Sourcing the Code: Searching for the Evolutionary Origins of Cannabinoid Receptors, Vanilloid Receptors, and Anandamide

John M. McPartland
Patty Pruitt

ABSTRACT. Two cannabinoid (CB) receptors are known in humans, CB1 and CB2. They are phylogenetically ancient. Studies suggest CB receptors occur in mammals, birds, amphibians, fish, sea urchins, mollusks, leeches, and Hydra vulgaris. The CB receptor genes from some of these animals have been cloned and sequenced. These sequences were used to construct a phylogenetic tree of CB genes. The gene tree is rooted in an ancestral CB gene that predates the divergence of vertebrates and invertebrates. Thus the primordial CB receptor evolved at least 600 million years ago, a date broadly consistent with the Cambrian explosion. Since then, one clade of invertebrates, the Ecdysozoa, has secondarily lost the genes coding CB receptors. There is no evidence that animals obtained CB genes from other organisms via horizontal gene transfer. We hypothesize that the primordial CB receptor diverged from a related G-protein coupled receptor, and it linked with a pre-existing ligand, anandamide. Anandamide serves as a ligand for CB receptors as well as vanilloid (VR) receptors. VR receptors regulate the sensation of pain, and may also modulate mood and memory. Our phylogenetic analysis suggests that VR receptors evolved before CB receptors, so anandamide first served as a VR ligand. We speculate that CB receptors, lacking se-
lective constraints, subsequently acquired a mutation that coupled them with 2-AG. A better understanding of CB and VR receptors may enable us to combine their beneficial effects. Dual-signaling ligands such as anandamide have excellent therapeutic potential as analgesics, vasodilators, and anti-cancer agents.

**INTRODUCTION**

*Cannabis and Cannabinoid Receptors*

The capacity of cannabis to alter human consciousness was discovered at least 12,000 years ago (Abel 1980). More recently, Gaoni and Mechoulam (1964) isolated ∆⁹-tetrahydrocannabinol (Δ⁹-THC) as the primary psychoactive ingredient in cannabis. Because Δ⁹-THC is highly lipophilic, it can act as a solvent on cell membranes. Researchers initially thought it simply “sloshed” neurons in a very nonspecific manner, like alcohol. Following the discovery of opioid receptors, however, Devane et al. (1988) demonstrated that cannabinoids bind to a selective, high-affinity membrane receptor. The cannabinoid receptor (now termed CB₁) was cloned, and its DNA sequence uncoded (Matsuda et al. 1990). The human gene encoding CB₁, *CNR1*, is a nucleotide sequence 1755 base pairs (bp) in length, and translates into a protein consisting of 472 amino acids (reviewed by Felder and Glass 1998). The chain of amino acids winds into a series of seven transmembrane domains (α-helices), connected by alternating intra- and extracellular loops, terminating with an extracellular amino group and an intracellular carboxyl group (Figure 1). This serpentine topology is characteristic of all G-protein-coupled receptors (GPCRs), such as the receptors for endothelial sphingolipids (EDG-1), melanocortin, adenosine, some glutamate receptors, acetylcholine (muscarinic, but not nicotinic receptors), serotonin (all 5-HT classes except 5-HT₃), epinephrine (alpha- and beta-adrenergic receptors), GABA_B, dopamine, opioids, ACTH, CCK, VIP, FSH, LH, TSH, parathyroid hormone, calcitonin, glucagon, oxytocin, vasopressin, angiotensin II, and substance P.
A second CB receptor, termed CB₂, was discovered by Munro, Thomas, and Abu-Shaar (1993). The CB₂ gene, \textit{CNR2}, codes for a nucleotide sequence 1776 bp in length, and translates into a protein consisting of 360 amino acids. When the sequences of \textit{CNR1} and \textit{CNR2} are aligned for comparison, Munro, Thomas, and Abu-Shaar (1993) reported that they are identical at only 44% of their translated amino acid residues.

\textbf{CB Receptor Phylogenetis}

Evidence suggests that CB receptors are phylogenetically ancient, because homologs of human CB receptors are found in many other animals. A homolog is defined in biological systematics as a similar structure, behavior, or other trait shared by different species. Homologous traits permit us to make inferences about a series of events that happened in the past, known as evolution, which cannot be directly observed. The concept of “homologous series” was described by Vavilov (1922). It was Vavilov’s elucidation of homologous series that led to his breakthroughs in \textit{Cannabis} plant taxonomy (Vavilov 1926).

In the field of phylogenetics, homologs are divided into two groups: \textit{Orthologs} are homologous genes found in different organisms, derived by descent from a common ancestor. \textit{Paralogs} are homologous genes found in a given organism, derived by a gene duplication event. Murphy
et al. (2001) cloned and sequenced CNR1 orthologs from 62 species of placental mammals, sampled across all extant orders within that clade. CNR1 orthologs have been cloned and sequenced from earlier vertebrates, including the zebra finch, *Taeniopygia guttata* (Soderstrom and Johnson 2000), the newt salamander, *Taricha granulosa* (Soderstrom et al. 2000), and the puffer fish, *Fugu rubripes* (Yamaguchi, Macrae, and Brenner 1996). The puffer fish expressed a pair of paralogs, $F CB_1A$ and $F CB_1B$. CB$_2$ genes, which are paralogs of CB$_1$ genes, have not been well-studied in other animals. CNR2 orthologs have been cloned from rodents (*Rattus norvegicus, Mus musculus*), but are absent in puffer fish (Yamaguchi, Macrae, and Brenner 1996).

Invertebrates may also express CB receptors; Stefano, Salzet, and Salzet (1997) cloned and sequenced a CB$_1$ gene fragment from the leech, *Hirudo medicinalis*. Unfortunately this has been the only attempt to clone a CB gene from an invertebrate. Other invertebrates display evidence of CB receptors, although the evidence is based on non-molecular methods, such as radioligand binding studies. This data set comprises the sea urchins *Strongylocentrotus purpuratus* and *Paracentrotus lividus*, the leech *Theromyzon tessulatum*, the mollusk *Mytilus edulis*, and even the most primitive animal with a nerve network, the cnidarian *Hydra vulgaris* (review by Salzet et al. 2000).

