Chapter 1

Introduction

1.1. The olfactory system

In the battle for survival, how an organism responds to various stimuli, such as food sources, potential mates and potential predators, is highly important in determining the evolutionary fitness of a particular individual. Sight, smell and hearing are all vital in an individual's struggle to survive, mate and successfully pass their selfish genes into the next generation. Of these senses the sense of smell, or olfaction, as the perception of odours is more properly known, is considered to be the most primitive. Lower organisms, for example *Caenorhabditis elegans*, are heavily reliant on their olfactory system to perceive the world and make the kind of choices likely to enhance their reproductive fitness. Even in mammals where hearing and sight have evolved, a functioning olfactory system enhances an individual's chances of survival; dogs and mice rely on odours to locate food, recognize territory, identify kin, and find a receptive mate.

The vertebrate olfactory system is well adapted for the detection and recognition of small odorant molecules. It is considered to be able to detect as many as 10,000 different odorants, and it has the ability to detect both subtle differences between chemical stereoisomers and the vastly different chemical structures that odorant molecules may possess. It is also able to detect some odorants at very low airborne concentrations of less than several parts per trillion (Snyder *et al.*, 1988). This ability to deal with very different olfactory inputs in vertebrates is located in one major neural structure, the major olfactory epithelium (MOE) which is a specialized neuroepithelium located in the posterior cavity of the nose. A secondary olfactory organ, the vomeronasal organ (VNO), is located further towards the front of the nose: this organ is

considered to be only involved in the detection of pheromones, chemical signals conveying social and sexual information.

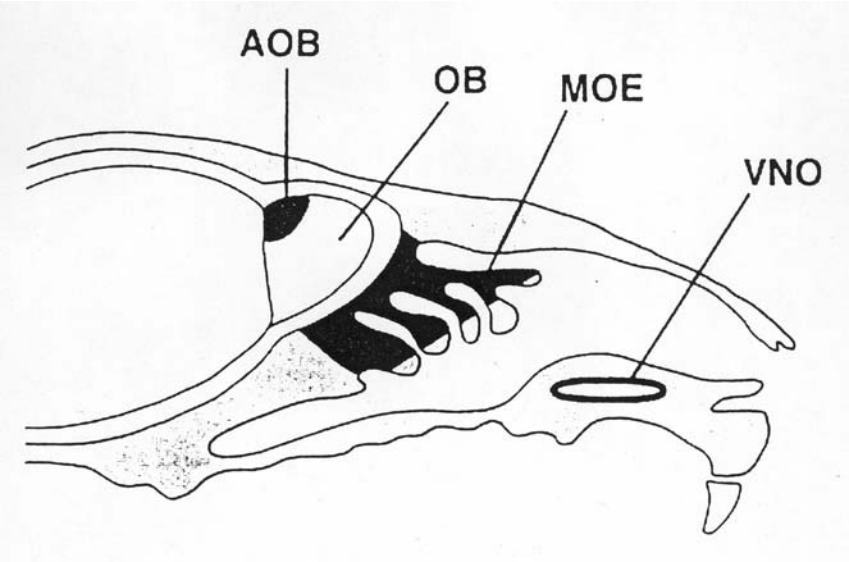


Figure 1.1: Anatomy of the olfactory system. A sagittal view of the rat head, showing the location of the vomeronasal organ (VNO) and the major olfactory epithelium (MOE) within the rat nose. The two structures project their axons to different structures within the rat brain: the sensory neurons of the MOE projecting axons to the olfactory bulb (OB), whilst sensory neurons of the VNO project to the accessory olfactory bulb (AOB).

Figure from Buck and Axel (1991).

1.2. The major olfactory epithelium: anatomical organization and signal transduction

The sensory epithelium of the MOE is made up of three cell types: olfactory sensory neurons (OSNs), their precursor basal cells, and sustentacular cells (which have glia-like, supportive functions within the epithelium). Olfactory sensory neurons are bipolar cells that project a single unmyelinated axon to the olfactory bulb located in the anterior part of the skull. A single dendrite is projected to the epithelial surface where it terminates in a dendritic knob and projects specialized cilia into the nasal lumen. These cilia, which lie in the thin layer of mucus that cover the tissue, provide a large surface area where olfactory receptors are exposed to a stream of warmed, moistened and possibly concentrated odorants. Throughout an individual's lifetime,

OSNs appear to continuously undergo cell death and be regenerated from the precursor basal-cell population (Levy *et al.*, 1991).

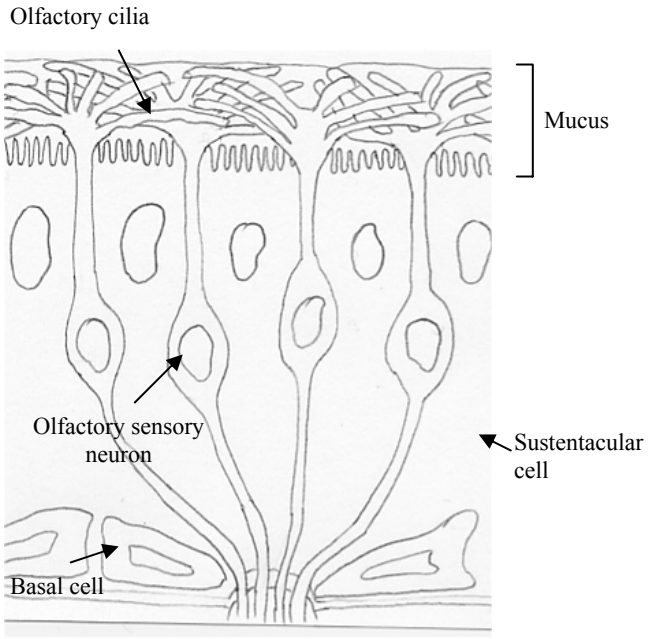


Figure 1.2: The structure of the major olfactory epithelium: odorant molecules enter the nose and dissolve in the layer of mucus covering the OSN cilia, and odorant binding occurs. OSNs are regenerated from basal precursor cells, whilst sustentacular cells have a supportive function within the structure.

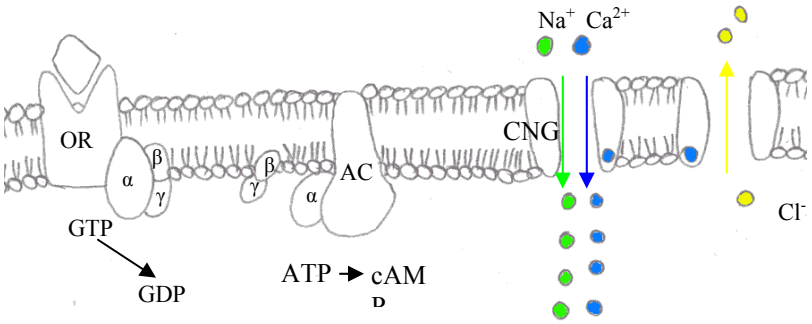
Adapted from Firestein (2001).

