What causes Alzheimer’s disease?

Richard A. Armstrong
Vision Sciences, Aston University, Birmingham, UK

Abstract
Since the earliest descriptions of Alzheimer’s disease (AD), many theories have been advanced as to its cause. These include: (1) exacerbation of aging, (2) degeneration of anatomical pathways, including the cholinergic and cortico-cortical pathways, (3) an environmental factor such as exposure to aluminium, head injury, or malnutrition, (4) genetic factors including mutations of amyloid precursor protein (APP) and presenilin (PSEN) genes and allelic variation in apolipoprotein E (Apo E), (5) mitochondrial dysfunction, (6) a compromised blood brain barrier, (7) immune system dysfunction, and (8) infectious agents. This review discusses the evidence for and against each of these theories and concludes that AD is a multifactorial disorder in which genetic and environmental risk factors interact to increase the rate of normal aging (’allostatic load’). The consequent degeneration of neurons and blood vessels results in the formation of abnormally aggregated ‘reactive’ proteins such as β-amyloid (Aβ) and tau. Gene mutations influence the outcome of age-related neuronal degeneration to cause early onset familial AD (EO-FAD). Where gene mutations are absent and a combination of risk factors present, Aβ and tau only slowly accumulate not overwhelming cellular protection systems until later in life causing late-onset sporadic AD (LO-SAD). Aβ and tau spread through the brain via cell to cell transfer along anatomical pathways, variation in the pathways of spread leading to the disease heterogeneity characteristic of AD.

Key words: Alzheimer’s disease (AD), aetiology, aging, genetic factors, β-amyloid (Aβ), tau.

Introduction
Ever since the first description of pre-senile dementia by Alois Alzheimer in 1907 [5], the presence of cognitive impairment together with the formation of senile plaques (SP) and neurofibrillary tangles (NFT) have been regarded as the defining clinicopathological features of Alzheimer’s disease (AD) [13,103,126]. The term AD was first used in 1910 by Kraepelin based on the clinical and pathological description of Alzheimer’s original cases. Of the two cases described by Alzheimer, however, both had numerous SP, but only one had significant numbers of NFT [76], thus identifying the pathological heterogeneity characteristic of AD [4], and creating difficulties in establishing a single theory as to its cause.

Many theories as to the cause of AD have consequently been proposed. It is not the intention to discuss every theory but to concentrate on those most likely to be involved. Hence, the theories are discussed in eight categories: (1) acceleration of aging, (2) degeneration of anatomical pathways, including the cholinergic and cortico-cortical pathways, (3) environmental factors such as exposure to aluminium,
head injury, and malnutrition, (4) genetic factors including mutations of amyloid precursor protein (APP) and presenilin (PSEN) genes, and allelic variation in apolipoprotein E (Apo E), (5) a metabolic disorder resulting from mitochondrial dysfunction, (6) vascular factors such as a compromised blood brain barrier, (7) immune system dysfunction, and (8) infectious agents. This review discusses the evidence for and against each of these hypotheses and develops a general theory as to the cause of AD. The implications of this theory are considered with reference to future research and treatment of AD.

Theories based on aging

That AD may be an accelerated form of natural aging is based on the observation that the many pathological changes in AD are similar to those present in normal aging apart from their severity [156]. Hence, in cognitively normal brain, there is an age-related reduction in brain volume and weight, enlargement of ventricles, and loss of synapses and dendrites in selected areas [94]. Accompanying these changes are the characteristic pathological features of AD, including SP and NFT [6]. Hence, in 60 normal elderly cases [117], 32/60 had no SP, although 30/60 were younger than 65 years, 13/60 had SP in the hippocampus, mainly in sector CA1 and the subiculum, and 12/60 had SP in temporal cortex comprising mainly the 'primitive/mature' and 'burnt-out' types of plaque. It was concluded that it was not possible to distinguish early-stage AD from normal aging at post-mortem [117]. Similarly, SP have been observed in 60% of normal elderly cases, albeit at lower density than in AD [37]. Moreover, Arrigada et al. [29] reported SP in most normal individuals greater than 55 years of age and concluded that there could be a 'continuum' of pathological change from elderly non-demented brains, early stage AD, to advanced AD [16].

The density of neuritic plaques (NP), which incorporate dystrophic neurites (DN), and SP with a distinct 'core' ('classic' plaques), may not be significantly different in AD and aging [36]. A greater vascular involvement in the formation of SP has been observed in aging than in AD [36]. Hence, SP have been observed in the frontal and temporal cortex in 15/20 patients with critical stenosis, most often in the depths of the gyri [180]. Alterations in cerebral perfusion may therefore play a role in SP formation and SP may not always be causally related to dementia. A further study of non-demented patients with critical coronary artery disease suggested that some patients had similar densities of SP to AD, numbers of SP being directly proportional to the duration of arterial disease [181]. In addition, there may be a close link between ischaemic brain episodes and sporadic cases of AD (SAD) suggesting that neovascular factors could aggravate the progression of the disease [150,151].

The most important molecular constituent of the SP is β-amloid (Aβ) [72], an approximately 4 kDa peptide arising by constitutive cleavage of a trans-membrane amyloid precursor protein (APP) [79]. Three subtypes of Aβ deposit are commonly observed in AD, viz., diffuse, primitive, and classic deposits [11,55]. Studies of Aβ deposition have also demonstrated a clear overlap between AD and normal aging. Hence, Aβ deposits were present in non-demented individuals greater than 60 years but were rare before this age [118]. After 60 years of age, Aβ deposits were present in a variety of diseases as a result of aging, especially in the temporal cortex, thus blurring the distinction between AD and related disorders [118]. In 14 non-demented elderly cases, Aβ deposits were present in the temporal lobe in 8/14, but only in cortical gyri, the CA sectors of the hippocampus and dentate gyrus being spared [9]. Moreover, there was a considerable variation in the density of deposits in control cases and a significant overlap with AD. The pattern of clustering of Aβ deposits was also similar in control and AD cases, i.e., in cortical gyri, deposits were aggregated into clusters regularly distributed parallel to the pia mater, suggesting a similar pathogenesis [9]. In a further study of centenarians [56], Aβ deposits were recorded in the parahippocampal gyrus (PHG) of patients, whether demented or not, but the hippocampus was unaffected, suggesting a little relationship between lesion density and severity of mental deficits.