Conversely, other invertebrates lack CB receptors, as evidenced by genome studies and radioligand binding studies. McPartland, Glass, and Mercer (2000) screened the entire genome of the fruit fly *Drosophila melanogaster*, which had been recently sequenced (Rubin et al. 2000), and found no genes with sequences resembling those of CNR1 and CNR2. Several low-identity sequences were located, but they exhibited crippling amino acid substitutions at critical residues known to confer CB receptor specificity. For example, in transmembrane helix 3 (Figure 2), CB$_1$ binding depends on a lysine residue at position 3.28 (Song and Bonner 1996) and a valine at 3.32 (Song et al. 1999); CB$_2$ receptors uniquely have key binding residues at methionine 3.34 (Chin et al. 1999), serine 3.31 and threonine 3.35 (Tao et al. 1999). The *D. melanogaster* sequences with closest identity to CNR1 and CNR2 had substitutions at all these positions.

Similarly, McPartland (2001) screened the entire genome of the nematode worm *Caenorhabditis elegans* and encountered similar crippling substitutions (Figure 2). These studies suggest the genes for CB receptors have been lost in *D. melanogaster* and *C. elegans*, or they mutated into unrecognizable pseudogenes. These negative results have been confirmed by radioligand binding studies of the insects *Apis mellifera*.
The human CB1 sequence in this region is identical to the CB1 orthologs cloned from monkey, rat, mouse, finch, newt, and fish CB1A. Fish CB1B contains four substitutions, and leech CB1 contains five substitutions; substitutions are printed in reverse (white on black). Drosophila (fruit fly) and C. elegans (nematode worm) sequences determined by BLAST as most identical to CB1 (McPartland et al. 2000c, 2001) are rife with substitutions. The human CB2 sequence contains eight substitutions in this region. Amino acid residues known to confer CB1 receptor specificity, such as the lysine at 3.28 and valine at 3.32, are in long boxes. CB2-specific serine (3.31), methionine (3.34), and threonine (3.35), are in short boxes.

(McPartland, Mercer, and Glass 2000), D. melanogaster (McPartland, Glass, and Mercer 2000), Gerris marginatus, Spodoptera frugiperda, and Zophobas atratus (McPartland et al. 2001). The apparent lack of an endocannabinoid system in insects opens many experimental possibilities. To better understand the role of CB receptors in health and disease, perhaps insects can serve as experimental animals akin to knockout mice (Di Marzo et al. 2000).

Sourcing CB Receptors from Horizontal Gene Transfer

The discovery of CB receptors led to an enigmatic question: why do animals have receptors for a cannabis compound? Pirozynski (1988) suggested that these kinds of “puzzling phenomena” could be caused by horizontal gene transfer (HGT). HGT is the nonsexual transmission of DNA between genomes of unrelated, reproductively isolated organisms (Rosewich and Kistler 2000). Scientists steeped in Darwinian theory tend to dismiss the significance of HGT, perhaps because of its
Lamarckian attributes, i.e., the acquisition of inheritable traits from the environment (in this case, from other organisms in the environment). Nevertheless, the sequencing of over 20 prokaryote genomes since 1995 has revealed the importance of HGT in the Archaea (e.g., *Methano-ococcus jannaschii*, *Archaeoglobus fulgidus*), and the Bacteria (e.g., *Haemophilus influenzae*, *Escherichia coli*, *Xylella fastidiosa*). Over 200 human genes may have been obtained from bacteria via HGT (International Human Genome Sequencing Consortium 2001), although this estimate is controversial (Saltzbeg et al. 2001).

McEno et al. (1991) proposed that humans acquired CB genes via HGT, from *Cannabis sativa*; McEno conjectured that Δ9-THC originally served as a ligand for CB receptors in the plant. HGT between distantly-related eukaryotes can be vectored by parasites capable of bridging both hosts. For example, the bacterium *Agrobacterium tumefaciens* is a potential vector, thanks to its extrachromosomal “Ti” plasmids, which are flawless gene conveyors. *A. tumefaciens* normally acts as a plant pathogen and it readily infects *Cannabis* spp. (McPartland, Clarke, and Watson 2000), but the pathogen also infects humans (Hulse, Johnson, and Ferrieri 1993). *A. tumefaciens* is capable of vectoring DNA into mammalian nuclei (Ziemienowicz et al. 1999), but the reverse also occurs. An ortholog of a human gene has been found in the bacterium (Whitehouse et al. 1998).

Fungi are potential HGT vectors; over a dozen fungal pathogens are known to infect both *Cannabis* spp. and humans (McPartland and Pruitt 1997). HGT among fungi may be quite common, and cases have been confirmed under experimental conditions (Rosewich and Kistler 2000). Incidences of HGT have been described between fungi and plants (e.g., the gene encoding taxol production, Stroble et al. 1996), and between fungi and animals (e.g., 6-hydroxynicotine oxidase, Schenk and Decker 1999).

**Endogenous Cannabinoid Ligands**

The HGT hypothesis lost some of its cachet when Devane et al. (1992) discovered an endogenous cannabinoid ligand that was produced in brain tissue, which they named anandamide (Figure 1). Since then, two other endogenous cannabinoid ligands have been found, sn-2 arachidonylglycerol (2-AG) (Mechoulam et al. 1995) and sn-2 arachidonoylglyceryl ether (2-AGE) (Hanus et al. 2001). These compounds are called “endocannabinoids” (DiMarzo and Fontana 1995), to differenti-
ate them from exogenous, plant-derived “phytocannabinoids” (Pate 1999). Endocannabinoids display a profile of biological activities similar to that of Δ9-THC, such as activation of CB receptors, inhibition of adenylate cyclase and calcium channels, hypothermia, analgesia, hypomobility, and catalepsy (Felder and Glass 1998). Endocannabinoids can explain why we have receptors that are sensitive to cannabis compounds; the plant ligands are simple mimics of our own, endogenous ligands.

The HGT hypothesis continues to have adherents. Ephick (1998) hypothesized a HGT mechanism to explain the existence of a CB gene in the leech. HGT may be implied by “puzzling phenomena” seen in other plants (other than Cannabis spp.). Soderstrom et al. (1998) discovered primitive plants (green algae) produce compounds that can bind to CB receptors. Tomato, soybean, and barley lipoxygenase enzymes can metabolize anandamide, a function that mammalian lipoxygenases cannot perform (van Zadelhoff, Veldink, and Vliegenhart 1998). Pine trees produce an analog of 2-AG, and the compound exhibits cannabinimetic activity (Nakane et al. 2000). These studies suggest many plants, presumably including Cannabis spp., can metabolize materials that have affinity for CB receptors. No one has looked to see if plants have the receptors.