Within the MOE, the process of olfaction signal transduction is believed to occur as follows. Firstly, odorants enter the nasal cavity where they dissolve in the mucus that covers the luminal surface of the olfactory epithelium. Once dissolved, the odorant molecules are able to bind to specific olfactory receptors (ORs) that are located on the cilia of the dendrites of olfactory sensory neurons. When the odorant molecule ligand has bound to the OR, these cell-surface transmembrane receptors undergo a conformational change, which facilitates interaction with a guanine triphosphate (GTP) binding protein (G protein). G proteins are composed of three subunits (α, β, and γ). The interaction between an olfactory receptor protein and a G protein (likely to be the olfactory specific G_{olf}) results in the exchange of GDP for GTP on the α subunit, and a subsequent dissociation of this subunit from the β and γ subunits. The activated α subunit

interacts with a second messenger enzyme, which in the case of G_{olf} is adenylyl cyclase. The cyclase converts adenine triphosphate (ATP) into cyclic adenine monophosphate (cAMP).

G-protein-dependent elevation of cyclic AMP (cAMP) leads to cAMP binding to the intracellular face of an ion channel, allowing the channel to conduct cations such as Na^+ and Ca^{2+}. Membrane depolarization triggers an action potential which is projected to the olfactory bulb. In addition to this pathway, identified using genetically altered mice lacking the various components of this transduction cascade (G_{olf}, adenylyl cyclase and cyclic nucleotide-gated channel (Brunet *et al.*, 1996, Belluscio *et al.*, 1998, Wong *et al.*, 2000)), OSNs also have an amplification mechanism. The calcium ions trigger the opening of an ion channel that allows negatively charged chlorine ions to leave the cell, increasing the net positive charge across the membrane (Kleene and Gesteland, 1991). As well as increasing the membrane depolarization, the calcium ions also act in the negative feedback pathway, acting (probably with calmodulin) on the ion channel to decrease its sensitivity to cAMP (Kurahashi and Menini, 1997). Additional down regulators of the signal include a RGS (regulator of G-protein signalling) protein which acts on adenylyl cyclase to decrease its activity (Sinnarajah *et al.*, 2001), and a kinase that phosphorylates activated receptors sending them into a desensitized state (Dawson *et al.*, 1993, Schleicher *et al.*, 1993).

Information about these action potentials generated in the OSNs is transmitted to the olfactory bulb. In the olfactory bulb, axons of olfactory sensory neurons form synapses with the dendrites of secondary neurons (mitral cells and tufted cells) and interneurons (periglomerular cells) within structures known as glomeruli. Mitral and tufted cells integrate the input from the neurons and local inhibitory currents before relaying this information to the olfactory cortex and other central brain areas via the lateral olfactory tract.



Adapted from Firestein (2001).

Figure 1.3: Sensory transduction in olfactory sensory neurons (OSNs). The OR ligand binds triggering disassociation of the GPCR and the α-subunit activates an adenylyl cyclase (AC). The cyclase converts ATP to cAMP which acts to open the cyclic nucleotide gated (CNG) ion channel. Calcium ions flowing into the cell trigger the opening of an ion channel allowing chlorine ions to leave the cell, increasing the net positive charge across the membrane.

1.3. The molecular basis of the olfactory system.

In spite of all the knowledge about the olfactory system, the molecular basis of olfaction remained elusive for many years. It was, then, a pioneering piece of work from Buck and Axel (1991) that identified a large family of genes considered to encode olfactory receptor genes. Buck and Axel's work on olfactory receptor genes was based on three assumptions. Firstly, the likely involvement of G proteins in the olfactory systems suggested ORs were likely to share structural and sequence similarities with the G-protein coupled receptor (GPCR) superfamily. The superfamily of GPCRs have highly conserved regions within their 7 transmembrane domain structure, allowing degenerate primers to be designed based on these conserved regions. Secondly, Buck and Axel hypothesized that ORs were members of a multigene family of considerable size and diversity because a large number of chemicals with differing structures can be detected. Thirdly, it was suggested that ORs were likely to be expressed only in olfactory sensory neurons.

Buck and Axel, therefore, used degenerate oligonucleotides known to anneal to conserved regions of G-protein coupled receptor genes together with cDNA sequences from rat olfactory epithelium to identify various putative olfactory receptor genes. Looking for a multigene family, they focused on a pair of primers that appeared to amplify a large number of genes. Further experiments on this multigene family revealed that it fitted the third specification: Northern blot analysis revealed that members of the multigene family were expressed exclusively in the olfactory epithelium of rat.

The 18 olfactory receptor proteins isolated in these experiments were aligned, and an additional observation seemed to support the idea that these were olfactory receptors. This observation was that, although the proteins were found to share structural and sequence similarities with the G-protein coupled receptor superfamily of neurotransmitters and hormone receptors, including seven hydrophobic stretches considered to represent seven transmembrane domains, in contrast to other GPCRs where maximum sequence conservation is seen within the transmembrane domains, these novel genes showed striking divergence within the third, fourth and fifth transmembrane domains. This divergence, within the transmembrane domains considered to be involved in ligand binding in other 7 transmembrane proteins (Kobilka, 1992), was consistent with the theory that maximal diversity between olfactory receptors would be expected to be found in the ligand binding regions, allowing the multigene family to interact with a large number of odorant molecules.

A further similarity with some other 7 transmembrane protein genes was that these putative olfactory receptor genes were found to have no introns within their coding sequences. This lack of introns allowed an estimate of the number of OR genes within the genome to be made. Southern blots were probed with OR genes that did not cross-hybridize, and it was found that each gene hybridized to 1 to 17 bands. In all a total of 70 bands were detected which led to an

idea that the rat genome contained about 100 to 200 OR genes. Later estimates produced a much higher number of suggested OR genes: in 1992, for example, Buck suggested the size of the multigene family could be estimated as between 500 to 1000 genes (Buck, 1992). These type of estimates suggested the OR gene family would be the largest family in the mammalian genome, with 0.8-1.6% of the 60,000 mammalian genes likely to be OR genes.

The initial identification of OR genes, therefore, revealed they were genes consisting of small (generally less than 1 Kb), intronless open reading frames. Consensus amino acid motifs could be highlighted: LHTPMY in intracellular loop 1, MAYDRYVAIC at the end of the third predicted transmembrane domain, SY at the end of transmembrane domain 5, FSTCSSH in transmembrane domain 6, and PMLNPF in transmembrane 7. Hypervariable regions, which could correspond to ligand binding sites were highlighted in transmembrane domains 3, 4 and 5. From initial experiments it was also known that a given OR gene would cross-hybridize to a small set of related OR genes. Sets of related OR genes suggested the OR gene repertoire could be subdivided into subfamilies (Lancet *et al.*, 1993).

The idea that transmembrane domains 3, 4 and 5 represent the hypervariable binding region of the protein has been supported by data from three dimensional models of other GPCR proteins (Pilpel and Lancet, 1999) which suggest that these three α-helical barrels arrange themselves into a pocket, one third of the way into the membrane. Studies on adrenergic receptors (Kobilka *et al.*, 1988) and the binding site of retinal in rhodopsin (Palczewski *et al.*, 2000) also suggest that these three domains are likely to be the binding site of these proteins.