Whether NFT occur as a result of normal aging is more controversial. The most important constituent of the paired helical filaments (PHF) and straight filaments which comprise cellular NFT in AD is the microtubule-associated protein (MAP) tau. Some studies suggest that the majority of cognitively normal individuals have minimal tau-immunoreactive NFT [105] and also less astrocytosis and microglial reaction [51]. By contrast, Bouras et al. [40] found that all non-demented cases had NFT in lamina II of the entorhinal cortex (EC) and in sector CA1 of the hip-
What causes Alzheimer's disease?

Moreover, in non-demented individuals, NFT were more numerous in the medial temporal lobe and in cortical association areas when a memory deficit was present suggesting NFT could be the pathological substrate for memory loss in non-demented as well as demented cases [82]. In the medial temporal lobe, the perforant path appears to be sensitive to tau-immunoreactive pathology in AD and these changes are distinct from those seen in normal aging even in the oldest individuals [67].

Two further aging processes may be involved in AD. First, an age-related breakdown of myelin [33], although other studies suggest that myelin loss occurs late in AD and is secondary to neuronal degeneration [43]. Second, the loss of cells in the locus caerules (LC), which provides noradrenolone to the cortex via terminal varicosities, and stimulates microglia to suppress production of Aβ [87]. Tau-immunoreactive NFT appear early in the LC in aging, mild cognitive impairment (MCI), and in AD apparently forming a continuum [80]. Loss of cells in the LC may therefore induce an age-related impairment of the blood brain barrier, thus implicating vascular factors in AD.

These studies suggest that the differences between AD and the normal elderly are largely quantitative rather than qualitative and there may be a ‘continuum’ of pathological change connecting these cases [17]. Nevertheless, the distribution of the pathology may differ in AD and control brain, being more localised to areas of the temporal lobe in aging and with a more extensive spread into the hippocampus and cortical association areas in AD [53,129,139]. An important question, therefore, is whether AD is an exacerbation of normal aging resulting from enhanced spread of the pathology along anatomical pathways.

Theories based on the degeneration of anatomical pathways

Cholinergic hypothesis

A specific degeneration of the cholinergic neurotransmitter system was one of the earliest theories as to the cause of AD [148]. Several studies prior to 1980 showed significant losses of acetylcholine in the AD brain [41]. Subsequent studies also showed reductions of choline acetyltransferase (CAT), values in the cerebral cortex being 30-50% of those of controls [175], and in levels of acetylcholinesterase (ACHE) especially late in the disease [148]. In the cerebral cortex, there were correlations between the density of SP and levels of CAT, and in the hippocampus between the density of NFT and CAT [147,148]. Moreover, loss of neurons was reported in the nucleus basalis of Meynert (nBM), a site of diffuse cholinergic projection to the cortex [202], and it was suggested that cortical SP may develop on the axon terminals of nBM cells [194]. Hence, AD was considered as the ‘cholinergic’ analogue of Parkinson’s disease (PD); as in the latter, there was degeneration of dopamine neurons originating in the substantia nigra, together with their cortical projections.

Rosser and Mountjoy [163], however, cautioned against extrapolating data from neurotransmitter system studies obtained post mortem as significant changes often occur at the extremes of life. In addition, it became clear that several neurotransmitter-specific neurons contribute neurites to SP [186] and that multiple neurotransmitter deficits were common in AD [46]. Hence, Bowen [42] found increased 5-hydroxytryptamine (5-HT) turnover in AD and suggested that this was due to a selective loss of cortical 5-HT neurons and the subsequent denervation of ascending projections. Hence, although degeneration of the cholinergic system undoubtedly occurs in AD, it is only a part of a more widespread neuronal degeneration affecting many anatomical pathways.

Cortico-cortical pathways

Several lines of evidence suggest degeneration of anatomical pathways connecting different parts of the cerebral cortex occurs in AD [53,139]. A major feature of the anatomical structure of the cerebral cortex is the replicated local neural circuit represented by ‘columns’ or ‘modules’ [130]. The diameter of individual cortical modules varies between 500 µm and 1000 µm, depending on the region, and there are specific connections between ordered sets of columns [90,130]. First, there is a reciprocal projection between a specific cortical area and a dorsal thalamic nucleus. Second, there are inputs to the cortex from generalised regulatory systems, e.g., from the basal forebrain, such as the nBM, and monoaminergic nuclei of the brain stem such as the LC. Consequently, the cortex is interlaced by clusters of fine noradrenergic fibres which repeat at 30-40 µm intervals. Third, homologous neocortical areas, with the
exception of the striate cortex and primary somatosensory cortex, are reciprocally connected in both hemispheres via commissures, both the commissural and ipsilateral association fibres terminating in vertically oriented columns 200-500 μm in width, alternating in a regular sequence with zones of comparable width free of such connections. Fourth, within a hemisphere, different regions are connected by fibres of the short and long cortico-cortical projections [53,90]. The cells of origin of these projections are also clustered and occur in bands approximately 500-1000 μm in width [90].