**Parallels Between Cannabis spp. and Capsicum spp.**

The genus Cannabis evolved in central Asia. Presently the genus includes C. sativa L., C. indica Lamark, C. ruderalis Janischewsky, and C. afghanica Vavilov. About the same time that prehistoric humans discovered Cannabis spp., early Native Americans migrated across the Bering Strait land bridge, and moved into central America. There they encountered chili peppers, Capsicum spp., whose physiological effects were so desirable and habituating that chili pepper use permeated Mesoamerican culture for thousands of years. A physician to the fleet of Columbus brought chili peppers to Europe, and thereafter chili peppers spread worldwide. Capsicum spp. are now consumed daily by an estimated 25% of the world’s population (Szallasi and Blumberg 1999).

The pungent, burning sensation is caused by capsaicin, a vanilloid compound (Figure 1). Capsaicin-sensitive nerves are a subset of sensory neurons, nociceptors that give rise to small diameter, unmyelinated C fibers (rarely, A δ fibers). Because capsaicin is highly lipophilic, researchers initially believed capsaicin worked in a nonspecific manner,
by perturbing neural membrane lipids. But once again, radioligand studies demonstrated that capsaicin binds to a selective membrane receptor, and the receptor was subsequently cloned (Caterina et al. 1997). Caterina and colleagues called it the vanilloid receptor, VR1. The topology of VR1 differs significantly from that of CB receptors; VR1 is a cation channel (Figure 1). The chain of amino acids in VR receptors is longer (consisting of 839 residues), and winds into a series of six transmembrane domains (β-sheets). The loop between transmembrane regions 5 and 6 forms a membrane pore, and the amino terminal contains three ankyrin domains. Both terminals are Intracellular.

VR1 conveys information about a variety of noxious stimuli. It is activated by capsaicin and other compounds, by moderate heat (> 43°C), and perhaps by protons (tissue acidosis). VR1 activation triggers Ca\(^{2+}\) influx, causing a cascade of local inflammatory and vasodilatory reactions. Ca\(^{2+}\) influx also causes membrane depolarization, potentially generating an action potential. The signal propagates to dorsal horn ganglia, evoking the release of somatostatin, substance P, and calcitonin gene related peptide (CGRP) within the dorsal horn (Szallasi and Blumberg 1999).

With repeated exposure to these stimuli, VR1 receptors become desensitized. This phenomenon partially underlies the seemingly paradoxical use of capsaicin as an analgesic. The daily oral consumption of chili peppers by indigenous people may provide symptomatic relief for chronic caries and poor dentition. Synthetic capsaicin has been used to treat post-herpetic neuralgia (shingles), osteo- and rheumatoid arthritis, diabetic neuropathy, post-surgical pain, interstitial cystitis, vasomotor rhinitis, cluster headaches, many other forms of hyperalgesia and allodynia (Szallasi and Blumberg 1999).

The Cannabinoid and Vanilloid Connection

There is evidence of cross-talk between CB receptors and VR receptors. The recent discovery of VR1 receptors in many brain regions (e.g., preoptic area, locus ceruleus, medial hypothalamus, striatum) suggests that VR1 receptors may modulate emotions and memory (Szallasi and DiMarzo 2000). Indeed, VR1 and CB\(_1\) receptors may co-localize in the same neurons.

Significantly, anandamide acts as an agonist at VR1. No other endogenous ligands of VR1 have been discovered. Zygmunt et al. (1999) described anandamide as a partial VR1 agonist, whose affinity for VR1
nearly equals its affinity for CB1. Smart et al. (2000) described anandamide as a full agonist at VR1, but reported its binding affinity was 20 times less potent at VR1 than at CB1.

On the other side of the coin, olvanil, a synthetic ligand of VR1 (Figure 1), also serves as a CB1 ligand (Di Marzo et al. 1998). Olvanil binds tighter to CB1 (Ki = 1.6 µM) than does anandamide to CB1 (Ki = 1.9 µM). Di Marzo et al. (2001) subsequently synthesized arvanil, a “hybrid” molecule that grafted the vanillyl ring of capsaicin onto the C20:4 omega 6 fatty acid moiety of anandamide (Figure 1). Arvanil has four times more affinity than anandamide at CB1 (Ki = 0.25-0.52 and 1.9 µM, respectively), and three times more affinity than capsaicin at VR1 (Ki = 0.3 and 1.3 µM, respectively). This convolution of receptors and ligands of CB and VR has led to the suggestion that they might require unification under IUPHAR nomenclature (Szolcsányi 2000). Similar situations have arisen with glutamate, acetylcholine, GABA, and 5-HT receptors, all of which encompass metabotropic GCRPs as well as ionotropic channel receptors.

After Caterina et al. (1997) cloned and decoded VR1, they discovered a VR paralog (Caterina et al. 1999), the VR-like protein 1 (VRL-1). This receptor is not activated by vanilloid ligands or moderate heat; it responds to high temperatures (−52°C). Suzuki et al. (1999) subsequently cloned a VR paralog sensitive not to ligands or heat, but to mechanical pressure. They designated it the Stretch-Inhibitable Cation (SIC) channel. This gating mechanism is shared by the VR-related Osmotically Activated Channel (VR-OAC) (Liedtke et al. 2000). A flurry of VR-related receptors have been recently described, in some cases simultaneously by different labs, such as Liedtke et al. (2000) and Strotmann et al. (2000). Some VR receptors were described before their identity was recognized. The VRL-1 ortholog in the mouse was initially labeled a growth-factor-related channel (Kanzaki et al. 1999). A clade of VR-related receptors has been named Epithelial Ca2+ Channels (ECaC) (Hoenderop et al. 1999) and Ca2+ Transport channels (CaT1 and CaT2) (Peng et al. 1999). Most recently, Delany et al. (2001) identified VRL-2.