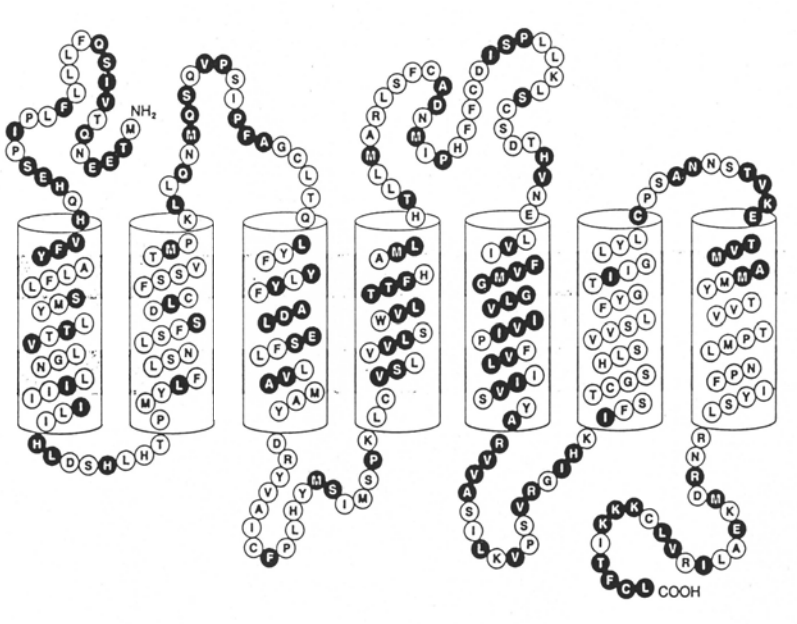
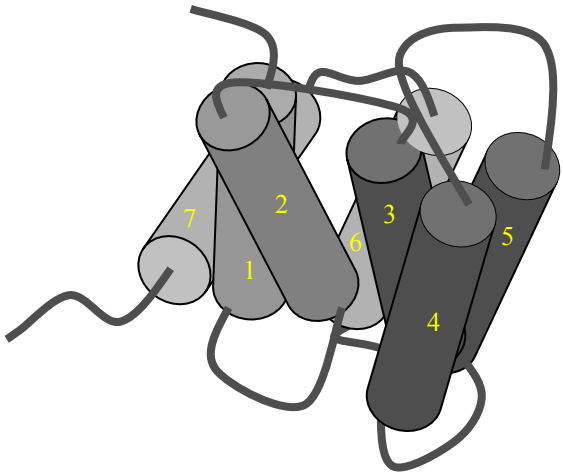


Figure 1.4a: Schematic diagram of the protein structure of an olfactory receptor: amino acids are represented as balls and the 7 predicted transmembrane regions are enclosed within cylinders. Conserved amino acids are shown in white, whilst highly divergent positions are black. This diagram shows the hypervariability of transmembrane regions, 3, 4 and 5, compared to the other transmembrane regions.

Figure from Buck and Axel (1991).

Figure 1.4b: Proposed three dimensional structure of an olfactory receptor protein, with degree of conservation within the transmembrane regions plotted according to conservation of amino acids shown in figure 1.3a. (Dark grey barrels suggest highly divergent region of protein, whilst lighter grey barrels suggest a greater degree of conservation.)

Adapted from Firestein (2001).



1.4. Cloning of further olfactory receptor genes in a variety of species

Following Buck and Axel's work, OR genes were identified in a number of species. Generally, degenerate PCR primers have been used to amplify related genes from genomic DNA, or (more rarely, since cDNA from olfactory mucosa is difficult to obtain) from olfactory epithelium cDNA

libraries. In rat, for example, further OR genes were identified through degenerate PCR of genomic DNA (Levy *et al.*, 1991, Raming *et al.*, 1993, Strotmann *et al.*, 1994b, Drutel *et al.*, 1995, Thomas *et al.*, 1996) and through degenerate PCR of testis cDNA (Vanderhaeghen *et al.*, 1997b). Rat OR genes were also identified using mRNA from the axon terminals of olfactory sensory neurons in the olfactory bulb (Singer *et al.*, 1998). In addition to OR genes identified in rat, OR genes were identified in various fish species: catfish (*Ictalarus punctatus*) (Ngai *et al.*, 1993b), zebrafish (*Danio rerio*) (Barth *et al.*, 1996, Byrd *et al.*, 1996, Weth *et al.*, 1996, Barth *et al.*, 1997), lamprey (*Lampetra fluviatilis*) (Berghard and Dryer, 1998), and mudpuppy (*Necturus maculosus*) (Zhou *et al.*, 1997). Chicken OR genes were cloned to provide an insight into OR genes in birds (Leibovici *et al.*, 1996), whilst information from the frog (*Xenopus laevis*) genome (Freitag *et al.*, 1995) suggests the OR gene repertoire in amphibians consists of some OR genes more closely related to those found in fish species (named Class I ORs by Freitag *et al.*, 1995), and some more closely related to those found in mammals (named class II ORs by the same group Freitag *et al.*, 1995). Information about the OR gene repertoire of other mammalian species has also been accumulated: there is a large amount of data about the OR gene family in mouse (Nef *et al.*, 1992, Ressler *et al.*, 1993, Asai *et al.*, 1996, Mombaerts *et al.*, 1996a, Sullivan *et al.*, 1996, Kubick *et al.*, 1997, Vanderhaeghen *et al.*, 1997b, Qasba and Reed, 1998), and a small amount of data about dog (Parmentier *et al.*, 1992, Issel-Tarver and Rine, 1996, Issel-Tarver and Rine, 1997, Vanderhaeghen *et al.*, 1997b), pig (Matarazzo *et al.*, 1998, Velten *et al.*, 1998) and various primates (Sharon *et al.*, 1999).

In humans, OR genes were initially found in a testis cDNA library: Parmentier *et al.* (1992) discovered that the orphan 7 TM receptor genes they had previously identified were orthologous to the rat OR genes identified by Buck and Axel. These genes were all likely to be functional, with complete, non-pseudogenic open reading frames, but later work on human OR genes revealed that a large number of the repertoire were pseudogenes (Selbie *et al.*, 1992) Some of

these pseudogenes were identified from a cDNA library made from olfactory tissue (Crowe *et al.*, 1996), suggesting some olfactory sensory neurons may express non-functional OR genes.

