1 Introduction

1.1 Literature overview

1.1.1 Transmissible spongiform encephalopathies

The transmissible spongiform encephalopathies (TSEs) are a group of fatal and infectious neurodegenerative diseases in humans and animals. This family includes Kuru (Gajdushek 1977), Creutzfeldt-Jakob disease and a variant of Creutzfeldt-Jakob disease (Collinge *et al.,* 1996; Poser 2002b), Gerstmann-Sträussler-Scheinker syndrome (Masters *et al.*, 1981), and fatal familial insomnia (Lugaresi *et al.,* 1986; Hsiao *et al.,* 1989) in humans. In animals the TSEs are propagated as scrapie in sheep, goats and moufflons (Dickinson 1976; Wood *et al.,* 1992) bovine spongiform encephalopathy in cattle (Wells *et al.,* 1987) transmissible mink encephalopathy in mink (Marsh 1992) chronic wasting disease in mule, deer and elk (Williams *et al.,* 1980). TSEs have been reported in domestic and wild cats, antelopes, and non-human primates grown in zoological gardens (Sigurdson *et al.,* 2003).

All TSEs are characterized by prolonged incubation periods, from months to years, preceding the onset of clinical illness. The progressive clinical symptoms appear incidentally in middle to late adult life and last from months to years prior to death (Chesebro 2003). The TSEs affect mainly the central nervous system (CNS), generally causing bilaterally symmetrical pathological changes (Bruce *et al.,* 1982). However, TSE related disorders of pancreas and cardiac muscles are also described (Carp *et al.,* 1998; Ashwath *et al.,* 2005). The earliest visible changes in the course of the disease are pronounced astrocytosis that includes hypertrophy and hyperplasia of astrocytes in gray matter (Rezaie *et al.,* 2001) and microglia activation (Williams *et al.,* 1997). The most prominent pathological change in TSEs is the formation of vacuoles in neurons, astrocytes, and oligodendrocytes, caused by loss of organelles and swelling of the cells (Landis *et al.,* 1981; Kim *et al.,* 1989; Jeffrey *et al.,* 1992b; Westaway *et al.,* 1994). The vacuolisation may lead to severe spongiform disruption of the tissue, which varies from scattered patches of small holes to diffuse porosity of the grey matter (Hadlow 1999). On the molecular level all TSEs are characterized by accumulation of the abnormal isoform of the cellular prion protein. When released from the surface of the neuronal cells, the abnormal prion protein diffuses through the extracellular space where it forms aggregates (Jeffrey *et al.,* 1995). In many cases these aggregates can be seen, after

immunostaining, as diffuse or compact amyloid plaques, which are a characteristic pathological change in the late stages of the disease (DeArmond *et al.,* 1985; Kitamoto *et al.,* 1986; Roberts *et al.,* 1988).

1.1.2 History of the TSEs

Until the 1960s scrapie of sheep and goat was the only known TSE (Hadlow 1999). The disease was first described 250 years ago in England, but there are suggestions that it was already present in northern Europe and Austro-Hungary in early 1900s (Brown *et al.,* 1998) or even in ancient China (Wickner 2005). Its infectious nature was proven in 1936 when Cuille and Chelle successfully transmitted scrapie to sheep (Chesebro 2003). In 1959, this endemic disease of sheep was proposed to be analogous to kuru, an epidemic neurological disorder found in the eastern highlands of Papua New Guinea (Gajdushek 1977). The brain pathology of people with kuru resembled another human neurological disease, Creutzfeld-Jakob Disease (CJD), a progressive presenile dementia, described in the early 1920s by two German neurologists H. G. Creutzfeld and A. M. Jacob (Poser 2002a). Two more human TSEs Gerstmann-Sträussler-Scheinker syndrome (GSS) and fatal familial insomnia (FFI) were described later (Masters *et al.,* 1981; Lugaresi *et al.,* 1986; Hsiao *et al.,* 1989). The human TSEs were not thought to be transmissible until 1965, when Gajdushek and co-workers demonstrated that chimpanzees developed a characteristic neurodegenerative disease after inoculation with Kuru or CJD brain homogenates (Chesebro 2003).

During the last two decades, the TSEs have attracted medical and public attention with the outbreak of the bovine spongiform encephalopathy (BSE), or “mad cow disease”, epidemic in 1986 (Wells *et al.,* 1987), and with the appearance of the variant Creutzfeldt-Jakob disease (vCJD) (Collinge *et al.,* 1996). The origin of the BSE remains unknown (Marsh 1993; Colchester *et al.,* 2005). The BSE epidemic broke out in the Southwest of England and rapidly spread through the United Kingdom and its trading partners in Europe via contaminated cattle food supplements (Wilesmith *et al.,* 1988; Pattison 1998). Similarly, the infection was transmitted to domestic cats and zoo animals (Bruce *et al.,* 1994). BSE is now spread in more than 25 countries. Since 1987, the epidemic accounts of 189 526 confirmed cases worldwide most of them (184 266) in UK[1].