The functional coupling of CB and VR receptors is complex. Activation of CB1 by anandamide is antinociceptive; it reduces capsaicin-evoked release of CGRP from the dorsal horn. Activation of VR1 by anandamide does just the opposite, in the same neurons (reviewed by Szolcsánynl 2000). The fact that anandamide serves as a ligand for CB and VR receptors presents an evolutionary riddle: which receptor was the ligand’s original target? Which was the inaugural receptor?
Purpose of This Study

The present study has three aims. First, it will examine the similarity between human CB genes and their homologs in other animals. CB gene sequences vary from species to species, due to accumulated mutations. For example, the CB₁ gene from the rhesus monkey (Macaca mulatta) is 100% identical to the human CNR₁ sequence, whereas the partial CB₁ gene cloned from the leech (H. medicinalis) shares only 58% identity with CNR₁. Their percent identity is proportional to the evolutionary distances between them. The primordial ancestors of humans and leeches diverged at least 600 million years ago (Lee 1999), so CB genes in the two species had over half a billion years to accumulate differences. In contrast, the CB genes in humans and monkeys had only 10 million years to accumulate differences. These differences will be used to construct a gene tree of CNR₁, CNR₂, and their related paralogs and orthologs.

Second, the present study will search for evidence of HGT-mediated CB gene migration. We can test this hypothesis by conducting radioligand studies on plants, such as C. sativa, to see if plant tissues have specific binding sites for tritiated cannabinoids. Better yet, we can search for plant genes that resemble CNR₁ or CNR₂; and the first plant to have its entire genome sequenced, Arabidopsis thaliana, has recently become available (Arabidopsis Genome Initiative 2000). The entire genomes of over 20 species of Bacteria and Archaea will also be scanned, in a search for potential HGT vectors.

Third, the CB versus VR question will be addressed. A VR gene tree will be constructed, and then compared to the CB gene tree, by quantifying their respective sequence divergences. Since the degree of sequence divergence is correlated with evolutionary time, this analysis should estimate the relative ages of CB genes and VR genes. It is assumed that anandamide originally evolved as the ligand of the older receptor.

MATERIALS AND METHODS

Construction of the CB Receptor Tree

The deduced amino acid sequences of curated CNR₁ and CNR₂ were obtained from GenBank™ (National Center for Biotechnology Information, www.ncbi.nlm.nih.gov), accession numbers g.i. 4502927 and
They were compared to the deduced amino acid sequences of selected whole clones of CB gene homologs deposited in GenBank™, with the following accession (g.i.) numbers: rhesus monkey CB₁, 9664881; rat CB₁, 111475; mouse CB₁, 733425, finch CB₁, 8575561; newt CB₁, 8575561; puffer fish CB₁A, 2494952; puffer fish CB₁B, 2494952; rat CB₂, 10719923; and mouse CB₂, 7447152. The leech sequence was obtained from Stefano, Salzet, and Salzet (1997). Sequences were aligned using gapped BLAST (Basic Local Alignment Search Tool) version 2.0 (Altschul et al. 1997), also available on the Internet (www.ncbi.nlm.nih.gov/blast/).

Homologies were calculated as percent identity (identical amino acid residues), aligned over a designated length of amino acid residues. Protein homology is considered significant in the presence of at least 30% identity, aligned over a stretch of at least 80% of the sequence length (Rubin et al. 2000). Significant sequences identified by BLAST were considered orthologs if they had greater sequence identity to human CB genes than to any other sequences in that given organisms (Tatusov et al. 2000). BLAST 2.0 uses a “SEG program” as a default filter to eliminate low-complexity regions within sequences (i.e., amino acid repeats). This can confound BLAST searches with sequences that have low-complexity regions, such as the βχχβ repeat in CB₁ (Reggio et al. 2000). Thus, BLAST searches were run with the SEG filter off.

A gene tree of 12 CB gene homologs was assembled, its branching pattern based on the percentage sequence identity measured between CNR1 and its paralogs and orthologs (Feng and Doolittle 1996). The CB gene tree contained little data concerning invertebrates, however, so the tree was supported with supplemental data obtained from whole-genome studies and non-molecular data. This data strongly implies the presence or absence of CB receptors in other organisms. The following non-vertebrates and their non-molecular data were compiled in the supplemental data: H. vulgaris specifically binds the selective CB₁ antagonist [³H]SR141716A, produces anandamide, and exhibits FAAH activity (De Petrocellis et al. 1999). FAAH (fatty acid amide hydrolase) is the enzyme that degrades anandamide (Duetsch and Chin 1993). S. purpuratus binds the synthetic CB ligand [³H]CP55,940 (Chang et al. 1993). P. lividus produces anandamide and exhibits FAAH activity (Bisogno et al. 1997). M. edulis specifically binds [³H]anandamide (Stefano, Liu, and Goligorsky 1996), produces anandamide (Sepe et al. 1998), and exhibits FAAH activity (Stefano et al. 1998). H. medicinalis binds [³H]anandamide (Stefano, Salzet, and Salzet 1997) and produces...
anandamide (Matias et al. 2001). *T. tessulatum* binds \[^{3}H\]anandamide (Stefano, Salzet, and Salzet 1997).

Conversely, some non-molecular studies imply an absence of CB receptors in non-vertebrates: Brains dissected from *A. mellifera* showed no specific binding of \[^{3}H\]CP55,940 and \[^{3}H\]SR141716A, showed no activation of GTP \(\gamma S\) by \(\Delta^{9}\)-THC or the synthetic CB ligand HU210, and contained no measurable levels of anandamide (McPartland, Mercer, and Glass 2000). Heads and bodies of *D. melanogaster* did not bind \[^{3}H\]CP55,940 and \[^{3}H\]SR141716A, contained little or no anandamide, and did not express FAAH (McPartland et al. 2001). A panel of insects including *G. marginatus*, *S. frugiperda*, and *Z. atratus* showed no specific binding to \[^{3}H\]CP55,940 and \[^{3}H\]SR141716A (McPartland et al., 2001c). The organisms included in this supplemental data will be integrated into the CB gene tree, placed in positions determined by an aligned phylogenetic tree. The phylogenetic tree is based on current taxonomic models (“The Tree of Life,” <http://ag.arizona.edu/tree>; Aguinaldo et al. 1997; Adoutte et al. 2000), and is designed to mirror the CB gene tree, to aid in its interpretation. The CB gene tree and the phylogenetic tree cannot perfectly match, however. Incongruencies between single-gene trees and phylogenetic trees arise because of gene duplications, gene lineage sorting (deep coalescence), or HGT (Maddison 1997).