1.5. The genomic organization of olfactory receptor genes.

Initial work on the genomic organization of olfactory receptor genes focused on a cluster of OR genes located on chromosome 17p (Ben-Arie *et al.*, 1994), where the first OR gene had been mapped (Schurmans *et al.*, 1993). 16 human OR genes located in a 350 Kb cluster were cloned (Ben-Arie *et al.*, 1994), but when one of the cosmids predicted to contain 6 of these OR genes was sequenced (Glusman *et al.*, 1996), only 3 OR genes mapping to this cosmid were found. This highlighted the potential problem involved in sequencing OR genes from cloned PCR products: it may be that many of the sequences could have been artificially generated by recombination between highly related nucleotide sequences (Meyerhans *et al.*, 1990, Mombaerts, 1999). Further work on this region of the genome lead to the characterization of a 412 Kb contiguous sequence, containing 17 OR genes, 10 with intact open reading frames, and 7 of which were predicted to be pseudogenes (Glusman *et al.*, 2000).

3 further human olfactory receptor genes were generated through sequencing a 36 Kb cosmid located on chromosome 19 (Trask *et al.*, 1998). Further work on human OR genes involved mapping genes to various locations within the human genome. Fan *et al.* (1995) mapped 3 OR genes to the Major Histocompatibility Complex class I region on chromosome 6p21, whilst human OR genes orthologous to 4 dog genes mapped to chromosomes 7q35, 11q11 and 19p13 (Issel-Tarver and Rine, 1997, Carver *et al.*, 1998). OR genes found within testis cDNA libraries were located to chromosomes 11p22, 17q21 and 19p13-19p31 (Vanderhaeghen *et al.*, 1997a), whilst the 36 Kb sequence block containing 3 OR genes and located on chromosome 19 was used as probe to discover further OR gene regions. Using this sequence, Trask *et al.* (1998) found the

block was duplicated on chromosomes 3q and 15q, and it also appears that additional copies of the block are present in different individuals (a human genome can contain between 7 to 11 copies of this block).

An attempt to try and locate all OR gene-containing sites within the human genome was made by Rouquier *et al.* (1998): a pool of OR gene fragments from degenerate PCR on genomic DNA was used as a probe in a series of fluorescence *in situ* hybridization (FISH) experiments on metaphase chromosomes. These experiments revealed that all chromosomes, except chromosome 20 and chromosome Y, contain OR genes, and suggested OR genes were present in between 25-53 locations within the human genome.

1.6. The expression of olfactory receptor genes within the olfactory system

The initial discovery of OR genes within the olfactory epithelium suggested that these genes were expressed in olfactory sensory neurons. Subsequent work confirmed that many OR genes are expressed in mature OSNs: *in situ* hybridization experiments suggested that specific OR genes were expressed in a subset of OSNs that have a characteristic bilateral symmetry (Nef *et al.*, 1992, Strotmann *et al.*, 1992, Raming *et al.*, 1993). Further work on expression of OR genes in the olfactory epithelium in mouse (Ressler *et al.*, 1993) and the rat (Vassar *et al.*, 1993) revealed that OSNs expressing a certain specific OR gene were located in one of several nonoverlapping zones. Within the specific zone, a random expression pattern with bilateral symmetry can be observed. Initial work suggested there were 3 zones of expression within the olfactory epithelium, but later studies suggested 4 expression zones in the mouse olfactory epithelium (Sullivan *et al.*, 1995, Sullivan *et al.*, 1996). Expression at the protein level, using an antiserum against a rat OR, also provides evidence for the zonal pattern of expression of OR genes (Koshimoto *et al.*, 1994). No physiological basis for this zonal organization has been uncovered, however, it is clear that

the zones can be discerned from the earliest embryonic stages at which OR genes are expressed (Sullivan *et al.*, 1995, Menco and Jackson, 1997). A large number of OR probes have all provided support for the zonal expression hypothesis (Ressler *et al.*, 1993, Vassar *et al.*, 1993, Strotmann *et al.*, 1994a, Strotmann *et al.*, 1994b, Sullivan *et al.*, 1996, Kubick *et al.*, 1997). However, one group of genes (OR37 group) does not show this zonal expression pattern; instead, they appear to be expressed in a 'patch' on the tips of some turbinates and not in the septum of the rat, mouse and guinea pig olfactory epitheliums (Strotmann *et al.*, 1992, Strotmann *et al.*, 1994a, Strotmann *et al.*, 1994b, Strotmann *et al.*, 1995). Zonal expression of OR genes in other species has not been demonstrated so clearly: in zebrafish, no observations regarding zones were made at either adult (Barth *et al.*, 1996, Byrd *et al.*, 1996) or embryonic (Vogt *et al.*, 1997) stages. Early observations in catfish also suggested that the expression of a particular OR gene was constant across the olfactory epithelium (Ngai *et al.*, 1993a), however, a later quantitative analysis suggested the existence of 3 or 4 expression domains (Weth *et al.*, 1996). In *Xenopus laevis* the two different classes of ORs appear to be expressed in different regions (Freitag *et al.*, 1995): this could represent a functional adaptation with one region specialized for detecting waterborne odors and the other airborne odors, or it could represent a zonal separation of expression similar to that observed in rat and mouse.

The expression of OR mRNA in the axon terminals of OSNs in the olfactory bulb means it is also possible to consider the distribution of neurons expressing a particular OR gene within the olfactory bulb. Probing the olfactory bulb with rat OR probes revealed that neurons expressing an OR gene appear to converge to a few discrete sites within the bulb (Ressler *et al.*, 1994, Vassar *et al.*, 1994). The location of these converging axons is bilaterally symmetric, and locations appear to be very similar between members of the same species. These experiments, however, did not provide single axon resolution: it may have been that individual axons expressing the same OR projected to different locations. These neurons would not have been detected because expression

of the OR can only be detected where there are a number of axons converging. A single axon approach, using a mutant OR gene together with the marker taulacZ, however, suggested all axons expressing this particular allele, converged to fixed, symmetric locations within the olfactory bulb (Mombaerts, 1996, Mombaerts *et al.*, 1996a, Mombaerts *et al.*, 1996b). This targeting of OSNs expressing a specific OR gene to specific glomeruli appears to be partially controlled by the OR gene. Evidence for this type of mechanism is available, for example, in one experiment a mutation introduced into an OR gene meant compared to the non-mutant form, a different glomeruli was targeted (Mombaerts *et al.*, 1996a). Further work on this mouse locus confirmed that OR genes are one component in this guidance process, although it appears there are other important components (Wang *et al.*, 1998).