[1] World Organisation of Animal Health http://www.oie.int/eng/info/en_esb.htm (release 10.08.2005)

Between 1994 and 1996, 12 atypical cases of CJD, called variant CJD (vCJD) have been diagnosed in the UK (Collinge *et al.,* 1996; Will *et al.,* 1996). Transmission studies showed that the vCJD is caused by the same agent as BSE (Bruce *et al.,* 1997). From 1996, 156 cases of vCJD have been reported in the UK, six in France and one each in Canada, Ireland, Italy and the United States of America[2].

1.1.3 Aetiology of the TSEs

Although the infectious nature of the TSEs was established in the 1930s, it is still not firmly proven which agent is responsible for the transmission of the disease. The scrapie agent is characterized by its ability to induce a well-defined neurological condition after a predictable, prolonged asymptomatic incubation period (Bolton *et al.,* 1991). Many structures have been proposed as hypothetical carriers of TSEs (Prusiner 1999; Bastian *et al.,* 2004; Broxmeyer 2004; Ebringer *et al.,* 2005).

The most widely discussed at present is the prion hypothesis. In 1966 it was suggested that the agent responsible for scrapie, which has a molecular weight of about 200 kDa is probably a protein (Alper *et al.,* 1966; Pattison 1966). A year later, J. S. Griffith proposed a theory, that scrapie disease could be caused by an abnormally folded protein molecule that alters the folding of another protein molecule and changes its biological properties (Griffith 1967). In 1982 Prusiner introduced the term "Prion" to describe a hypothetical "small *pro*teinaceous *in*fectious particle which is resistant to inactivation by most procedures that modify nucleic acids" (Prusiner 1982). The only component of the prion is supposed to be a protein derived from the prion protein (PrP) gene, a normal protein-making gene common to birds and mammals (Shaper *et al.,* 1997). The PrP gene, encodes the production of a protein called cellular prion protein (PrP^{C}). Once the disease-causing conformer of the PrP (PrP^{Sc}) enters the body, the normal PrP genes start producing the prion protein isoform-PrP^{Sc}, which accumulates in the brain and causes its degeneration and death. Since the aetiology and the pathogenesis involve modification of the prion protein, TSEs were also termed "prion diseases" (Prusiner 1996a). The term prion, however, is frequently used to signify the infectious TSE agent, regardless of its composition or structure (Weissmann 2004).

[2] The UK National Creutzfeldt-Jakob Disease Surveillance Unit http://www.cjd.ed.ac.uk/index.htm (release 06.09.2005)

For many years the attempts to directly relate the PrP to infectivity have failed (Somerville 2002). However, PrP^{C} from a healthy brain homogenate could be converted into a PrP^{Sc} like product *in vitro* via an autocatalytic process called protein-misfolding cyclic amplification (PMCA) (Saborio *et al.,* 2001). In support of the prion concept, it was recently shown that such replicates induce a specific disease in wild-type hamsters (Castilla *et al.,* 2005). Nevertheless, it remains possible RNA or other molecules present in the original brain preparations to be amplified along with the PrP^{Sc} and play a critical role in conferring infectivity and specificity to the new PrP^{Sc} (Zou *et al.,* 2005). Other reports in support of the "protein-only hypothesis" are those concerning the creation of a new TSE strain by polymerization of recombinant mouse PrP into an amyloid state (Legname *et al.,* 2004; Baskakov *et al.,* 2005; Legname *et al.,* 2005). Intracerebral (*i.c.*) inoculation of this product into transgenic mice, overexpressing N-terminally truncated mouse PrP induce a typical TSE (Legname *et al.,* 2005). If the latter work holds up, it will be a watershed in prion biology (Couzin 2004).

Originally, due to the long period before the appearance of clinical symptoms, the TSEs were classified as a group of diseases caused by an unconventional, slow virus with very unusual physico-chemical properties (Millson *et al.,* 1976; Gajdushek 1977). The agent is filterable, its inoculation is followed by an eclipse phase and its infectious units are multiplied up to 10^{12} per gram of brain tissue. Adaptation to new hosts (shortened incubation period upon serial passaging) and strains of varying virulence and pathogenesis are frequently observed (Gajdushek 1977). The upholders of the virus hypothesis suggest that the pathogen is a virus that induces amyloidosis in the brain tissue (Czub *et al.,* 1986; Diringer 1991; Manuelidis *et al.,* 1995; Narang 2002; Manuelidis 2003). The viral genetic information could be encoded in a nucleic acid fragment of 1.5 kBa, constantly found in highly infectious amyloid fractions (Diringer *et al.,* 1997). The nucleic acid is responsible for the infectivity, whereas PrP is needed for the agent replication possibly serving as a receptor (Arjona *et al.,* 2004). The nucleic acid, perhaps ssDNA, is sandwiched between two layers of protein (Narang 2002), often seen as tubulofilamentous virus-like particles in synaptic terminals and axons (Liberski *et al.,* 1989; Jeffrey *et al.,* 1992b; Narang 2002). However, none of the latter has been consistently related to infectivity, and no scrapie specific nucleic acid molecule has been identified so far (Schreuder 1993; Weissmann 2004; Safar *et al.,* 2005b).