**Investigation of HGT**

*CNRI* and *CNR2* sequences were compared with all cDNA sequences of *C. sativa* deposited at GenBank™, and the entire genome of *A. thaliana* (Arabidopsis Genome Initiative 2000), as well as a non-redundant search of all prokaryotic (Archaea and Bacteria) sequences deposited at GenBank™. A search of fungal cDNA sequences, including the entire genome of *Saccharomyces cerevisiae*, was completed previously (McPartland, Mercer, and Glass 2000), and selected fungi were subjected to radioligand binding studies (McPartland and Glass 2002). Lastly, the human genome sequence was searched for degenerate CB genes (http://genome.ucsc.edu, http://www.ensembl.org), because degenerative mutations are often the fate of duplicate genes (Lynch and Conery 2000). Sequences were aligned with BLAST, as described previously.

**VR vs. CB Receptors**

The deduced amino acid sequences of human VR receptor genes were obtained from GenBank™, with the following accession (g.i.)
RESULTS

CB Receptor Phylogeny

A list of CNR1 homologs is presented in Table 1, ranked by their percentage identity to the CNR1 sequence, as measured by BLAST 2.0. The CNR1 orthologs from 60 other vertebrates are also deposited at GenBank™, but most of these are gene fragments. Orthologs of CNR2 have been relatively ignored; Genbank™ contained only two: rat CB2 gene (sharing 81% identity with human CNR2), and mouse CB2 gene (sharing 82% identity with CNR2). The rat and mouse CB2 sequences share 93% identity with each other; rat and mouse CB1 sequences share 97% identity.

A CB gene tree, based on the percentage sequence identity between CNR1 and its homologs is presented in Figure 3. Vertebrates are over-represented in the CB gene tree, coupled with a dearth of non-vertebrate gene sequences. Consequently the gene tree was supported with supplemental data that strongly implies the presence or absence of CB receptors in 13 other organisms. Their placement in the CB gene tree was guided by a juxtaposed “Tree of Life.”

The CB gene tree is rooted in an ancestral CB gene. The first bifurcation of the tree represents the deep divergence between the CNR1 and CNR2 sequences, which share only 47% identity with each other, as measured by BLAST. This gene duplication event gave rise to separate paralogous lineages, the CNR1 orthologs and CNR2 orthologs. After
the divergence of CNRI and CNR2, a second gene duplication event gave rise to the puffer fish paralogs, CB1A and CB1B. A BLAST search of the human genome did not identify any degenerate CB receptor paralogs (psuedogenes) in Homo sapiens, suggesting that a similar gene duplication did not occur in humans, or the duplicate gene in humans subsequently mutated beyond recognition.

**Searching for Evidence of HGT**

Screening all cDNA sequences of C. sativa deposited at GenBank™, as well as the entire genome of A. thaliana, did not reveal any orthologs of human CB receptors. The A. thaliana sequence with best BLAST alignment, Mre11 protein (g.i. 5524769), exhibited only 28% identity, over a stretch of 83 amino acids (a mere 18% of the CNR1 sequence). A non-redundant search of all Archaea and Bacteria sequences deposited at GenBank™ did not disclose any gene products with significant identity to human CB genes.

**VR Receptor Phylogeny**

BLAST aligned human VR1 and a rat ortholog, sharing 85% identity. The SIC sequence was closely related, sharing 82% identity with VR1. The clade of VR-OAC receptors collectively shared 50% identity

---

**TABLE 1. Homologues of human CB₁ receptors, with percent identity calculated with BLAST 2.0 algorithm.**

<table>
<thead>
<tr>
<th>Species</th>
<th>Percent identity with human CB₁ gene sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monkey (Macaca mulatta) CB₁</td>
<td>100% of 472 amino acids</td>
</tr>
<tr>
<td>Rat (Rattus norvegicus) CB₁</td>
<td>97% of 473 amino acids</td>
</tr>
<tr>
<td>Mouse (Mus musculus) CB₁</td>
<td>97% of 473 amino acids</td>
</tr>
<tr>
<td>Finch (Taeniopygia guttata) CB₁</td>
<td>91% of 473 amino acids</td>
</tr>
<tr>
<td>Newt (Taricha granulosa) CB₁</td>
<td>83% of 473 amino acids</td>
</tr>
<tr>
<td>Puffer fish (Fugu rubripes) CB₁</td>
<td>72% of 468 amino acids</td>
</tr>
<tr>
<td>Puffer fish (Fugu rubripes) CB₁</td>
<td>59% of 470 amino acids</td>
</tr>
<tr>
<td>Leech (Hirudo medicinalis)</td>
<td>58% of 153 amino acids</td>
</tr>
<tr>
<td>Human CB₂</td>
<td>47% of 360 amino acids</td>
</tr>
</tbody>
</table>
with VR1; within the clade, rat VR-OAC shared 97% with the mouse ortholog and 95% with the human ortholog. VRL-2 shared more identity with VR1 (49%) then did the VRL-1 clade (collectively about 42% identical to VR1). Within the VR-1 clade, rat shared 92% with mouse and 79% with human. The greatest divergence was seen in with ECaC and CaT1, both clades shared 30% identity with VR1. The ECaC clade and the CaT1 clade shared 73% identity with each other. The ECaC

FIGURE 3. The CB receptor gene tree with supplemental data (A), mirrored by a phylogenetic tree of major taxonomic clades (B).

A. CB receptor gene tree, based on similarities between 12 cloned sequences; supplemental data includes 15 other organisms with evidence suggesting they express CB receptors (●) or they do not (○).
B. Phylogenetic tree of life, emphasizing the Deuterostomes (●), which are over-represented in the mirrored gene tree), the Lophotrochozans (○), and the Ecdysozoans (●). Gene duplications events are marked (+).
orthologs were 83% identical and the CaT1 orthologs were 89% identical. The similarity between these sequences and VR1 was used to construct a VR gene tree, illustrated in Figure 4. The VR gene family exhibits deeper divergences than the CB gene tree, at several levels.