A number of studies suggest that AD is linked to degeneration of these cortical pathways. First, Scheff and Price [168] measured loss of synaptophysin reactivity in the cortex in AD and related it to synapse loss in the temporal lobe, while Wakabayashi et al. [198] found a decrease in the synaptic marker SP6 in all regions of the AD brain. Second, Pearson et al. [139] studied the distribution of NFT in AD and observed they were located in the cell bodies that give rise to the cortico-cortical pathways, SP developing at their terminals and collateral branches. It was suggested that the disease could spread via these pathways in the orthograde and/or retrograde direction. This pattern of spread was also observed by Hiorns et al. [90] who concluded that in AD, severely affected regions were interconnected by cortico-cortical pathways. The most explicit description of this hypothesis, however, is by De Lacoste and White [53] in which the authors describe AD as a ‘disconnection syndrome’, characterised by disruption of all afferent/efferent connections between the hippocampus, cortex, and the rest of the brain. Hence, if cortico-cortical projections are selectively vulnerable, the pathology may spread in a stepwise fashion via these connections, SP arising from the distal axonal projections of degenerating NFT-bearing neurons [53,139]. In support of this hypothesis, Roberts et al. [160] postulated that regions of the medial temporal lobe are affected first, especially laminae II pre-α neurons of the EC, followed by association areas, and then the primary sensory areas. EC neurons exhibit high levels of APP, which is converted into Aβ at the synaptic terminal resulting in loss of synaptic connectivity [160]. Several studies also suggest that NFT in AD exhibit a spatial pattern consistent with their development in relation to the cortico-cortical pathways [25]. Hoesen and Solodkin [91], for example, showed that NFT selectively damage strips of the cortex and hippocampus, columns of NFT exhibiting a regular periodicity repeating every 80-120 μm and representing 4-5 cell diameters, and with a spacing of 300 μm. With a greater duration of disease, NFT gradually filled up these columns giving rise to clusters of NFT of increasing size, a result also reported by Armstrong [8].

**Cell to cell transfer of pathogenic proteins**

One of the first studies to suggest that the degeneration in AD could spread across normal synaptic projections, and involve the transfer of substances between neurons, was by Saper et al. [165]. More recent research confirms these ideas and suggests that pathogenic proteins, including tau, α-synuclein, the disease form of prion protein (PrPSc), and Aβ may be secreted from cells, enter other cells, and seed small intracellular aggregates within these cells [73,184] (Fig. 1). Hence, tau and Aβ could exit

![Fig. 1. Possible cell to cell transfer of pathogenic proteins and spread of the pathology along anatomical pathways. Tau and β-amyloid (Aβ) could exit cells via exocytosis (Exo) or secretion (Sec) and enter a new cell by endocytosis (Endo) or by interactions with membrane lipids (L). Transfer may also occur via tunnelling nanotubes (TNT) which connect various neurons.](image-url)
cells via exocytosis or secretion and enter a new cell by endocytosis or by interactions with membrane lipids. Transfer may also occur via tunnelling nanotubes (TNT) which connect various neurons [184]. For example, if tau spreads from cell to cell in the cortex, the resulting NFT may exhibit a spatial pattern which reflects this spread. Previous studies have suggested non-random distributions of NCI in the cerebral cortex of various disorders, the inclusions often exhibiting a distinct clustering pattern consistent with their spread via the cortico-cortical pathways [25].

**Theories based on environmental factors**

Many environmental factors have been linked to AD, but most studies relate to three such variables, viz., exposure to aluminium (Al), effect of head trauma, and the influence of diet and malnutrition.

**Aluminium**

Much of the evidence that Al is a cause of AD is circumstantial and controversial [21]. Epidemiological studies [77,120,133] have found little correlation between environmental Al and AD. In addition, out of 13 studies, in which gross brain tissue has been analysed for Al, 9/13 found enhanced levels in AD brain, while 4/13 found no significant differences compared with control brains [34]. The significance of enhanced Al levels in brain is also unclear since damaged brains may accumulate Al [21]. The presence of Al may also be linked to the formation of SP and NFT. Studies have either reported [61], or not reported [109], Al in association with the cores of classic Aβ deposits, but it is not clear whether this type of SP is especially frequent in AD [18]. Al has also been detected in pyramidal neurons containing NFT [142,143]. It has been suggested that Al can bind to DNA and affect processing of cytoskeletal proteins resulting in the formation of NFT. Nevertheless, there is little evidence that Al can bind significantly to DNA [57,120]. In addition, patients with renal failure, exposed to high concentrations of Al, do not develop cellular NFT [142]. Furthermore, whether or not exposure of experimental animals [52] or cell cultures [171] to Al results in NFT formation has been controversial. Experimentally induced cytoskeletal change usually results in the formation of straight 10 nm filaments, unlike the majority of NFT, which are composed of 20-24 nm twisted filaments [190], and which have a different molecular composition in AD [81]. Finally, neuropsychological conditions, which arise from acute exposure to Al, such as dialysis dementia [7] or bladder irrigation [131], in which a 1% alum solution is used to treat haemorrhagic cystitis, probably have little significance apart from demonstrating that Al in acute doses can be neurotoxic.

**Head injury**

Head trauma results in a primary injury which frequently spreads via inflammatory cytokines to initially unaffected regions, thus amplifying the original injury due to the activation of microglia and central nervous system immune cells [71]. Several observations suggest a link between head injury and AD. In survivors of head injury, APP is observed in neuronal perikarya and in DN surrounding Aβ deposits, as in AD [68]. The formation of Aβ from APP occurs within the synaptic terminal fold of axons, the presence of glia not being necessary for this conversion. Hence, the production of APP may be a component of the brain’s response to neuronal injury [68]. Subsequently, it was shown that specific neurons in the medial temporal lobe secreted large quantities of APP and that there were more APP-immunoreactive neurons in these areas in head injury patients [124]. Hence, an increased expression of APP in head trauma cases may be an acute phase response to neuronal injury [161], the overexpression of APP leading to the deposition of Aβ. Several acute phase proteins are localised within Aβ deposits in AD including amyloid-P, complement factors, and α-antichymotrypsin [98]. Furthermore, Regland and Gottfries [154] proposed that APP maintains cell function by supporting neuronal growth and survival. The possible neurotrophic action of APP is supported by the observation that it shares structural features with the precursor for the epidermal growth factor [154]. NFT may also be a part of the neurons response to injury [205]. These studies suggest that the formation of pathological proteins as a result of brain injury is one method by which AD pathology develops and is then propagated within the brain by cell to cell transfer.