Some researchers believe that the prions are formed when the host PrP associates with a foreign pathogenic factor to create a "virino" particle. This is a chimeric molecule consisted of a still-to-be-discovered nucleic acid and PrP^{Sc} as a shelf protein (Dickinson *et al.,* 1988; Liberski *et al.,* 2004). In 1990 Weissmann (Weissmann 1990) proposed the so-called "unified theory". It suggests that the agent is a molecular chimera in which PrP^{Sc} confers infectivity, and an unidentified oligonucleotide is responsible for the strain-specific characteristics.

Quite intriguingly, the TSEs in humans can be present as sporadic, genetic or infectious disorders (Prusiner 1997). The infectious forms of the diseases include horizontal transmission of the agent by intake of contaminated products including human growth hormone and gonadotropin, transplantations of infected tissues or blood infusion (Prusiner 1997), surgical interventions, improperly sterilized depth electrodes (Collins *et al.,* 1999; Will 2003). This group includes Kuru (Poser 2002a), the cases of vCJD (Bruce *et al.,* 1997; Hill *et al.,* 1997), and infectious CJD (Will 2003). Familial (inherited) TSEs are associated with the presence of an autosomal dominant genetic alteration of the PrP gene (Weissmann 2002; Gambetti *et al.,* 2003b). The mutations in the PrP gene possibly induce alterations in the thermodynamic properties of PrP. The mutated PrP^{C} might more likely than normal PrP^{C} spontaneously convert into PrP^{Sc} thereby initiating a cascade of PrP^{Sc} -catalyzed PrP conversions (Bamborough *et al.,* 1996). Remarkably, more than 50 different mutations in the human PrP gene, resulting in nonconservative substitutions, have been found to segregate with the inherited prion diseases GSS, familial CJD and FFI (Prusiner 1997; Gambetti *et al.,* 2003b). Most of them are amino acid substitutions like the Pro/Leu substitution at codon 102 (P102L) in GSS and the Asn/Asp substitution (D178N) common for FFI, but some are insertions of supernumerary repeats in the octarepeat region (Collins *et al.,* 2001; Weissmann 2002; Wadsworth *et al.,* 2003). Interestingly, a particular mutation in a single PrP pedigree can be present with symptom complexes typical of CJD in some patients and of FFI or GSS in others (Gambetti 1996; Prusiner 1996b).

The causes for about 85% of the CJD and very few GSS and FFI cases remain unknown (Masel *et al.,* 1999; Chesebro 2003; Wadsworth *et al.,* 2003). In these spontaneously arising or sporadic forms of the prion disease, no mutations of the PrP gene and no epidemiological evidence for exposure to a TSE agent through contact with TSE-infected people or animals can be found (Prusiner 1997; Chesebro 2003). Sporadic CJD (sCJD) occurs at an incidence of about 1 case per million population per year worldwide, with an equal incidence in men and

women (Wadsworth *et al.,* 2003). The sporadic TSEs are thought to be initiated by spontaneous misfolding of the PrP^{C} to PrP^{Sc}, possibly as a result of a somatic mutation in the prion gene in a single cell (Prusiner 1997; Chesebro 2003).

1.1.4 The prion protein

1.1.4.1 Properties of PrP^{C} and PrP^{Sc}

The PrP has a dual physiological nature: (i) it has a "normal", not yet defined, cellular function; (ii) it is associated with the propagation of the TSEs in infected individuals. The biological role of the PrP^{C} is still unclear. Current hypotheses include protective function against oxidative stress (Milhavet *et al.,* 2000), regulation of copper homeostasis (Quaglio *et al.,* 2001; Qin *et al.,* 2002), involvement in cell adhesion processes or cell signalling activity (Wechselberger *et al.,* 2002 188; Lasmezas 2003). It is also unclear, what is the pathogenic function of PrP^{Sc}. The exact mechanism of neurodegeneration in the prion diseases is still poorly understood (Unterberger *et al.,* 2005). It was shown that autophagy is involved in the formation of the spongy change in the TSEs (Liberski *et al.,* 2004). Recently published data suggest, that the glycosylphosphatidylinositol (GPI) anchor of PrP^{Sc} could be responsible for the neurotoxicity in the TSEs (Chesebro *et al.,* 2005).