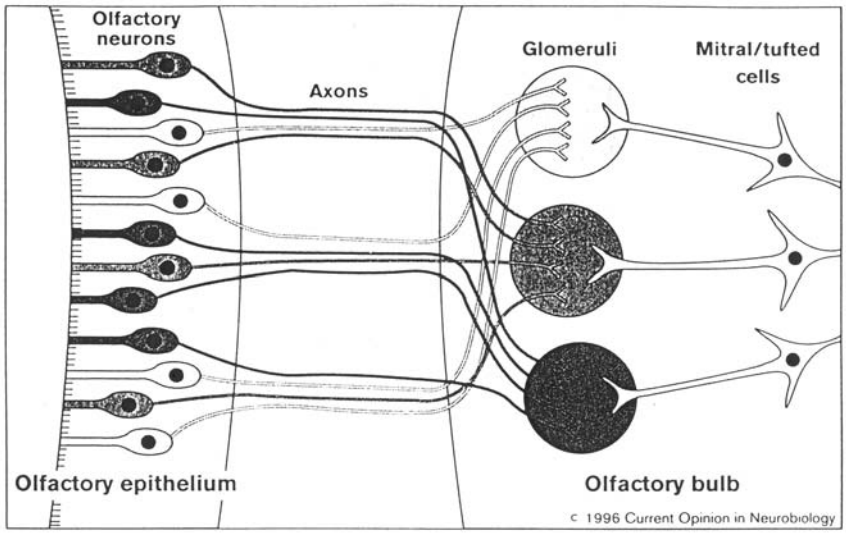


Figure 1.5: OR targeting within the olfactory bulb. OSNs expressing the same OR gene target the same glomerulus within the olfactory bulb, for example, all neurons expressing the 'white' OR within the MOE project to the 'white' glomerulus within the OB.

From Mombaerts (1996).

The fixed location of glomeruli relating to one specific OR suggests odor perception is encoded by a combination of activated glomeruli. This is supported by physiological observations in mammals (Hildebrand and Shepherd, 1997) and by observations from the zebrafish olfactory bulb where 80 defined glomeruli (Baier and Korsching, 1994) show stereotyped patterns of glomerular activation in response to various odorant molecules (Friedrich and Korsching, 1997).

OR genes, therefore, are expressed in olfactory sensory neurons, and this expression can be detected in the major olfactory epithelium, and also, at a much lower level in the olfactory bulb. Both the olfactory epithelium and the olfactory bulb show distinct patterns of expression; within the olfactory epithelium, OSNs expressing a specific OR gene are located in a constant expression zone, whilst in the olfactory bulb, OSNs expressing the same OR gene converge to specific glomeruli. The expression of OR genes, however, does not appear to be restricted to the olfactory system: there is evidence that olfactory receptor genes are expressed in other tissues.

1.7. The expression of olfactory receptor genes outside the olfactory system

Expression of OR genes within the testis has been observed since the initial identification of olfactory receptor genes; the human counterparts to the Buck and Axel's rat OR genes were cloned from a testis cDNA library (Parmentier *et al.*, 1992). In addition to the expression of OR genes in human testis, OR genes have been reported to be expressed in the testis of dog, mouse and rat (Vanderhaeghen *et al.*, 1993, Walensky *et al.*, 1995, Vanderhaeghen *et al.*, 1997a, Vanderhaeghen *et al.*, 1997b, Walensky *et al.*, 1998). OR genes also appear to be expressed in spermatids and spermatozoa: RNA has been found in postmeiotic round spermatids (Parmentier *et al.*, 1992, Walensky *et al.*, 1998), whilst antisera detected OR protein in late round and elongated spermatids, and on the tail midpiece of mature spermatozoa (Vanderhaeghen *et al.*, 1993, Walensky *et al.*, 1995). Expression of OR genes within the testis may result from the aberrant regulation of transcription. Alternatively, there may be some functional reason for the expression of these genes within the testis, for example, they may regulate sperm maturation, sperm motility or sperm attraction to oocytes. The signal transduction machinery associated with ORs is present in the cells of rat testis (G-protein receptor kinase 3, β-arrestin (Walensky *et al.*,

1995), adenylyl cyclase III (Gautier-Courteille *et al.*, 1998)) suggesting that ORs may have some biological role within the testis.

Expression of OR genes has also been reported in other non-olfactory tissues, for example, the heart of rat (Drutel *et al.*, 1995), the notochord of chick (Nef and Nef, 1997), the brainstem of rat (Raming *et al.*, 1998) and mouse (Conzelmann *et al.*, 2000), and the erythroid cells of human and mouse (Feingold *et al.*, 1999). Expressed sequence tag (EST) libraries have also suggested that OR genes are expressed in other non-olfactory organs, such as the lungs, heart, liver, placenta, colon, ovaries, as well as sperm cells. This could be due to genomic contamination of EST libraries, or it could be due to erroneous transcriptional processes. Alternatively, it may be that these OR genes have been recruited for a non-olfactory purpose in different tissues. One non-olfactory purpose was proposed by Dreyer (1998) who suggested that olfactory receptors may be important in providing part of a molecular addressing code required during development. This model advances the idea that cells assemble organisms through a so-called 'area code' that functions like the country, area, regional and local portions of the telephone dialing system. The olfactory receptor genes can be seen as the last digits in this cell 'area code', an idea that fits in with the predicted large size of the olfactory receptor gene family and the apparent widespread tissue distribution of these genes.

1.8. The regulation of olfactory receptor genes

The size of OR gene repertoire has been estimated to be about 500-1000 genes in rat (Buck, 1992), 400 in dog (Parmentier *et al.*, 1992), 600 in mouse (Qasba and Reed, 1998), and up to 100

in catfish and zebrafish (Ngai *et al.*, 1993b, Barth *et al.*, 1996, Weth *et al.*, 1996). This large number of OR genes means it is difficult to determine the number of OR genes expressed per olfactory sensory neuron, but *in situ* hybridization of the olfactory epithelium with OR probes has not shown any colocalization of OR genes, so it is generally considered that an olfactory sensory neuron expresses a single or a very small number of olfactory receptor genes (Mombaerts, 1999). Expression of OR genes within olfactory sensory neurons, however, is known to be restricted so only one allele of a given OR gene is expressed (Chess *et al.*, 1994). In a cross between mice containing 2 different allelic forms of 2 OR genes, it was found that the offspring expressed either the maternal or paternal allele in OSNs. Regulation of OR genes, therefore, must be tightly controlled in order to express 1 allele of 1 (or very few) OR genes, but little is known about the mechanisms controlling this regulation.

1.9. Ligands of olfactory receptor genes

In spite of the large amounts of data about OR genes that has been accumulated, the vast majority of OR genes remain orphan receptors: very little is known about ligand-receptor interactions. This scarcity of knowledge is due to the fact that ORs are very difficult to express on the surface of heterologous cells (Mombaerts, 1999). Whilst several 7 transmembrane proteins, such as the opsins and the β-2-adrenergic receptor have been successfully expressed on cell surfaces, OR proteins tend not to be incorporated in the plasma membrane, instead they are retained, nonfunctional, in intracellular compartments (McClintock *et al.*, 1997).

Expression of 1 rat OR protein in insect Sf9 cells has been achieved using a baculovirus vector; some odorants caused transient increases in intracellular second messengers but others produced no observable response (Raming *et al.*, 1993). Zebrafish OR genes were also successfully expressed in human embryonic kidney (HEK293) cells by fusing a N-terminal membrane import

sequence of a guinea pig serotonin receptor and an artificial c-myc epitope tag to the ORs (Wellerdieck *et al.*, 1997). These cells showed transient increases in intracellular calcium (detected with the calcium-sensitive dye fura-2) when exposed to fish food, but there was no response to amino acids, bile acids or progesterone (physiologically relevant odorants for fish).