**Diet and malnutrition**

Abalan [1] was one of the first authors to propose that AD could be caused by malnutrition. This
The hypothesis is based on clinical observation of AD patients who often exhibit emaciation and cachexia, urinary tract infections, terminal bronchopneumonia, and low triceps skinfold. Low serum albumin, iron, folate, tryptophan, vitamin B$_{12}$, and low cerebral metabolism of glucose and oxygen may also be present. These symptoms suggest a protein calorie malnutrition syndrome in AD which could result in the development of NFT due to chronic nutritional deficiencies of calcium and magnesium. A problem with this type of hypothesis, however, is in determining cause and effect, as malnutrition could be a consequence of the disease resulting from the mental state of the patient [1]. A more direct demonstration of a link between diet and AD has been reported by Sparks et al. [182] in which deposition of Aβ was induced in rabbits fed with high levels of dietary cholesterol. In addition, McCaddon and Kelly [121] found in a human family carrying a mutation of the APP gene (APP$_{717}$ Val-Glycine), that individuals with AD had a greater vitamin B$_{12}$ deficiency, compared with unaffected members. It was concluded that this link was unlikely to be secondary and to be a consequence of impaired dietary intake. A B$_{12}$ deficiency could then result in a reduction of monoamine transmitters and in cholinergic activity.

**Theories based on genetics**

In the 1990s, strong evidence emerged of the connection between familial AD (FAD) and specific genetic factors. Hence, small numbers of cases were linked to APP mutations [48,74] and a larger subgroup to PSEN1/2 mutations [112,173], while others genes are currently unidentified [179]. In addition, allelic variation in the Apo E locus on chromosome 19 was identified as a significant risk factor, especially in late-onset AD [185].

**APP**

A variety of Aβ peptides are formed as a result of secretase cleavage of APP [79]. The most common of these peptides is Aβ$_{42}$, found largely in discrete Aβ deposits, whereas the more soluble Aβ$_{40}$ is also found in association with blood vessels [125] and may develop later in the disease [54]. In addition, mutations of APP within the Aβ coding region may result in the deposition of Aβ$_{38}$ in vessel walls, especially in those cases with extensive cerebral amyloid angiopathy (CAA) [128]. Causation of AD could also be attributable to early soluble peptide oligomers [69], which vary with the type of mutant, thus providing a genetic basis for variations in pathogenesis among FAD cases. The discovery of Aβ led to the formulation of the ‘Amyloid Cascade Hypothesis’ (ACH), the most influential model of the molecular pathology of AD developed over the last 25 years [84] (Fig. 2). Essentially, the ACH proposes that the deposition of Aβ peptides is the initial pathological event in AD leading to the formation of senile plaques and neurofibrillary tangles (NFT), and then to cell death and dementia, both SP and NFT acquiring several additional proteins during their formation, such as apolipoprotein E (Apo E), glial fibrillary acidic protein (GFAP), ubiquitin (Ub), and complement. Other abbreviations: Amyloid precursor protein mutations (APPm), Presenilin genes 1 and 2 mutations (PSEN1/2m).

**Fig. 2.** The original Amyloid Cascade Hypothesis (ACH). The ACH proposes that the deposition of β-amyloid (Aβ) peptides is the initial pathological event in AD leading to the formation of senile plaques (SP) and neurofibrillary tangles (NFT), and then to cell death and dementia, both SP and NFT acquiring several additional proteins during their formation, such as apolipoprotein E (Apo E), glial fibrillary acidic protein (GFAP), ubiquitin (Ub), and complement. Other abbreviations: Amyloid precursor protein mutations (APPm), Presenilin genes 1 and 2 mutations (PSEN1/2m).
synaptic loss, and gliosis [66]. Second, FAD caused by APP<sub>177</sub> (valine–isoleucine) mutation have significant numbers of NFT thus supporting a link between APP and the cytoskeleton [110]. Third, cases linked to PSEN1 have greater numbers of SP and NFT compared with cases of sporadic AD (SAD) suggesting that PSEN1 may also increase tau deposition [172].

There are two main problems regarding the ACH as originally formulated. First, as demonstrated in head injury patients, SP and NFT may be reactive products resulting from cellular neurodegeneration rather than being its cause [27]. The results of animal experiments also suggest that the formation of Aβ could be a reactive process. Experimental lesions that damage the nucleus basalis in rat brain elevate APP synthesis in the cerebral cortex suggesting that the production of APP is a specific response to loss of functional innervation [199]. In addition, chemical lesions of the nucleus basalis using N-methyl-D-aspartate (NMDA) elevate APP synthesis in cortical polysomes [199] and, in areas of brain damaged by kainite [101], APP695 was recorded in DN near to the lesion. Intrathecal or intraparenchymal injections of a toxin also induced APP in hippocampal neurons subsequent to neuronal damage [97].

Lesion experiments may also induce pathological changes leading to NFT. Denervation of the dopamine pathways and septal lesions which affect both the cholinergic system and γ-aminobutyric acid (GABA) neurons projecting to the dentate gyrus, for example, result in loss of dendritic microtubule associated protein 2 (MAP2) and tau-immunoreactivity in dentate gyrus granule cells [193]. Hence, denervation may cause transsynaptic changes in dentate gyrus neurons which represent a precursor stage to NFT.