PrP^{C} is a highly conservative sialoglycoprotein that is expressed on the cell surface in neurons and glia of the brain and spinal cord, as well as in some peripheral tissues and in leukocytes (Caughey *et al.,* 1990; Bendheim *et al.,* 1992; Stahl *et al.,* 1993). PrP^{C} may have different topology as well- GPI-anchored, transmembrane, or soluble (Ermonval *et al.,* 2003). In contrast, PrP^{Sc} can be found in infected individuals in the central and the peripheral nervous systems, as well as in extraneural tissues such as spleen, lung, lymph nodes and muscles (Beekes *et al.,* 1998; Bolton 1998; Groschup *et al.,* 1999; McBride *et al.,* 1999; Thomzig *et al.,* 2003). In infected tissues PrP^{Sc} is located intracellularly in the Golgi apparatus, endosomes and lysosomes (Taraboulos *et al.,* 1990; McKinley *et al.,* 1991a; Arnold *et al.,* 1995), but PrP^{Sc} molecules have been detected on the cell surface as well (Caughey *et al.,* 1991a; Jeffrey *et al.,* 1992a). Scrapie and cellular PrP isoforms are encoded by the same chromosomal gene (Basler *et al.,* 1986). The PrP gene is a single-copy, contains an open reading frame in a single exon, and is transcribed to give a single species of mRNA of 2.4 kb. DNA of hamster PrP encodes for polypeptides of 254 amino acids (Figure 1.1 A) (Silveira *et al.,* 2004).

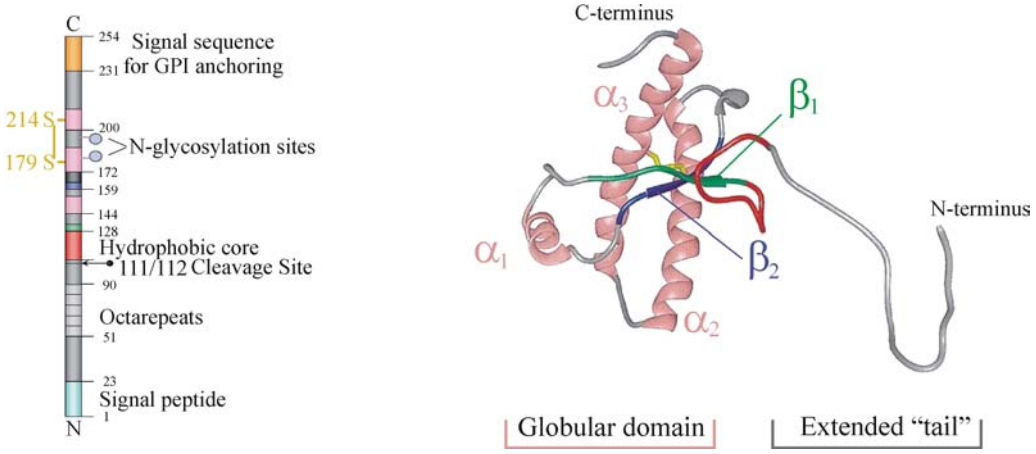

Figure 1.1 Domains (A) and tertiary structure of the PrP^{C} (B) (Liu *et al.*, 1999; Ermonval *et al.*, 2003)

An N-terminal signal peptide of 22 amino acids is removed from the PrP precursor during biosynthesis and additional 23 amino acids are removed from the C terminus of the protein during GPI addition at Ser231 (Bolton *et al.*, 1987; Turk *et al.*, 1988; Haraguchi *et al.*, 1989), resulting in a mature PrP polypeptide of 210 residues. The protein contains two consensus sites for Asn-linked glycosylation at residues 181 and 197. Addition of glycans at these sites generates three PrP glycoforms: unglycosylated, mono-glycosylated and di-glycosylated (Caughey *et al.*, 1989). High mannose glycans are added to the protein in the endoplasmatic reticulum and converted to complex glycans in the Golgi apparatus (Borchelt *et al.*, 1990; Caughey *et al.*, 1990). Although the relative proportions of the individual glycans between PrP^{C} and PrP^{Sc} may differ (Rudd *et al.*, 1999), both protein isoforms contain an identical set of at least 57 individual oligosaccharide chains, found on partially- and fully-glycosylated PrP molecules, which consist of numerous minor glycoforms (Rudd *et al.*, 1999; Pan *et al.*, 2002; Silveira *et al.*, 2004). In addition to the various glycoforms PrP^{C} may be present as different isoforms depending on proteolytic processing (full length or truncated). PrP^{C} has a physiological 111/112 cleavage site within the conservative hydrophobic region (Chen *et al.*, 1995). The PrP^{C} variants are specifically expressed in different brain regions (Liu *et al.*, 2001; Beringue *et al.*, 2003).