A more successful approach to detecting ligand-receptor interactions has been to use OSNs in order to achieve expression. Zhao *et al.* (1998), for example, used an adenoviral vector to drive expression of 1 OR and the green fluorescent protein (GFP) in a number of rat olfactory sensory neurons. Transepithelial potentials across the olfactory epithelium (electroolfactograms, EOGs) were measured in order to measure the response of the OR to 74 ligands, and it was found that octyl aldehyde raised the amplitude of the EOG above control levels. There was also a smaller response to 3 other aldehydes, heptyl aldehyde, nonyl aldehyde and decyl aldehyde. These responses were also measured in single-cell studies of the infected neurons.

Only one human OR gene has been functionally characterized: this gene, OR17-40, was expressed in HEK 293 cells and *Xenopus laevis* oocytes, and was found to be the ligand for helional and another structurally related molecule, heliotroplyacetone (Wetzel *et al.*, 1999).

1.10. The accessory olfactory system

The identification of genes coding for receptors in the olfactory epithelium was followed by the identification of more genes involved in olfaction. However, these genes were found in a functionally and anatomically distinct tissue, the vomeronasal organ (VNO). The VNO is located in a more anterior position within the nose than the MOE, and although it too is involved in

olfaction, it sends information via a separate pathway of neuronal projections. In rats and mice, the VNO is typically involved in stimulating innate behavioural responses upon the detection of pheromones, chemical signals conveying social and sexual information (Keverne, 1999). For example, experiments in rodents have linked the VNO to mating and aggressive behaviours in males (Clancy *et al.*, 1984, Meredith, 1986) and sexual development and onset of oestrus in females (Johns *et al.*, 1978, Reynolds and Keverne, 1979, Lomas and Keverne, 1982). In the VNO, work has revealed three families of receptors; Dulac and Axel (1995) identified seven novel seven transmembrane domain receptor sequences (named V1Rs), whilst a number of groups (Herrada and Dulac, 1997, Matsunami and Buck, 1997, Ryba and Tirindelli, 1997) identified another subfamily of pheromone receptors (V2Rs), and Pantages and Dulac (2000) found a third potential group (V3Rs). These three families are not closely related to each other, nor do they appear to be related to the family of OR genes. The V1R genes are predicted to be part of a family containing 30-100 genes, and expression of the family appears to be restricted to the apical cell layer within the VNO, whilst the V3R family (with approximately 100 members) are expressed in a distinct subset of VNO neurons. The V2R genes meanwhile, correspond to a family of about 100 genes that are related to the metabotropic glutamate receptors, the extra-cellular calcium-sensing receptor, and, more distantly to the GABA-B receptor. These similarities suggest the V2R proteins may function in a different way to the smaller OR, V1R and V3R proteins, with binding of ligands occurring in the large extracellular domain rather than somewhere within the 7 transmembrane structure. V2R genes appear to be expressed in a small fraction of vomeronasal sensory neurons (VSNs) in the basal region of the VNO. As with olfactory sensory neurons, expression of V1R, V2R, and V3R genes seems to be tightly controlled with only one or a few receptors expressed per VSN. The organization of expression of different classes of VRs in different regions within the VNO is also reflected within the accessory olfactory bulb (AOB) organisation: apical VSNs project to the rostral AOB whilst basal VSNs project to the caudal AOB.

The role of the accessory olfactory system in the human olfactory system is considered to be minimal. In contrast to rodents, there is only one generally accepted pheromone-like effect in humans, namely the synchronization of the menstrual cycle between women who live in close proximity to each other (McClintock, 1971), and whilst there is evidence that a VNO-like organ develops in the embryo, it is thought that the organ becomes vestigial in adults (Keverne, 1999). In addition, in rodents, one-third of V1R repertoire are predicted to be pseudogenes (Del Punta *et al.*, 2000), and two-thirds of the V2R gene family are predicted to be pseudogenes (Herrada and Dulac, 1997, Matsunami and Buck, 1997). Although functional examples of V1Rs (Rodriguez *et al.*, 2000) and V3Rs (Pantages and Dulac, 2000) have been found in the human genome, it has therefore been suggested that the majority of the VR genes found in the human genome are pseudogenes (Giorgi *et al.*, 2000).

1.11. The Major Histocompatibility Complex (MHC)

The Major Histocompatibility Complex (MHC) is an extended complex of gene clusters located on human chromosome 6p21.3 that is considered to contain a remarkably high number of genes with immunological function. The region is of major biomedical importance, owing to its role in tissue transplantation rejections and its role in influencing susceptibility to a variety of autoimmune diseases, such as insulin dependent diabetes mellitus, multiple sclerosis, systemic lupus erythmatosus and rheumatoid arthritis (Thomson, 1995). A complete map of the human MHC was published by the MHC sequencing consortium in 1999: at this time 224 gene loci (128 of which were predicted to be expressed) had been identified (Figure 1.6).

The MHC has traditionally been divided into three areas: class I, class II, and class III. The most telomeric region is the class I region which contains, amongst other genes, the classical class I

genes, HLA-A, HLA-B, and HLA-C. These classical class I genes encode the heavy (α) chain, which together with β2 microglobulin (β2m, locus found on chromosome 15) make up MHC class I molecules. These MHC class I molecules are expressed by virtually all nucleated cells within the human body, and they play a role in the immune system through their ability to bind peptides (antigens) and present these peptides to T-cells. Peptides loaded into class I molecules are generally derived from endogeneous source by the proteasome, of which LMP2 and LMP7 (both found within the MHC, (Driscoll *et al.*, 1993)) are subunits. These peptides are then transported in the endoplasmic reticulum by the TAP1/TAP2 molecule (also encoded by genes within the MHC, Ortmann *et al.*, 1994) where they are bound to the MHC class I molecules which proceed to the cell surface via the Golgi apparatus. At the cell surface the MHC class I molecule-peptide complex is accessible to T cells possessing the CD8 surface antigen. If a T cell is expressing a T cell receptor (TCR) that recognizes the specific peptide an immune response may be initiated: this generally results in the lysis of the cell expressing the peptide.