The second objection to the ACH is that there is no generally accepted mechanism to explain how Aβ deposition leads to NFT [59]. There have been several attempts to establish such a mechanism but none have become universally accepted. First, Aβ may promote the formation of intracellular tau, although the mechanism of this interaction was uncertain [70]. Second, there may be a synergistic interaction between NFT and Aβ [136,177]. Third, when foetal rat hippocampal and human cortical neurons were treated with Aβ, fibrillar forms of Aβ could apparently induce tau phosphorylation. Hence, amyloid fibril formation might alter the phosphorylation state of tau resulting in the loss of microtubule binding capacity.

Studies have also suggested that SP and NFT occur in distinct but independently distributed patterns in AD [19,93]. Studies of the spatial patterns of SP and NFT show them to be clustered, the clusters often being regularly distributed parallel to the pia mater [8]. Clusters of SP and NFT, however, are frequently distributed independently of each other, i.e., neither in nor out of phase, which would not support a direct pathogenic link between them. In addition, SP and NFT may be separated in the brain both in space and in time [119]; in the entorhinal cortex, for example, NFT may precede the appearance of SP [59]. Perez et al. [140], however, showed that Aβ<sub>25-35</sub> could result in tau aggregation and that a decrease in Aβ aggregation was induced by tau peptides. Hence, aggregation of tau may be correlated with disassembly of Aβ which could explain the lack of spatial correlation of the SP and NFT [19].

In transgenic experiments [13], the presence of APP mutations alone or in combination with PSEN1 can induce Aβ deposits in normal brain. Apart from evidence of hyperphosphorylated tau in DN associated with the plaques, these mutations do not appear to induce tau pathology or a significant inflammatory response. Hence, the presence of tau transgenes in the form of a triple model appears to be necessary to completely replicate AD pathology [13]. Hence, the ACH as originally formulated does not appear to provide a complete explanation for AD.

**PSEN genes**

The most common type of FAD is linked to mutations of the PSEN genes [112,173] and the effect of these mutations is also assumed, albeit more indirectly, to lead to the enhanced deposition of Aβ [79]. Full length PSEN is composed of nine trans-membrane domains located on the endoplasmic reticulum membrane. Endoproteolytic cleavage of PSEN and assembly into γ-secretase complex is followed by transport to the cell surface, thus potentially influencing APP processing [92]. Hence, mutant PSEN1 could enhance 42-specific-γ-secretase cleavage of normal APP resulting in increased deposition of amyloid-forming species [188]. PSEN may also act through loss of function by a reduction in γ-secretase activity [206].

PSEN, however, may have a variety of other functions. The PSEN1 gene may be involved in notch signalling [183] and therefore important in cell differen-
tiation. \textit{PSEN1/2} genes may also be involved via the perturbation of cellular calcium homeostasis [207] or in interactions with the transcriptional coactivator \textit{cAMP-response element binding (CREB-binding)} protein which plays a key role in regulating gene expression [64]. \textit{PSEN1} may also mediate neuroprotective functions by ephrin-B [32] and a decline in such protection could be involved in AD.

\textbf{Apo E}

An allelic variation in \textit{Apo E} has been identified as a major risk factor in late-onset AD, individuals with AD having 2-3 times the frequency of allele e4 compared with cognitively normal individuals [185]. In addition, allele e4 may accelerate the development of AD pathology within the aging brain [137] and hence, is often associated with an earlier disease onset [75].

The relationship between the deposition of \textit{A\beta} and \textit{Apo E} genotype has been controversial. The majority of studies, however, report increased amyloid deposition in individuals expressing allele e4 [35,38,152]. In addition, the clustering pattern of NP may reflect the degeneration of specific cortico-cortical and cortico-hippocampal pathways [10]. Cellular NFT often occur in regularly distributed clusters along the cortex [8] and NP may develop on the dendrites and axon terminals of NFT containing cells [53,139]. Individuals expressing allele e4 are associated with the development of NP in smaller and denser clusters compared with the other \textit{Apo E} genotypes, which may reflect a more specific pattern of neurodegeneration [22].

\textbf{Other genes}

\textit{APP} and \textit{PSEN1/2} genes together account for less than 5% of cases of AD [78,90]. Late-onset AD (LO-AD) is likely to be a multifactorial disease with the potential involvement of multiple genes. Hence, genome wide studies have revealed genes on chromosomes 6, 9, 10, 11, 12, 14, 18, 19, and the X chromosome could be involved in AD [44,176] with the most compelling case for a gene on chromosome 12 [138]. Further studies map this locus to 12q13, a region which encompasses the vitamin D receptor (VDR) gene [200]. This gene is a major mediator of the activity of vitamin D and insufficiency of this vitamin has been implicated as a risk factor for AD. In addition, the glyceraldehyde-3-phosphate dehydrogenase (\textit{GAPDH}) gene has been implicated in LO-AD [3]. A major problem with postulating genetic factors as the cause of AD, however, is the similarity of FAD and SAD phenotypes [49,134] and therefore, no specific gene mutation is an essential condition to generate the AD phenotype.