1.1.4.2 Structure of the cellular prion protein PrP^{C}

The secondary structure of PrP^{C} has been analysed by circular dichroism (CD) and by Fourier-transform infrared (FT-IR) spectroscopy, revealing a predominance of α-helical content (42%) but a little β-sheet, too (3%) (Pan *et al.*, 1993). Additional studies, using high-

resolution nuclear magnetic resonance (NMR) spectroscopy of the intact polypeptide chain of recombinant hamster PrP, showed a three-dimensional structure, consisting of two distinctly different chain segments of approximately equal size. The C-terminal half of the polypeptide chain forms a well-structured globular domain, whereas the N-terminal half forms a flexible extended "tail" (Figure 1.1 B) (Donne *et al.,* 1997). The globular domain comprises a core (residues 125-228) and a largely unstructured N-terminus (residues 90-119). The transition region between the core domain and the flexible N-terminus, i.e., residues 113-128, does not adopt any regular secondary structure. The globular core consists of three α-helices (residues 144-154-α_1, 172-193-α_2, and 200-227-α_3), two short antiparallel β-strands (residues 129-131-β_1 and 161-163-β_2) and a helical turn (residues 165-169), between β_2 and helix α_2. Helices α_2 and α_3 are linked by a loop formed by a disulfide bond between the Cys179 and Cys 214 residues (Turk *et al.,* 1988: Liu, 1999 #240; Riek *et al.,* 1997).

1.1.4.3 Conversion of PrP^C to PrP^{Sc}

The conversion of PrP into its pathological isoform is a posttranslational event that occurs after the precursor reaches the cell surface (Caughey *et al.,* 1991a; Baron *et al.,* 2002). Presumably, the PrP^C unfolds to some extent, binds specifically to the PrP^{Sc}, and refolds under its influence (Prusiner 1991). Aggregates of 12-24 PrP molecules appear to be the most efficient initiators of the conversion process (Silveira *et al.,* 2005) possibly serving as seeds (Glover *et al.,* 1997; Collins *et al.,* 2004). However, additional "supporting" factors such as chaperons (DeBurman *et al.,* 1997), specific raft domains of the membrane (Baron *et al.,* 2002) or even nucleic acids (Cordeiro *et al.,* 2001; Deleault *et al.,* 2003) may be required for the process of *in vivo* conversion.

The alteration of PrP^C into PrP^{Sc} is conformational (Stahl *et al.,* 1993) and involves a substantial increase in the amount of the β-sheet structure of the protein, with a small decrease in the amount of the α-helix. Infrared spectroscopy indicates 30% α-helix, 43% β-sheet, 11% turns and 16% random coil structural composition for PrP^{Sc}, and 25% α-helix, 54% β-sheet 9-10% turns and 11-16% coil structural content for PrP27-30(Caughey *et al.,* 1991b; Gasset *et al.,* 1993; Pan *et al.,* 1993). Two-thirds of the β-sheet absorption is localized in a region that often reflects intermolecular associations characteristic for amyloids (Caughey *et al.,* 1991b; Safar *et al.,* 1993b). The generation of the PrP^{Sc} isoform involves primary changes in the N-terminus of the protein, including folding of a portion of the N-terminal tail from residues 90

to 121 and part of the first α-helix into a β-sheet, while the C-terminal α-helices remain unchanged (Peretz *et al.,* 1997). The structural rearrangement is accompanied by the acquisition of abnormal physicochemical properties as insolubility in mild detergents and partial resistance to digestion with PK (Harris 1999). PrP^{Sc} aggregates are hydrophobic and characterized by a strong tendency for aggregation with one another and with various cellular components. In contrast to PrP^{C}, whose half-life is of 4 to 6 h (Caughey *et al.,* 1989), PrP^{Sc} appears to be metabolically stable for a period of about 24 h (Borchelt *et al.,* 1990). Consequently, PrP^{Sc}-that, like the full-length PrP^{C} has a molecular weight of 33-35 kDa (McKinley *et al.,* 1986), undergoes a proteolytic cleavage that removes a part of the N terminus within the region attacked by PK yielding a core protein with a molecular weight of 27-30 kDa-PrP27-30 (Harris 1999).

PrP27-30 was first co-purified with infectivity in extracts derived from brains infected with the 263K strain of the scrapie agent (Bolton *et al.,* 1982; Diringer *et al.,* 1983; Prusiner *et al.,* 1983). The presence of PrP27-30 in fractions enriched for scrapie infectivity was accompanied by the identification of rod-shaped particles, measured 10 to 20 nm in diameter and 100 to 200 nm in length (Diringer *et al.,* 1983; Prusiner *et al.,* 1983). These particles can be visualised by electron microscopy as amyloid fibrils known as scrapie-associated fibrils (Merz *et al.,* 1981) or more widely, as prion rods (Prusiner *et al.,* 1983). The prion rods as all amyloid fibrils display pathognomonic green birefringence when viewed in polarized light after staining with Congo red (Prusiner *et al.,* 1983; McKinley *et al.,* 1991b).