Class II genes include the LMP2, LMP7, TAP1 and TAP2 genes which have functions as detailed above. The MHC class II region, however, takes its name from the classical and non-classical MHC class II genes. The non-classical class II genes include HLA-DOA and HLA-DOB

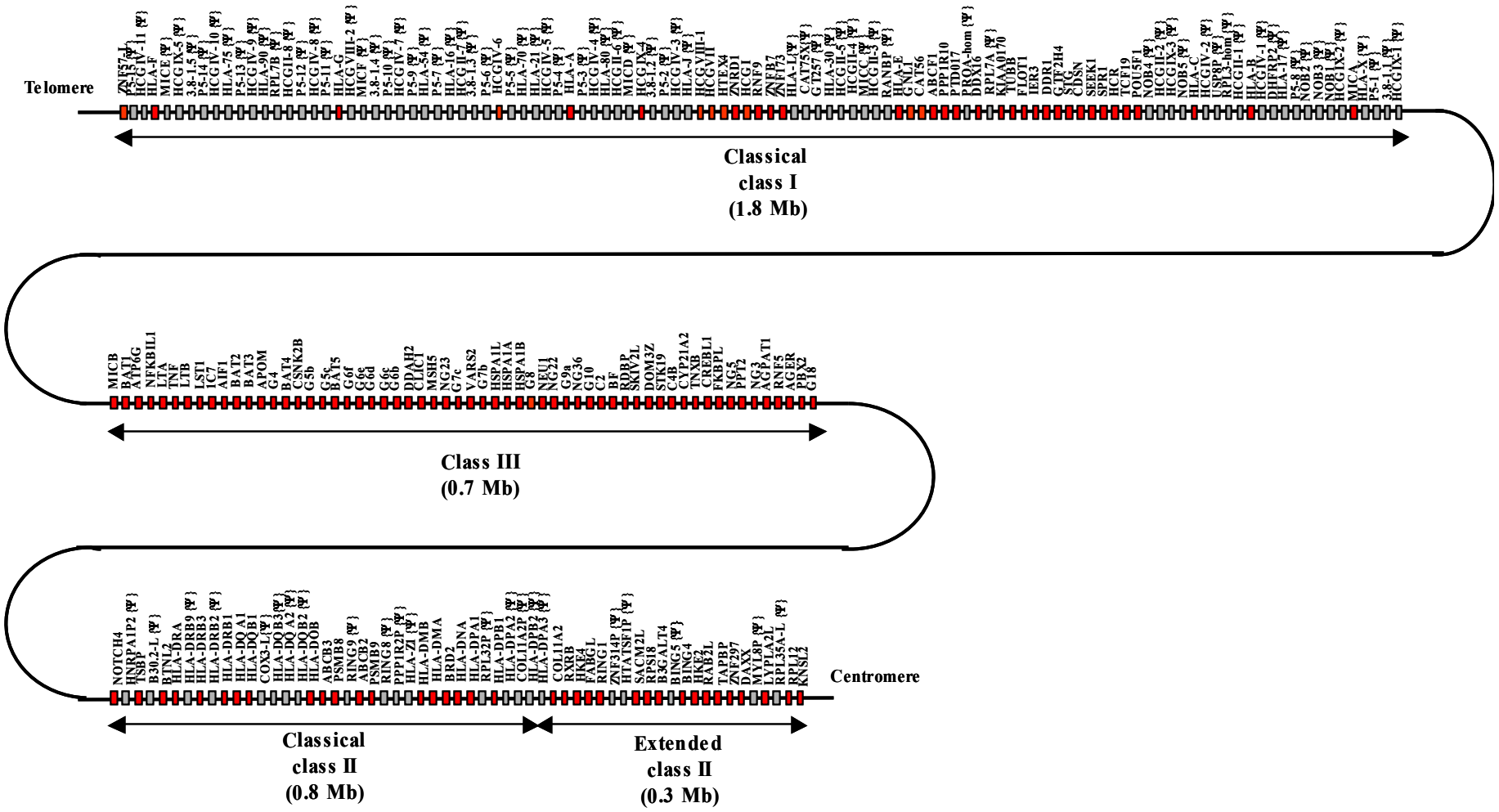


Figure 1.6: Gene map of the human MHC. (Adapted from The MHC Sequencing Consortium, 1999.) Expressed genes are highlighted in red, whilst pseudogenes are indicated by the Ψ symbol. The gene names are taken from the 1999 map: several have been changed subsequently by HUGO.

which produce protein products that combine to form the molecule HLA-DO. This molecule acts to suppress the ability HLA-DM has to facilitate peptide loading (Weber *et al.*, 1996). The classical class II genes (HLA-DP, HLA-DQ, HLA-DR) either encode proteins with 2 α chains (HLA-DPA, HLA-DQA, HLA-DRA) or proteins with 2 β chains (HLA-DPB, HLA-DQB, HLA-DRB): the corresponding α and β chains combine to form class II MHC molecules. MHC class II molecules differ from MHC class I molecules in that the groove of the peptide binding region (PBR) is open-ended, allowing longer peptides (generally 12-24 amino acids) to be bound. In contrast to MHC class I molecules, class II molecules are expressed only on a limited number of cell types, known as antigen presenting cells (APCs). These include B lymphocytes, macrophages, dendritic cells and activated T lymphocytes. Class II molecules bind predominantly to peptides from extracellular sources. Prior to peptide binding, these molecules are assembled in the endoplasmic reticulum (ER) with a membrane-bound chaperone protein (known as the MHC class II-associated invariant chain or γ chain) acting to stabilize the complex. This γ chain is degraded by proteases in the trans-Golgi reticulum, with the exception of a small fragment (the class II associated invariant chain peptide, CLIP) which is buried in the PBR. CLIP is only displaced just prior to binding: this reaction is catalysed by the product of 2 other non-classical class II genes, HLA-DM (Denzin and Cresswell, 1995, Sloan *et al.*, 1995). After binding the MHC class II molecule-peptide complex is transported to the cell surface where, if it is recognized by a specific T lymphocyte carrying the CD4 surface antigen, an adaptive immune response is triggered.

The MHC class III region is located between class I and class II. It is the most gene dense area of the MHC, containing a number of genes involved in the complement cascade of natural immunity, the interferon-inducible heat shock proteins, and a number of genes involved in the inflammation response.

Between species, there is a conservation of some of the basic genes within all 3 regions suggesting there is an evolutionary advantage in conserving the MHC as an unit. This MHC 'unit' can be observed in species evolving after the divergence of the jawless vertebrates (for example, hagfish or lamprey). The three regions of the human MHC appear to have been subject to different evolutionary mechanisms: whilst MHC class II and class III genes often appear to have direct orthologs, the MHC class I appears to have expanded and contracted in different species. This is discussed in more detail in the introduction to Chapter 5.

Work on the three 'classical' regions of the MHC revealed that sequence conservation and possibly linkage disequilibrium extended further than the three classical regions; immediate flanking regions were termed the extended class I and extended class II regions of the MHC (Stephens *et al.*, 1999). The extended class I region of the MHC is where Fan *et al.* (1995) located a cluster of OR genes; these genes were also found to be conserved in the syntenic region in mice and rats (Szpirer *et al.*, 1997). The conservation of this OR cluster within the MHC across three species suggested there may be some functional reason for this conservation.