\textbf{Theories based on mitochondrial dysfunction}

Theories that AD may be due to mitochondrial dysfunction have a long history [39] and rely on several lines of evidence. First, an early change in AD is the presence of swollen and distorted mitochondria accompanied by a decline in the cerebral metabolic rate [39]. Second, there may be a deficiency of several enzymes involved in carbohydrate utilization in AD including such mitochondrial markers as phosphofructokinase (PFK) and pyruvate dehydrogenase [39]. Third, part of the familial aggregation of AD may be accounted for by excess maternal versus paternal inheritance, consistent with mitochondrial inheritance. Hence, genomic analysis of 1007 individuals suggested that possession of specific haplotypes could reduce the risk of AD, and therefore, there is a possibility of haplotypes that would increase the risk [159]. The possibility of a genetic origin of mitochondrial dysfunction in AD has led to the formulation of the ‘mitochondrial cascade hypothesis’ (MCH) in which it is the mitochondrial dysfunction that is the root cause of the events leading to the ACH [189]. Nevertheless, it is possible that mitochondrial dysfunction is a further consequence of the ACH [189]. Fourth, ‘adaptor protein evolutionarily conserved signalling intermediate in Toll pathway’ (ECSIT) is a protein which may act as a molecular sensor, ascertaining cell homeostasis in response to oxidative damage by \textit{A\beta} [178]. Protective molecules could then be specifically activated in response and this failure could result in severe mitochondrial damage, promote apoptosis, and result in synaptic dysfunction and neuronal death [178].

\textbf{Theories based on blood brain barrier dysfunction}

The involvement of the cerebral blood vessels in the pathogenesis of AD has been controversial [30]. Some studies have found spatial correlations between \textit{A\beta} deposits and blood vessels suggesting that degeneration of blood vessels or diffusion of
substances from vessels could be involved in the formation of Aβ deposits [127,145,146]. By contrast, other studies conclude that the spatial correlations observed between Aβ deposits and blood vessels are fortuitous and arise because of the presence of high densities of capillary profiles and Aβ [102,115]. In the cerebral cortex of cases of SAD, however, of the three Aβ deposit subtypes, only the classic Aβ deposits exhibited a consistent spatial relationship with blood vessels [15,23]. Classic deposits were clustered specifically around larger blood vessel profiles, such as the vertically penetrating arterioles, the number of classic deposits declining exponentially with distance from the vessels [15].

A number of factors could explain the correlation between Aβ deposits and blood vessels in AD (Fig. 3). First, Aβ could develop in association with the basement membranes or smooth muscle of blood vessel walls [145,204]. Tian et al. [192] found that blood vessels underwent degenerative changes in AD accompanied by Aβ deposition and loss of smooth muscle cells. Second, Aβ could be released by axon terminals or reactive glial cells juxtaposed to vessel walls [96]. Attems et al. [30], for example, found that Aβ 42 was deposited within the glia limitans rather than the capillary walls. Third, diffusion could occur from degenerating arterioles or from clusters of capillaries surrounding the larger blood vessels. Blood vessels with collapsed or degenerated endothelia are evident in more than 90% of AD cases [100] and occur concurrently with Aβ deposition. The integrity of the brain microvasculature may be related to neuronal degeneration and especially to age-related cell losses in the LC [87,116]. In AD transgenic mice, deposition of Aβ in blood vessels is associated with endothelial cell activation and apoptosis [170], which could encourage diffusion. Endothelial cell injury could also depend on duration of dementia [192].

If diffusion is involved in the pathogenesis of the classic Aβ deposits, then various plasma proteins may be implicated. Amyloid fibrils have been observed projecting directly from blood vessels towards classic-type deposits [196] suggesting that Aβ itself could diffuse from blood vessels. However, most Aβ deposited in the cortex is likely to be of neuronal origin [4,12]. A number of other plasma proteins could be involved including amyloid-P, α-antichymotrypsin, antitrypsin, antithrombin III as well as complement factors [96,97] and Apo E [165]. The mRNAs of these proteins, however, is found in brain tissue suggesting that they are made in situ [97]. The likely exception is amyloid-P since the liver is the only tissue of the body which exhibits its mRNA [97]. Plasma proteins, such as amyloid-P, may act as ‘molecular chaperones’ and be involved in the aggregation of Aβ to form plaque amyloid. Hence, secondary damage to cerebral vessels as a result of Aβ deposition from neurons could result in the leakage of plasma proteins from the larger diameter blood vessels. These proteins could then enhance the condensation of already formed Aβ to form a solid amyloid core resulting in the observed pattern of dispersion of classic deposits around blood vessels.

Fifth, deposition of Aβ around the larger blood vessels could be a result of impaired drainage. Extracellular fluid is drained from the brain to the cervical lymph nodes via the perivascular channels [191] and therefore, Aβ around vessels could be attributable to overloading of this system. Such a process could also lead to enhanced deposition of Aβ in the immediate vicinity of the blood vessel and a negative exponential distribution of amyloid deposits. Sixth, another

![Fig. 3. Relationships between the cerebral microvasculature and the development of cored (classical) senile plaques (SP). β-amyloid (Aβ) could develop in association with the basement membranes or smooth muscle (S) of blood vessel walls or Aβ could be released by axon terminals or reactive glial cells juxtaposed to vessel walls. Diffusion of proteins such as amyloid-P could also occur from degenerating arterioles or from clusters of capillaries surrounding the larger blood vessels and influence cored SP development. Deposition of Aβ around the larger blood vessels could also be a result of impaired drainage. Other abbreviations: Capillary end feet (CEF), Neuron (N).](image-url)
function of Aβ may be to serve as an acute response to vessel damage and to seal vascular leakage [106]. Vascular pathology may also be involved in the pathogenesis of Aβ deposits in FAD. Hence, plasma Aβ may be increased in FAD as well as in SAD [187]. In addition, a proportion of FAD cases exhibit CAA similar to that observed in SAD, viz., the deposition of Aβ in and around the major blood vessels. Mutations of the APP gene within the Aβ region are often associated with familial CAA [174] while APP mutations outside the Aβ region result in a pathology deficient in Aβ40, and with lower levels of CAA [108]. In addition, possession of one or more ApoE ε4 alleles significantly increases the risk of AD, and could also be linked to vascular pathology. In transgenic mice, for example, ApoE markedly promotes CAA associated vessel damage [65].