1.1.4.4 Extraction and purification of prion rods

The original attempts for purification of the scrapie agent were based on the assumption that the agent is a virus that could be isolated and purified from tissue components by virtue of its size, density, surface charge and specific affinity for certain macromolecules. Early studies suggested that the scrapie agent was distributed throughout virtually all subcellular fractions, associated with cellular membranes (Malone *et al.,* 1979). Scrapie-infected brains were homogenized and centrifuged at low speed to sediment nuclei, unbroken cells and tissue fragments. The supernatant suspension containing the agent was then subjected to various treatments designated to concentrate and purify viruses. The use of equilibrium sucrose or sodium chloride density gradients revealed multiple peaks of infectivity, found throughout the gradients ranging from 1.08 to more than 1.30 g/cm^3. This indicated that some forms of the agent might be associated with lipids (Prusiner 1982) and that, when applied to crude

suspensions of membranous material from brain, the technique is not useful for the isolation of the scrapie agent (Rohwer *et al.,* 1979). In cesium chloride density gradients the agent showed highest accumulation at 1.32 g/ml, but in the presence of cesium chloride 99% of the infectivity was lost (Millson *et al.,* 1976). Since the initial purification of many biological macromolecules involves a series of differential centrifugations, Prusiner and colleagues studied the sedimentation properties of the scrapie agent in fixed-angle rotors (Prusiner *et al.*, 1977). These studies showed that the agent from both murine spleen and brain sedimented in the range from 60S to 1000S. On this basis, a partial purification scheme was developed (Prusiner *et al.,* 1987). The preparation was free of cellular membranes and enriched for the scrapie agent 20- to 30-fold in respect to protein and DNA.

The insolubility of the disease agent in mild detergents (Hunter 1979; Millson *et al.*, 1979) as well as its resistance to nucleases, carbohydrates, lipases and proteases (Hunter 1979) especially its partial PK resistance (Prusiner *et al.,* 1981), its hydrophobic properties and the tendency to form aggregates, implicate that the centrifugation technique, coupled with enzymatic treatment and the use of detergent for extraction, should be useful in the purification of the agent (Diringer *et al.,* 1983). Since 1983 several protocols for prion purification have been proposed. A large-scale protocol was developed by Prusiner's group (Prusiner *et al.*, 1983) and small-scale protocols and their modifications were exploited by other authors (Diringer *et al.,* 1983; Hilmert *et al.,* 1984; Bolton *et al.,* 1985; Bolton *et al.,* 1987). Since small- scale protocols for purification of highly infectious fractions are sufficiently effective and simple, they gained popularity (Diringer *et al.,* 1997). The scrapie specific activity obtained by the purification procedure developed by Prof. Diringer and co-workers (Hilmert *et al.,* 1984), appears to be higher than those from previous methods (Bolton *et al.,* 1991). However, the purified fractions contained about 4 to 8 µg PrP27-30 per brain and considerable amounts of host derivates e.g. proteins and nucleic acids (Hilmert *et al.,* 1984). Further modifications of the latter purification method resulted in a new, easily applicable and reproducible purification protocol developed for extraction and purification of PrP27-30 from Syrian hamsters infected with 263K scrapie strain, which requires 20 g brain material, to obtain fractions with clearly defined chemical composition of 21±3 µg total protein and 19±1 µg amyloid contents, about 2 ng DNA and 1 ng RNA per gram of hamster brain and high infectivity 3-20 x 10^8 LD_{50} (Diringer *et al.,* 1997).

The purified prion rods consist mainly of PrP27-30 (Prusiner *et al.,* 1981; Bolton *et al.,* 1982; Diringer *et al.,* 1997). However, the following cell derived chemical components, intimately associated with the purified prion aggregates, are found in finally purified PrP27-30 samples: (i) two host sphingolipids- galactosylceramide and sphingomyelin (Klein *et al.,* 1998); (ii) a polysaccharide polymer, consisting of 80% 1,4-, 15% 1,6- and 5% 1,4,6-linked glucose units and characterized by a very high molecular mass (Appel *et al.,* 1999); (iii) small amounts of nucleic acids (Kellings *et al.,* 1992; Diringer *et al.,* 1997).