DISCUSSION

CB Receptor Phylogenetics

The CB gene tree (Figure 3) suggests an ancient CB gene underwent a duplication event, giving rise to present-day CNR1 and CNR2. The duplication event must have occurred prior to the divergence of vertebrates and invertebrates, because at least one product of the duplication, the CB1 gene, has orthologs in both vertebrates (fish, amphibians, birds, mammals) and invertebrates (leech). The divergence between CNR1 and CNR2 (47% identity) is greater than that between CNR1 and the leech CB1 gene (58% identity), suggesting the duplication event is older than the leech CB1 gene. The study of CB2 genes in other animals needs further attention. The antiquity of the duplication event is supported by the wide separation of CNR1 and CNR2 in the human genome, on chromosomes 6q14-15 and 1p35-36, respectively.
The CNR1-CNR2 sequence divergence corresponds to a dissimilarity in physiology. Whereas CB$_1$ receptors are primarily expressed by cells in the central nervous system, CB$_2$ receptors are located in immune cells (B-cells, monocytes, T-cells, etc.) and immune tissues (tonsils, spleen, etc.). This ramification enables the cannabinoid signaling system to span the psycho-neuro-immune axis (aka, “the mind-body connection”). CB$_1$ and CB$_2$ have also diverged in their pharmacology. CB$_1$, for instance, binds anandamide with four-fold greater affinity than does CB$_2$; CB$_2$ binds $\Delta^9$-THC with 16-fold greater affinity than does CB$_1$ (reviewed by Felder and Glass 1998). Some synthetic cannabinoids bind to CB$_1$, with nearly no affinity for CB$_2$, on the order of 1000-fold selectivity (Di Marzo et al. 2001).

Because CB receptors are present in vertebrates as well as very primitive metazoans (*Hydra vulgaris*), the CB gene must have evolved prior to the divergence of these organisms’ ancestors, which happened at least 600 million years ago (Lee 1999). This dates to the earliest multicellular animals, which were experiencing a rapid evolutionary burst in concert with their new needs for cell-to-cell communications. Phylogenetic studies indicate many neuroreceptor GPCRs appeared during this evolutionary horizon, in parallel with the “Cambrian explosion” of metazoan animals documented in the fossil record (Peroutka and Howell 1994; Xue 1998).

Yet, Figure 3 illustrates a lack of CB receptors in some animals. The paucity of data concerning non-vertebrates makes it difficult to discern the broader taxonomic boundaries between the “haves and the have-nots.” Salzet et al. (2000) concluded that CB receptors are conserved in animals from Cnidarians (*Hydra*) to humans (*Homo*). Elphick and Egertová (2001) concluded that CB receptors evolved after Deuteroostomes (e.g., vertebrates and echinoderms) diverged from Protostomes (annelids, molluscs, nematodes, insects, cnidarians, and poriferans). Their conclusion required the rejection of many previous studies: Elphick (1998) dismissed the leech CB$_1$ gene, characterizing it as a primordial CB/melanocortin hybrid, and not a functional CB$_1$ receptor. The CB gene tree (Figure 3) disputes his hypothesis, because the leech gene evolved after the divergence of CB$_1$ and CB$_2$ genes, so it cannot be a primordial predecessor. The leech sequence clearly segregates to the CB$_1$ lineage. This can be confirmed by close inspection of the leech sequence in transmembrane helix 3 (Figure 2). The leech sequence has a lysine residue at position 3.28, which is critical for CB$_1$ function (Song and Bonner 1996), but the sequence has substitutions at CB$_2$-specific
sites, such as methionine 3.34 (Chin et al. 1999), serine 3.31, and threonine 3.35 (Tao et al. 1999). The leech receptor’s functionality has also been confirmed by radioligand binding studies (Stefano, Salzet, and Salzet 1997).

Subsequently, Elphick and Egertová (2001) dismissed the leech CB1 gene as a possible artifact arising from DNA contamination. They disregarded the *H. vulgaris* binding studies by De Petrocellis et al. (1999) because the stereoselectivity of this binding site was not tested. They rejected the *M. edulis* and *H. medicinalis* binding studies (Stefano, Liu, and Goligorsky 1996; Stefano, Salzet, and Salzet 1997) as “non-standard” because they used tritiated anandamide.

Alternatively, all the aforementioned studies can be accepted. From this perspective, McPartland, Glass, and Mercer (2000) concluded that CB receptors were present in Deuterostomes and *some* Protostomes. The lack of CB receptors in insects was interpreted as a sorting event that occurred in the course of insect evolution. They hypothesized that insects secondarily lost CB receptors because of a lack of ligand; anandamide is a metabolite of arachidonic acid, and insects produce little or no arachidonic acid in their tissues, in contrast to Deuterostomes and most other invertebrates.

The addition of nematode genome data (McPartland 2001), however, permits a more elegant interpretation of the CB gene tree, based on new animal taxonomy: Aguinaldo et al. (1997) proposed that Protostomes diverged into two clades. The Lophotrochozoa (lophophore-bearing animals with trocophore larvae) include the annelids, molluscs, platyhelminthes, and rotifers. The Ecdysozoa (animals that undergo molting during their life cycle) include the nematodes, arthropods (insects and crustaceans), onychophorans, and priapalids.

Thus, Figure 2 suggests that CB receptors may be absent in the Ecdysozoa, but retained in other invertebrates (the “higher” echinoderms and lophotrochozoans, and the “lower” cnidarians). Why CB receptors were secondarily lost in a clade of molting animals is open to conjecture. The mechanism driving this sorting event may be due to phospholipid biochemistry, or due to cellular modifications associated with molting, such as the loss of locomotory ectodermal cilia. Indeed, a recent study of β-thymosin orthologs (Manuel et al. 2000) demonstrated that the conserved version of these actin-binding polypeptides was absent in Ecdysozoan organisms (*D. melanogaster* and *C. elegans*), but present in bookend clades, including the Deuterostomes (sea ur-
chins and vertebrates), Lophotrochozoans (leeches and mussels), and the lower Poriferans (sponges).

A few reports in the literature conflict with the Ecdysozoa hypothesis. Egertová, Cravatt, and Elphick (1998) reported 5% specific binding of $[^3]H$CP55,940 in muscles of the locust, *Schistocerca gregaria*, but they questioned their own findings. Howlett et al. (2000) detected specific binding of $[^3]H$CP55,940 in *D. melanogaster* heads, but the binding was not displaced by CB$_1$-specific SR141716A or CB$_2$-specific SR144528. Previously, Howelett et al. (1990) reported no binding of $[^3]H$CP55,940 in a mollusc (*Aplysia californica*) and a vertebrate (a lamprey reported as *Ichthyomyzon intercostus* but probably *I. unicuspis*).

