putative CDR residues are shown by *circles.* Positions populated mainly by hydrophobic side chains are marked with *white circles,* while positions populated mainly by hydrophylic side chains are marked with "*". A schematic odorant is shown on the putative binding site.

micrometers away from the external environment and thus exposed to more extreme temperature and pH changes.

In order to study the structure of the odorant binding site, to identify potential odorant contact residues, and to examine the notion that they should harbor excess intersequence variability,[8,11] we have constructed a homology-derived three-dimensional model for OR proteins, using rhodopsin as a template. Amino acid variability analysis confirmed that the transmembranal segments 3, 4, and 5 are highly variable (FIG. 5A).

Significant new insight has been subsequently obtained by performing a Fourier analysis on the patterns of sequence variability. This analysis has indicated that most of the variability is positioned along one side of these helices (FIG. 5A,B). Superimposing the variability information onto the three-dimensional model suggests that many of the variable residues may face the interior of the receptor barrel. The set of 20 interior-facing, variable residues thus defined is proposed to serve as the odorant complementarity determining region (CDR), in analogy to immunoglobulins.[28] This prediction is further supported by the finding that a considerable portion of the proposed OR CDR residues overlap, in a multiple alignment, with ligand-binding residues identified experimentally in other GPCRs.

Importantly, there exists a subset of variable OR positions with no such correspondence to ligand-binding positions in other GPCRs. This suggests the existence of a specialized ligand pocket unique to olfactory receptors, located in the cleft between transmembranal segments 4 and 5 (FIG. 5B). The existence of such an OR-unique ligand binding interface has important implications to our general understanding of GPCR function. In the future, when ligands are identified for specific ORs our proposed structural model could be verified.

REFERENCES

1. LI, W. H. 1997. Molecular evolution. Sinauer Associates, Inc. Sunderland, MA.
2. DROUIN, G. & M.M. DE SA. 1995. The concerted evolution of 5S ribosomal genes linked to the repeat units of other multigene families. Mol. Biol. Evol. **12:** 481–493.
3. DOENECKE, D., W. ALBIG, C. BODE, B. DRABENT, K. FRANKE, K. GAVENIS & O. WITT. 1997. Histones: Genetic diversity and tissue-specific gene expression. Histochem. Cell Biol. **107:** 1–10.
4. MCCORMACK, W. T., E. A. HURLEY & C. B. THOMPSON. 1993. Germ line maintenance of the pseudogene donor pool for somatic immunoglobulin gene conversion in chickens. Mol. Cell. Biol. **13:** 821–830.
5. BENEVOLENSKAYA, E. V., G. L. KOGAN, A. V. TULIN, D. PHILLIP & V. A. GVOZDEV. 1997. Segmented gene conversion as a mechanism of correction of 18S rRNA pseudogene located outside of rDNA cluster in D. melanogaster. J. Mol. Evol. **44:** 646–651.
6. FITCH, D. H., W. J. BAILEY, D. A. TAGLE, M. GOODMAN, L. SIEU & J. L. SLIGHTOM. 1991. Duplication of the gamma-globin gene mediated by L1 long interspersed repetitive elements in an early ancestor of simian primates. Proc. Natl. Acad. Sci. USA **15:** 7396–7400.
7. OHTA, T. 1991. Multigene families and the evolution of complexity. J. Mol. Evol. **33:** 34–41.
8. LANCET, D. & N. BEN-ARIE. 1993. Olfactory receptors. Curr. Biol. **3:** 668–674.
9. LANCET, D. & U. PACE. 1987. The molecular basis of odor recognition. Trends Biochem. Sci. **12:** 63–66.
10. REED, R. R. 1990. How does the nose know? Cell **60:** 1–2.
11. BUCK, L. & R. AXEL. 1991. A novel multigene family may encode odorant receptors: A molecular basis for odor recognition. Cell **65:** 175–187.
12. LANCET, D. 1991. Olfaction. The strong scent of success [news]. Nature **351:** 275–276.
13. CHESS, A., I. SIMON, H. CEDAR & R. AXEL. 1994. Allelic inactivation regulates olfactory receptor gene expression. Cell **78:** 823–834.
14. LANCET, D. 1986. Vertebrate olfactory reception. Annu. Rev. Neurosci. **9:** 329–355.