1.1.4.5 Structure of the prion rods

The amyloids are filamentous aggregates formed by non-natively folded peptides or proteins associated with over 20 human diseases including type 2 diabetes, Atrial amyloidosis, and neurodegenerative diseases like Alzheimer and Parkinson (Dobson 2001). According to the hierarchical assembly model the amyloids are formed in a nucleated polymerization process in which nascent protofilaments elongate through the addition of monomeric, partly folded intermediates to the ends of the growing protofilaments (Jarrett *et al.,* 1983; Dobson 2001; Khurana *et al.,* 2003). The rod-like protofilaments then interact with one another to form protofibrils consisting of two or more intertwined protofilaments (Sunde *et al.,* 1997b; Khurana *et al.,* 2003; Tycko 2004). Finally the protofibrils intertwine to form mature fibrils. The driving forces for the assembling of protofilaments in protofibrils and protofibrils in fibrils, as well as the clamping of fibrils arise from the surface of the protofilaments and involve both hydrophobic and electrostatic interactions (Khurana *et al.,* 2003). The morphology and properties of all amyloid fibrils are remarkably similar, even when the precursor proteins in different diseases do not share sequence homology or related native structure (Sunde *et al.,* 1997b; Jimenez *et al.,* 2002).

The tertiary structure of PrP^{Sc} has not yet been described. Since the amyloid fibrils are insoluble and noncrystalline solid materials, they are incompatible with x-ray crystallography and liquid state NMR (Nguyen *et al.,* 1995; Riesner 2003). Therefore no high-resolution molecular structure of an amyloid fibril has yet been determined experimentally. X-ray fibril diffraction of PrP27-30, showed a 4.7 Å reflection typical for a cross-β structure (Nguyen *et al.,* 1995).

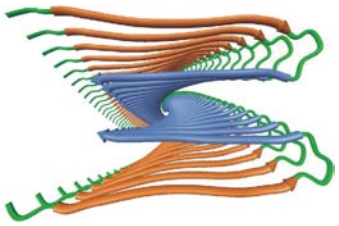

Figure 1.2 Cross β-sheet organisation of an $A\beta_{1-40}$ protofilament (Tycko 2004).

In the cross-β structure (Figure 1.2) a double β-strand is involved in a fibrous protein fold (Sunde *et al.,* 1997a) in which the direction of the intermolecular bond between in register ordered antiparallel (Li *et al.,* 1999; Tjernberg *et al.,* 1999) or parallel, β-sheets, coincide with the direction of the long fibril axis, whereas the β-strands are orientated perpendicularly to the long fibril axis (Tycko 2004; Nelson *et al.,* 2005). The paired sheets are stabilized by closely enmeshed self-complementing side chains protruding from the two sheets, termed a steric zipper, (Nelson *et al.,* 2005) or by hydrogen bonded amino acid side chains (Tycko 2004).

An alternative structural model for the formation of the prion rods was proposed, after 2D crystals found in purified PrP27-30 fractions were investigated by electron crystallography (Wille *et al.,* 2002; Govaerts *et al.,* 2004). The model revealed that PrP^{Sc} can adopt either a cross-β sheet or parallel β-helical architecture.

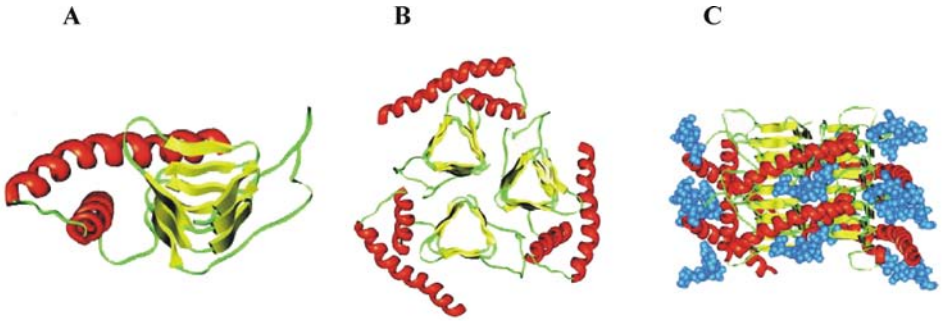

Figure 1.3 Prion rod constructed of monomers with a left-handed β-helical fold (Govaerts *et al.,* 2004).

The N-terminal sequence of PrP (residues 89–174) is compatible with a parallel left-handed β-helical fold (Figure 1.3 A) (Govaerts *et al.,* 2004). Left-handed β-helices readily form trimers, providing a natural template for a trimeric model of PrP^{Sc} (Figure 1.3 B). The trimeric model accommodates the PrP sequence from residues 89–175 in a β-helical conformation with the C terminus, retaining the disulfide-linked α-helical conformation (residues 177–227)

observed in the normal cellular isoform. In addition, the proposed model matches the structural constraints of the PrP27-30 crystals, appropriately positioning residues 141–176 and the N-linked sugars. Two discs of PrP27-30 assemble through polar backbone interactions between the lower β-helical rung of the top disk and the upper rung of the bottom disk (Figure 1.3 C). This assembly provides enough room for the α-helices to stack and the N-linked sugars to extend away from the centre of the structure (Govaerts *et al.,* 2004).