Theories based on immunology

There are several lines of evidence supporting the presence of immune system dysfunction in AD [123,132,162]. Complement proteins associated with the ‘classical’ pathway [123], brain reactive antibodies [132], immunoglobulins [62,95], circulating immune complexes (CIC) in peripheral tissue and cerebral blood vessels [86], helper/inducer and cytotoxic/suppressor T-cells [123], and abundant reactive microglia [123] have all been observed in AD. Complement system proteins may act as pattern recognition molecules mediating the uptake of Aβ by glial cells which express complement receptors [195]. Significant alterations in the major histocompatibility locus (MHC) antigens have also been demonstrated [197] suggesting either increased resistance or susceptibility to an antigen in AD [50,88]. These immune reactions could be a response to the pathological processes of AD and especially to the deposition of Aβ. However, of more interest is whether immune reactions to an environmental antigen could be a cause of AD. Studies suggest that exposure to metals can induce an immune response in peripheral tissues [20]. First, mercury can induce antibodies against renal antigens and inhibit RT6+ T-cells [107]. Second, exposure to cobalt, Al [21], tin, zirconium, and beryllium have been associated with lung inflammatory disease [63]. Lymphocyte transformation, an increase in T-cells, and the development of a granulomatous pneumonitis have been observed in these disorders [63]. Third, Al phosphate and hydroxide are used as vaccine adjuvants [83], i.e., they enhance the immune response to an antigen [60]. Adjuvants containing Al can activate complement proteins [153], prime helper T-cells for IgE production [104] and induce antibody [155]. Since many metals can be absorbed into the brain [34], it is possible that a metal-induced immune activation could occur in brain and contribute to the pathogenesis of AD.

The essential features of a possible hypothesis have been proposed by Sartvetnick and Fox [166] (Fig. 4). The immune system can respond to antigens
PAD4 in astrocytes and neurons respectively results in a citrullinated protein expression in the hippocampus and cerebral cortex. Hence, neuronal loss in AD could result from the loss of cellular contents, including citrullinated proteins, resulting in an auto-immune response and the production of auto-antibodies [2].

**Theories based on infectious agents**

Indirect evidence that infection could be a cause of AD has been reported by Wisniewski *et al.* [203] who suggested that invasion by a virus could cause activation of microglia and pericytes and ultimately, amyloid deposition. In addition, Libikova *et al.* [113] suggested that the virus herpes simplex (HSV), to which antibodies may be observed in the cerebral spinal fluid (CSF) in AD, could induce abnormal protein formation and result in PHF and NFT. This type of theory has been given further credibility by observations of marked structural and biochemical alterations in regions associated with olfaction, most notably the olfactory bulb and EC [58]. The olfactory system is a possible point of entry into the brain for an infectious agent or protein [117]. Hence, experimentally introduced viruses into body cavities and organs often result in high titres of the virus in olfactory epithelium, olfactory bulb, and other brain regions. Nevertheless, it is unlikely that HSV is the responsible virus, as it does not typically enter the brain via the olfactory system [164].

Recent research on cell to cell transfer of pathogenic proteins has re-established interest in infectious agents as a cause of AD and related neurodegenerative diseases [73,184]. The similarity in the properties of Aβ and PrPSc in AD and Creutzfeldt-Jakob disease (CJD), respectively [26], has also contributed to this debate. Such cell to cell transfer raises the possibility, first proposed by Braak and colleagues [85], that in PD, a pathogenic agent introduced via ingestion and/or inhalation, may transfer along axons to basal areas of brain, the brain stem, and then to the cerebral cortex. Hence, in PD, α-synuclein could be the target of the unknown agent causing protein misfolding and subsequent spread via connecting cells (Fig. 1). α-Synuclein taken up from the extracellular space can induce aggregation of other α-synuclein proteins in recipient cells. By analogy with PrPSc, nucleation or seeding activity of α-synuclein may result in a core of transferred α-synuclein surrounded by additional layers of cytoplasmic proteins.
mic α-synuclein contributed by the host cell. Aβ and tau may also have these properties and therefore, cell to cell transfer of an externally acquired infectious agent or protein could be involved in AD.

Discussion

Consideration of the theories

Many theories have been advanced as to the cause of AD and the objective of this section is to select those, which on the basis of the evidence, are most likely to be involved and then to assemble them into a plausible theory. Any theory of AD would need to explain a number of key observations: (1) the heterogeneity characteristic of AD, (2) the overlap between AD and aging, (3) the similarity of FAD and SAD, (4) that Aβ and tau may be reactive proteins, (5) that some cases of AD have few SP and/or NFT, (6) that the pathology is spatially related to cerebral blood vessels, and (7) the immunological responses.

One of the major problems in assessing the validity of any theory is that it is not clear whether it is a primary or secondary event. The theories which have had the most longevity and impact are the cholinergic hypothesis and the ACH. Both are involved in the pathogenesis of AD, but do not provide a complete theory. Hence, there is evidence of cholinergic dysfunction in AD, but only as part of a multi-system degeneration. In addition, the ACH has two problems, viz., SP and NFT may be reactive rather than causal [27], and there is no accepted mechanism to explain how Aβ leads to NFT. Hence, the cholinergic hypothesis is unlikely to be the primary cause of AD and the ACH in its present form requires modification to be included in a general theory.

Of the remaining theories, interest in Al and other metals as a cause of AD has declined, although there is some evidence that Al could accumulate in the brain, interact with processes that result in Aβ deposition and tau formation [20], and that metals may cause an immune reaction in brain. In addition, hypotheses derived from the study of malnutrition and head injury, suggest that these are likely to be risk factors for AD rather than primary causes. Immunological changes and mitochondrial dysfunction in AD are also likely to be reactions to earlier pathological processes such as the formation of abnormal proteins.