15. BEN-ARIE, N., D. LANCET, C. TAYLOR, M. KHEN, N. WALKER, D. H. LEDBETTER, R. CARROZZO, K. PATEL, D. SHEER, H. LEHRACH & M. A. NORTH. 1994. Olfactory receptor gene cluster on human chromosome 17: Possible duplication of an ancestral receptor repertoire. Hum. Mol. Genet. **3:** 229–235.
16. SULLIVAN, S. L., M. C. ADAMSON, K. J. RESSLER, C. A. KOZAK & L. B. BUCK. 1996. The chromosomal distribution of mouse odorant receptor genes. Proc. Natl. Acad. Sci. USA **93:** 884–888.
17. BEN ARIE, N., M. NORTH, M. KHEN, T. MARGALIT, H. LEHRACH & D. LANCET. 1992. Mapping the olfactory receptor 'sub-genome:' Implications to human sensory polymorphisms. Presented at the Cold Spring Harbor Symposium on Genome Mapping and Sequencing. May, 1992.
18. GLUSMAN, G., S. CLIFTON, R. ROE & D. LANCET. 1996. Sequence analysis in the olfactory receptor gene cluster on human chromosome 17: Recombinatorial events affecting receptor diversity. Genomics **37:** 147–160.
19. BRAND, V., S. ROUQUIER, P. DE JONG, T. S., J. DEMAILLE & D. GIORGI. 1996. Olfactory receptor gene clusters on human chromosome 3. HUGO's Human Gen. Meeting [abstract 60]: 14.
20. FAN, W., Y. C. LIU, S. PARIMOO & S. M. WEISSMAN. 1995. Olfactory receptor-like genes are located in the human major histocompatibility complex. Genomics **27:** 119–123.
21. BUETTNER, J. A., D. LANCET & G. A. EVANS. 1996. Organization of olfactory receptor genes on human chromosome 11. Am. J. Hum. Genet. 59 (Suppl.): A299.
22. OLSEN, A. S., R. R. REED, B. F. BRANDRIFF, L. A. GORDON, S. M. G. HOFFMAN, K. SCHRADER, S. TSUJIMOTO, A. V. CARRANO & H. MOHRENWEISER. 1993. Assembly and analysis of a cosmid contig map of the olfactory receptor gene family on chromosome 19. Am. J. Hum. Genet. **51** (Suppl.): A243.
23. PILPEL, Y. & D. LANCET. 1997. Functional sites of olfactory receptors: A multisequence modeling approach. Proteins. Submitted.
24. BEN-ASHER, E., S. HORN-SABAN, G. GLUSMAN, J. LAMMARDIN, A. OLSEN, S. W. CLIFTON, B. A. ROE, J. DEMAILLE & D. LANCET. 1997. DNA sequencing of three olfactory receptor gene clusters on human chromosomes 17 and 19. Microb. Comp. Genomics **2:** 188.
25. BECK, S., S. ABDULLA, R. P. ALDERTON, R. J. GLYNNE, I. G. GUT, L. K. HOSKING, A. JACKSON, A. KELLY, W. R. NEWELL, P. SANSEAU, E. RADLEY, K. L. THORPE & J. TROWSDALE. 1996. Evolutionary dynamics of non-coding sequences within the class II region of the human MHC. J. Mol. Biol. **255:** 1–13.
26. MARCHESE, A., T. V. BEISCHLAG, T. NGUYEN, H. B. NIZNIK, R. L. WEINSHANK, S. R. GEORGE & B. F. O'DOWD. 1995. Two gene duplication events in the human and primate dopamine D_5 receptor gene family. Gene **154:** 153–158.
27. BENEVOLENSKAIA, E. V., G. L. KOGAN, M. D. BALAKIREVA, D. FILIPP, I. P. ARMAN & V. A. GVOZDEV. 1994. Variability of rDNA genes, detected as a result of analyzing a pseudogene nucleotide sequence in *Drosophila melanogaster.* Genetika **30:** 318–325.
28. WU, T. T., E. A. KABAT & H. BILOFSKY. 1975. Similarities among hypervariable segments of immunoglobulin chains. Proc. Natl. Acad. Sci. USA **72:** 5107–5110.
29. HIGGINS, D. G., J. D. THOMPSON & T. J. GIBSON. 1996. Using CLUSTAL for multiple sequence alignments. Methods Enzymol. **266:** 383–402.