The literature contains another level of non-molecular evidence that we did not include in the supplemental data, that of pharmacological studies. Many researchers have reported changes in organisms after giving them Δ$_9$-THC; sometimes this data is used to infer the presence of CB receptors in the affected organisms. This inference may not be true, because Δ$_9$-THC causes many non-receptor effects (reviewed by McPartland and Russo 2001). Hence, pharmacological studies can only hint at the presence of receptors.

For instance, pharmacological studies have demonstrated that Δ$_9$-THC is antifungal; it inhibited the growth of *S. cerevisiae* (El Sohly et al. 1982) and *Phomopsis ganjae* (McPartland 1984). But *P. ganjae* shows no specific binding with $[^3]H$CP55,940 and $[^3]H$SR141716A, and BLASTing the entire genome of *S. cerevisiae* found no CB gene orthologs (McPartland and Glass 2001). Accordingly, the antifungal effects of Δ$_9$-THC are not mediated by CB receptors (the mechanism may be Δ$_9$-THC stimulation of phospholipase A$_2$ or inhibition of cytochrome P$_{450}$ enzymes, both of which are non-CB receptor effects).

Some pharmacological studies contradict other lines of evidence. Acosta-Urquidi and Chase (1975) exposed *A. californica* to Δ$_9$-THC, which produced a change in the slug’s nerve action potentials. This hints at a CB receptor-mediated effect, and it agrees with DiMarzo et al. (1999) who detected 2-AG and FAAH-like activity in *Aplysia*, but it conflicts with the negative binding studies reported by Howlett et al. (1990).

Two pharmacological studies are particularly evocative in their support of the Ecdysozoa hypothesis: Rothschild and Fairbairn (1980) demonstrated behavioral changes in moths (*Pieris brassicae*) exposed to Δ$_9$-THC. Nearly identical behavior, however, was aroused by cannabidiol (CBD), a ligand with little affinity for CB receptors. This sug-
gests that the behavioral changes were not mediated by CB receptors. Nevertheless, $\Delta^9$-THC and CBD must activate moth olfactory receptors. Waser (1999) fed $[^3H]\Delta^9$-THC to ants (*Formica pratensis*); their brain tissues accumulated the tritiated material, but the ants showed no significant changes in behavior. In contrast, ants fed $[^3H]LSD$ were severely altered; this is because $[^3H]LSD$ binds to well-known serotonin and dopamine receptors in insect brains (Blenau, May, and Erber 1995).

The Ecdysozoa hypothesis requires further testing. The best evidence for these experiments would be genetic cloning studies. Radioligand binding studies are subject to false negative results if receptor levels are low, especially against a high noise background (i.e., high levels of non-specific binding). When radioligand studies are positive, however, they have fine predictive value for the presence of CB receptors. The same cannot be said for the extraction of endocannabinoids from animals. McPartland et al. (2001) extracted 2-AG from neural tissues of *A. mellifera* and *D. melanogaster*, even though overwhelming evidence suggests these organisms lack CB receptors! This conundrum was clarified by Hoyle (1999), who demonstrated that there is greater evolutionary pressure to conserve receptor ligands than to conserve the neuroreceptors themselves. Endocannabinoids in insects may protect them from predators who do have CB receptors. The defense glands of an aquatic beetle, *Agabus affinis*, were recently shown to contain 2-AG (Schaaf and Dettner 2000). These glands discharge when the beetle is seized by a fish. In a feeding assay with minnows (*Phoxinus phoxinus*), spiking pellets with 100 µg of 2-AG deterred pellet consumption (Schaaf and Dettner 2000). Similarly, parasitic Ecdysozoans may secrete 2-AG in order to blunt the immune reactions of their hosts. In these cases, the presence of endocannabinoids may be a case of convergent evolution, homoplasy rather than homology.

**Searching for Evidence of HGT**

Our inability to find CB homologs among *C. sativa* genes is not surprising, because little of this plant’s genome has been deposited at GenBank™. The absence of CB homologs in the entire genome of *A. thaliana*, however, is telling. Although the genera *Cannabis* and *Arabidopsis* belong to different plant families, evidence suggests that the Plant Kingdom displays considerable synteny (conservation of gene order). All of the deposited *C. sativa* genes have orthologs in the *A. thaliana* genome (pairwise BLAST searches, data not shown). The gene encoding
Δ⁹-THC synthesis (when it is found), may also have an ortholog in the *A. thaliana* genome; *A. thaliana* has genes that code for the production of alkaloids and phytoalexins that *A. thaliana* is not known to synthesize (Arabidopsis Genome Initiative 2000). The *A. thaliana* genome is full of surprises, including genes obtained from bacteria via HGT, and the orthologs of dozens of human disease genes, such as Niemann-Pick, Wilson, breast cancer, cystic fibrosis, and hyperinsulinism (Arabidopsis Genome Initiative 2000). None of these human orthologs, however, codes for a GPCR protein. Only 27 *A. thaliana* genes code for proteins that resemble GPCRs (Arabidopsis Genome Initiative 2000), so the lack of CB gene orthologs in *A. thaliana* is reasonable.

GPCRs are similarly rare in Prokaryote genomes, so we should not be surprised by the lack of CB homologs in these potential HGT gene vectors. Fungi, however, may code for dozens of GPCRs, and the ligands signaling these receptors have yet to be identified (Bölker 1998). A previous study, however, determined that none of the GPCR-related genes in *Saccharomyces cerevisiae* code for CB orthologs (McPartland and Glass 2001). Tritiated cannabinoid ligand binding studies on a panel of fungi were also negative. Taken together, these results suggest that CB receptors evolved in primitive animals, and did not radiate via HGT from fungi, plants, or prokaryotes.

**Anandamide as a Ligand for VR Receptors vs. CB Receptors**

Comparing the CB gene tree (Figure 3) with the VR gene tree (Figure 4) illustrates deeper divergences in the latter. For example, human and rat orthologs of CB₁ share 97% identity, whereas human and rat orthologs of VR₁ share only 85% identity. The VR gene tree has diverged into six major branches, while the CB gene tree has only two: CB₁ and CB₂. The lowest branch of the CB tree has 47% similarity, whereas the lowest branch of the VR tree has 30% similarity, again indicative of deeper divergence. The deeper sequence divergences reflect deeper physiological divergences. CB₁ and CB₂ still recognize each other’s ligands (although their relative affinities have diverged), whereas the VR homologs have widely diverged in their gating mechanisms.