By contrast, there is a clear connection between AD and normal aging [158], all features of AD can be observed to some degree in a cognitively normal brain. An important aspect of aging is age-related decline in LC neurons, and its effects on the blood brain barrier could play a role in the disease [87,116]. In addition, there is strong evidence that degeneration of specific anatomical pathways and the spread of the pathology by cell to cell transfer is important in AD.

A possible theory

It is hypothesised first, that the primary factor in AD is an age-dependent breakdown of anatomical systems and pathways within the brain and the consequence loss of synapses [47] (Fig. 5). The extent of this aging effect, which begins early in life, is mediated by the degree of lifetime stress (the ‘allostatic load’). The brain is the ultimate mediator of stress-related mortality through hormonal changes resulting in hypertension, glucose intolerance, cardiovascular disease, and immunological problems [47]. The consequence is gradual synaptic disconnection, neuronal degeneration, and the upregulation of genes determining various reactive and breakdown products such as Aβ and tau [27,149,154,199]. These changes are most manifest initially in the EC which has a high proportion of APP reactive cells. In addition, cell losses in the LC may begin to affect the integrity of the blood brain barrier resulting in the release of proteins with chaperone-type activity. Second, in small numbers of families, specific APP or PSEN mutations directly influence the outcome of this age-related degeneration by determining the solubility and/or toxicity of the molecular product. Cells have mechanisms to protect against the accumulation of misfolded and aggregated proteins including the ubiquitin system and the phagosome-lysosome system. Neuronal degeneration in individuals with specific mutations results in the rapid formation of pathogenic Aβ and tau, and then phases of ‘secondary’ neurodegeneration, which overwhelms the protection systems. Early onset FAD is the consequence of this process. Once initiated, Aβ and tau are then likely to be propagated through the brain via cell to cell transfer from their origin in the medial temporal lobe to the hippocampus and cerebral cortex.

By contrast, in individuals without a specific genetic mutation, but where more complex genetic, e.g. Apo E, and environmental risk factors are present, the outcome of age-related loss of synapses is
mainly soluble and smaller quantities of insoluble proteins which are degraded by the cellular protection systems and do not significantly accumulate to form SP and NFT. With advancing age, however, the protective systems become less effective resulting in slowly accumulating quantities of Aβ and tau. The result of these insidious processes is that the cellular protection systems do not become overwhelmed until much later in life, the consequence being late onset SAD. Subsequently, there are a number of responses to neurodegeneration including a further disruption of the blood brain barrier, an immunological response, and mitochondrial dysfunction which could initiate phases of secondary degeneration. Disease heterogeneity [24] is attributable to variations in the pathways of spread of the pathology from its origin in the medial temporal lobe.

**Predictions of the theory**

The theory suggests that it is aging and the diseases associated with aging that provide the trigger initiating the cascade of events leading to AD, rather than the initial deposition of Aβ, and makes a number of predictions. First, the hypothesis predicts that significant signs of neuronal degeneration in AD should precede those of Aβ deposition and that the effect of Aβ is secondary rather than primary in causing neurodegeneration. Second, that the pathogeneses of SP and NFT are not directly linked and the two lesions essentially arise independently. Third, in transgenic experiments, the effect of the transgene will be age-dependent. In a model which incorporates an APP: V7171 mutation, for example, there was an age-related loss of pyramidal neurons in the hippocampus CA sectors including at sites devoid of plaque deposition [169] consistent with this prediction.

**Implications**

First, the theory suggests that AD is not a disease linked primarily to defective genes but a complex syndrome dependant on the rate of aging and indirectly influenced by genetic risk factors and the environment. Second, the hypothesis questions whether the presence, distribution, and molecular determinants of SP and/or NFT should continue to play an important role in pathological diagnosis. Hence, if SP/NFT are the products of brain degeneration and not its cause, they will represent relatively late stages in pathogenesis and there may be cases of AD that are difficult to classify because they may have insufficient numbers of SP and NFT or exhibit early developmental stages of these pathologies. In addition, if SP and NFT represent the consequenc-
es of specific types of neurodegeneration [11] rather than being characteristic of AD, then many cases may show combinations of pathological features, i.e., there will be a considerable degree of overlap between different disorders. Numerous examples of such cases have been reported in the literature, e.g., dementia with Lewy bodies (DLB) with associated AD pathology, Creutzfeldt-Jakob disease (CJD) with AD, Pick’s disease (PKD) with AD, and these cases are often difficult to classify [17,28]. Third, should significant effort continue to be devoted to immunotherapy and other treatments designed to remove Aβ from the brain? Such treatments could limit the degree of secondary degeneration induced by Aβ. Nevertheless, Aβ might be beneficial to the nervous system by promoting neurogenesis [114] and having a range of other protective functions [111]. In addition, excessive removal of Aβ could reduce chelation within the brain and result in enhanced oxidative stress [31].

Conclusions

Many theories have been advanced as to the cause of AD including those based on: (1) aging, (2) degeneration of anatomical pathways, (3) environmental factors, (4) genetic factors, (5) mitochondrial dysfunction, (6) vascular factors, (7) immune system dysfunction, and (8) infectious agents. It is proposed that AD is a multifactorial disorder in which external and internal factors act together to increase the rate of normal aging (‘allostatic load’). The consequent degeneration of neural pathways and blood vessels results in the formation of abnormally aggregated proteins such as β-amyloid and tau which then spread, via cell to cell transfer, from the medial temporal lobe to affect association areas of brain and then the primary sensory areas. Variation in the pathways of spread leads to the disease heterogeneity characteristic of AD.

References


What causes Alzheimer’s disease?