1.12. MHC genetic diversity and reproductive selection

The most prominent hypothesis as to why an OR cluster appears to be conserved in its position next to the MHC across several species is the idea that there is some connection between MHC genetic diversity, reproductive selection, and olfaction. Products of genes within the MHC have critical roles during immune recognition, binding self and foreign peptide fragments for presentation to T lymphocytes (Babbitt *et al.*, 1985, Bjorkman *et al.*, 1987). The 'red queen' hypothesis, namely that pathogens will constantly evolve to defeat immune defences, means that the MHC has been forced to generate a huge amount of genetic diversity to be able to deal with the constant pathogenic onslaught. This genetic diversity can be generated through preferential

selection of heterozygotes over homozygotes (heterozygote-advantage or overdominance) or through preferential selection of relatively rare genotypes (negative frequency-dependent selection), or, more likely, through some combination of the two (Doherty and Zinkernagel, 1975, Hughes and Nei, 1988, Potts and Wakeland, 1990, Takahata and Nei, 1990, Slade and McCallum, 1992). In addition to these selection pressures generated through pathogen-driven selection, however, genetic diversity can also be generated through disassortative mating preferences (Potts and Wakeland, 1993). Offspring produced through disassortative mating choices are likely to be genetically fitter than other offspring since they will have a reduced inbreeding load and they will have an increased resistance to genetic disease arising from their increased MHC heterozygosity.

Within mice, experiments have shown that both inbred male mice and and outbred male mice preferentially mate with females with dissimilar MHC genotypes (Yamazaki *et al.*, 1976), and further experiments revealed that the detection of this dissimilarity is primarily through the mice distinguishing genotypic identity through smelling the odour of conspecific's urine (Yamaguchi *et al.*, 1981). Selective mating preference appears to be acquired through comparison of a potential mate's odour with remembered familial odours, since mice raised by foster parents will mate with other mice with an odour dissimilar to that of the family nest, rather than with mice dissimilar to themselves (Yamazaki *et al.*, 1988).

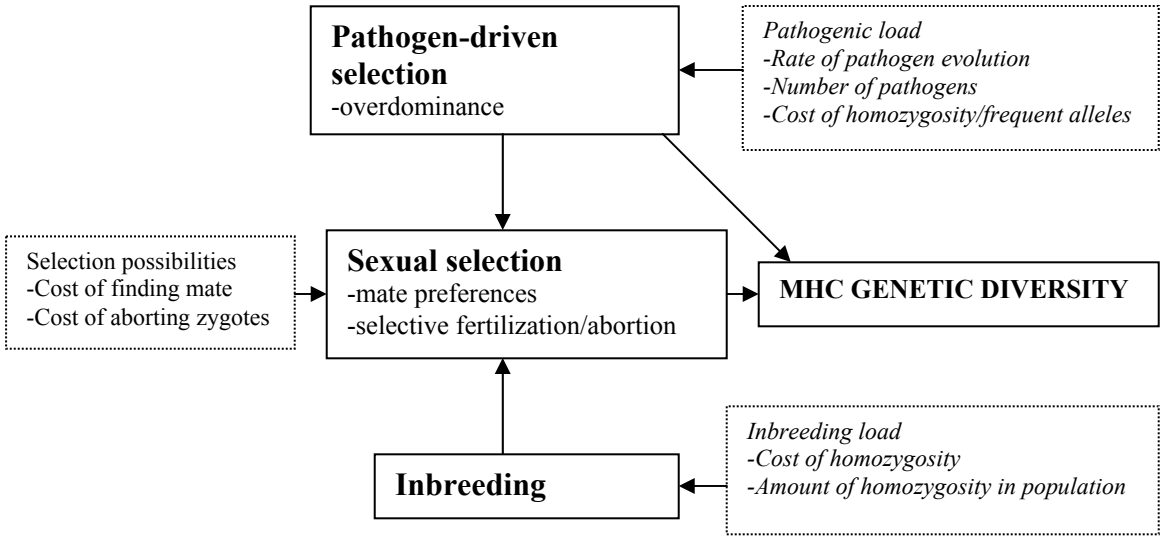


Figure 1.7: Factors contributing to MHC genetic diversity. Selective forces are shown in bold, with some of the variables shown in italics. Adapted from Potts and Wakeland (1993).

Another form of sexual selection has been observed in mice. In female mice, maintenance of pregnancy is dependent on the odour type to which the female mouse is exposed in early pregnancy. Female mice carrying an embryo with a similar MHC genotype will selectively abort the embryo if exposed to a dissimilar odour type (Yamazaki *et al.*, 1983). The exact contribution of this mechanism to the generation of genetic diversity is debatable, since there would appear to be a high cost involved in aborting a foetus, but nonetheless this is a clearly observed behaviour in laboratory mice.

Whether these mechanisms of sexual selection exist within other vertebrates in something that requires further work. Within humans, there is little conclusive evidence suggesting a link between MHC genotype and sexual selection (Tiwari and Terasaki, 1985), and there is contradictory evidence over whether similar MHC couplings produce a higher rate of abortions or have a lower fecundability than dissimilar couplings (Christiansen *et al.*, 1989, Ober *et al.*, 1992, Pennesi *et al.*, 1998).

The idea that there is a connection between sexual selection for dissimilar MHC genotypes and olfaction, therefore, makes the conservation the physical linkage between a cluster of OR genes and the MHC very interesting. It is possible that a combined MHC-olfactory haplotype is inherited. This combined haplotype could confer on an organism an increased sensitivity to detect (through olfaction) MHC alleles within their haplotype.

1.13. Aims of this thesis

The aims of this thesis were, firstly, to identify all OR genes located telomeric of the human MHC classical class I region. The target region was delineated by the HLA-F locus and the HFE locus: this region is known as the 'extended MHC class I' region in human. In addition to analyzing the sequence in the human extended class I region, another aim of this thesis was to map, sequence and identify the syntenic OR cluster region in mouse. Here, the target region was demarcated by the Gabbr1 locus and the breakpoint in synteny that occurs telomeric of this locus on mouse chromosome 17 (well centromeric of the HFE locus on human chromsome 6).

After identification of these genes in the human and mouse target regions, the syntenic regions were compared to identify orthologs and other conserved segments of sequence that could have functional roles. Analysis of the regulatory regions and expression profiles of the human MHC-linked ORs, using both experimental and *in silico* approaches was also performed. The MHC-linked OR genes are located just telomeric of the most highly polymorphic region of the human genome, the MHC, so polymorphism within these genes was also considered. A phylogenetic analysis comparing the MHC-linked ORs against other ORs within the human genome was also considered essential to reveal whether the MHC-linked ORs can be considered unique within the human genome. In addition to the MHC-linked OR genes, a number of pheromone receptor (VR)

pseudogenes were also found to be located telomeric of the MHC. These pseudogenes, which once formed part of the mammalian olfactory system, were analysed in order to investigate further the relationship between the pheromone and olfactory receptor genes.

These various strands of the thesis were expected to provide an unique insight into the function and evolution of the MHC-linked olfactory receptor genes. By combining data from these pillars of the thesis, a number of key issues were considered including:

- The relationship between the MHC, the 'extended MHC class I' region and the ORs. (Chapter 3 and Chapter 4)
- The syntenic relationships of the MHC-linked ORs in mouse and human (Chapter 5)
- The regulation of the MHC-linked ORs (Chapter 6)
- The polymorphisms of (some) MHC-linked ORs (Chapter 7)
- The relationship between MHC-linked ORs and other ORs in the human genome (Chapter 8)
- The relationship between the MHC-linked VR and OR genes (Chapter 9).