Since the degree of divergence is correlated with evolutionary time, this analysis suggests the primordial VR receptor predated the primordial CB receptor. We therefore infer that anandamide originally evolved as the ligand of the older receptor. This analysis is speculative, because it is based on two assumptions. First, it is based on the neutrality theory
of molecular evolution, which predicts that the rate of genetic divergence will be constant across time and different species (Kimura 1986). Neutrality theory, like HGT theory, is at odds with orthodox Darwinians, who maintain that evolutionary change at the molecular level is due entirely to natural selection. Our second assumption is that CB genes and VR genes evolved at similar rates. In other words, they pass a “relative-rate” test used to calibrate the molecular clock (International Human Genome Sequencing Consortium 2001).

Our proposal that anandamide originally evolved as a VR1 agonist agrees with data reported by Szallasi and DiMarzo (2000), who noted that regions of the brain with high levels of anandamide correlate with the regional expression of VR1. To wit, some anandamide-rich areas, such as the brainstem, have correspondingly few CB1 receptors (Szallasi and DiMarzo 2000), suggesting that the primary target of anandamide in these regions may be VR1 receptors (Di Marzo et al. 2000).

From another perspective, Sugiura et al. (1999) also argued that CB1 was not the original receptor for anandamide. Instead, they presented evidence that CB1 was originally a 2-AG receptor, based on binding studies and ligand extraction studies. Gonsiorek et al. (2000) presented similar data for CB2.

Arguments contrary to our proposal focus on the fact that anandamide has less affinity for VR1 than it does for CB1 (Smart et al. 2000). We interpret this as evidence that the receptors are continuing to evolve. VR1 may be evolving away from the ligand and towards a temperature-gated mechanism. Indeed, a splice variant of VR1 was recently described (Schumacher et al. 2000), and it completely lost its ability to bind capsaicin. Splice variants are alternative ways in which a gene’s protein-coding sections (exons) are joined together to create a messenger RNA molecule and its translated protein.

CB genes also continue to evolve. Shire et al. (1995) described a CNR1 splice variant. Tsai, Wang, and Hong (2000) described a CNR1 microsatellite polymorphism. Microsatellites are mutated DNA loci that contain nucleotide repeats; the CB1 microsatellite is an AAT triplet repeat. Gadzicki, Muller-Vahl, and Stuhrmann (1999) described a CB gene with a single nucleotide polymorphism (SNP, pronounced “snip”). A SNP is a point mutation in the DNA sequence. There are a lot of them. The SNP Consortium (http://snp.cshl.org), has identified 1.42 million SNPs in the human genome (International Human Genome Sequencing Consortium 2001), including over a dozen SNPs of CNR1 and CNR2.
Concluding Remarks

Gene duplication events, splice variants, and SNPs are the most common mechanisms generating the evolution of new genes (Lynch and Conery 2000). Results presented in this paper suggest these mechanisms, and not HGT, generated the genes for CB receptors.

The CB gene tree traces the origin of CB receptors back at least 600 million years. The primordial CB gene probably diverged from a closely-related GPCR, such as EDG-1. These GPCRs are gated by ligands derived from fatty acids. They evolved from older GPCRs gated by biogenic amines, which first appeared when plants and animals diverged, about 1200 million years ago (Peroutka and Howell 1994). All GPCRs may be predated, however, by the ionotropic glutamate receptor (iGluR) (Chiu et al. 1999). The iGluRs are ligand-gated ion channels, related to VR1. VR1 belongs to the TRP family of ion channels, whose ancestors can be found in *D. melanogaster* and *C. elegans*, and some of these receptors are activated by arachidonic acid, the precursor of anandamide (Harteneck, Plant, and Schultz 2000).

Unlike the evolutionary fate awaiting most new genes, the primordial CB gene survived. Our results suggest the CB gene survived because it linked with a pre-existing VR1 ligand, anandamide. Duplicate receptors are powerful sources of biological novelty, because the second receptor is not under constraints to maintain its original ligand and can accept mutations (Baker 1997). The primorial CB receptor, lacking selective constraints, eventually acquired a mutation that coupled it with 2-AG. This new receptor-ligand couplet gained novel functions, which apparently were advantageous. New receptors activated by new ligands often become fixed and stabilized by the selective forces of evolution (Goh et al. 2000). Nevertheless, one clade of animals, the Ecdysozoans, has secondarily lost the genes coding CB receptors. Investigating the physiology of these animals lacking an endocannabinoid system will shed light on this system’s role in our own physiology.

Finally, a better understanding of CB and VR receptors may enable us to combine the beneficial effects of *Cannabis* spp. and *Capsicum* spp. The synthetic CB-VR “hybrid” ligand, arvanil, has excellent therapeutic potential as an analgesic, a vasodilator, and as a potent anti-proliferative agent against human breast cancer and prostate cancer (Di Marzo et al. 2001). It is intriguing to regard whole cannabis as a CB-VR “hybrid,” because the plant contains eugenol and guaiacol (McPartland and Russo 2001). Eugenol and guaiacol are capsaicin congeners used as dental analgesics, and they may activate VR1 re-
ceptors (Ohkubo and Shibata 1997). We need to elucidate the close association of these compounds and their multiplicity of neuroreceptor targets. The endocannabinoid system, like the "high" it can engender, is not a linear business.

REFERENCES

Chin, C. N., J. W. Murphy, J. W. Huffman, and D. A. Kendall. 1999. The third transmembrane helix of the cannabinoid receptor plays a role in the selectivity of aminoalkylindoles for CB$_2$, peripheral cannabinoid receptor. *J Pharmac Exper Therap* 291:837-44.


Song, Z. H., C. A. Slowey, D. P. Hurst, and P. H. Reggio. 1999. The difference between the CB(1) and CB(2) cannabinoid receptors at position 5.46 is crucial for the selectivity of WIN55212-2 for CB(2). *Mol Pharmacol* 56:834-40.


RECEIVED: 05/15/2001
ACCEPTED IN FINAL FORM: 08/04/2001