1.1.5 Strains of the transmissible spongiform encephalopathies

The TSE agents, like conventional microorganisms, exhibit strain variations (Bruce *et al.,* 1991), which were first recognized some 40 years ago, when experimental sheep scrapie was transmitted and serially passaged in goats and mice (Pattison *et al.,* 1961). Two separate passage lines were maintained that produced strikingly different clinical signs in unselected goats from the same herd, either “drowsy” or a “scratching” syndrome. The strain properties have been observed for TSEs in several mammalian species but are most thoroughly documented in experimentally infected mice (Priola 2001; Bruce 2003). The mouse model reveals the existence of more than 20 different TSE isolates from a range of sheep, goat, cattle or human sources (Priola 2001). TSE strains differ in their clinical characteristics (Carp *et al.,* 1984) and their ease of transmission between species (Kimberlin *et al.,* 1978). The strains retain their identity after propagation in different host species or PrP genotypes, even when passaged from different brain regions or spleen (Carp *et al.,* 1997), showing that these agents carry their own strain-specific information. Furthermore, the strains interact with genetic factors in the host to influence the characteristics of the disease e.g. the length of the incubation time, clinical symptoms, brain pathology and species tropism (Fraser *et al.,* 1967; Outram 1976; Kimberlin *et al.,* 1978).

The strains often show specific variations related to the physicochemical properties of the PrP^{Sc}. PrP^{Sc} from different strains differ in molecular weight (Bessen *et al.,* 1992; Parchi *et al.,* 1996), glycoforms ratio and the PrP27-30 profile in one- (Somerville *et al.,* 1997; Thomzig *et al.,* 2004) or two-dimensional (Pan *et al.,* 2001; Zanusso *et al.,* 2002) gel electrophoresis. The strain specific PrP^{Sc} shows variations in its resistance to thermal (Somerville *et al.,* 2002) and chemical inactivation, relative PK resistance of PrP27-30 (Bessen *et al.,* 1994; Kuczius *et al.,* 1999; Horiuchi *et al.,* 2002), conformational stability and insolubility in detergents (Somerville *et al.,* 1989; Bessen *et al.,* 1994; Peretz *et al.,* 2001). Additionally, different strains

show PrP^{Sc} with different membrane topology (Hegde *et al.,* 1999) and ultrastructure of their aggregates (McKinley *et al.,* 1991b; Jeffrey *et al.,* 1995).

1.1.5.1 Strain differentiation in TSE

The most obvious difference between TSE strains is in their incubation periods in inbred mice. There are also striking differences in the type, severity and distribution of the pathological changes they cause in brain (Fraser *et al.,* 1967). These two aspects of the TSE strain variation have been first explored in the course of the development of quantitative methods for strain discrimination (Priola *et al.,* 2003).

1.1.5.2 Incubation periods and histopathology

The incubation period of the TSEs in mice, depends on the strain (Bruce *et al.,* 1987) and on genetic factors of the host (Dickinson *et al.,* 1968; Kingsbury *et al.,* 1983). TSEs in ordinary mice are characterized by long asymptomatic incubation periods, followed by progressive neurological symptoms. In general, when a single strain of scrapie is injected *i.c.* into a group of mice of a single inbred line, the incubation period between the initial infection and the terminal stage of the disease is remarkably uniform for all individuals in the group and is also highly repeatable. This method has yielded precise and reproducible results for a wide range of mouse scrapie models, giving standard errors of less than 2% of the mean incubation period (Bruce 2003). The strain differentiation capacity of the mouse model was considerably increased by the use of mice carrying different alleles of the *Sinc* gene, the major host gene that controls the incubation period (*Sinc* from "*s*crapie *inc*ubation") (Dickinson *et al.,* 1968; Dickinson *et al.,* 1971). The *Sinc* gene with its alleles *a* and *b* was found to encode a prion protein that differs by two amino acids: Leu/Phe at codon 108 and Thr/Val at codon 189 (Brown *et al.,* 2000). The TSE strains interact with the two PrP alleles and produce in both homozygotes and in the heterozygote F_1 cross a characteristic difference in the incubation periods up to hundreds of days (Dickinson *et al.,* 1971; Fraser *et al.,* 1985). Genes other than the PrP gene influence the incubation period (Kingsbury *et al.,* 1983) but usually to a lesser extent (Race *et al.,* 1990). The incubation period is widely used to differentiate TSEs strains in the hamster model as well (Kimberlin *et al.,* 1975; Marsh *et al.,* 1975; Lowenstein *et al.,* 1990; Thomzig *et al.,* 2004).

Another fundamental method for TSE strain discrimination is based on the neuropathological changes observed in infected brain. Using the severity of vacuolation in

different sections of both grey and white matter, a “lesion profile” can be constructed characteristic for each combination of TSE strain and host genotype (Fraser *et al.,* 1967; Fraser *et al.,* 1968; Kimberlin *et al.,* 1975; Marsh *et al.,* 1975; Outram 1976; Fraser *et al.,* 1985; Lowenstein *et al.,* 1990; Thomzig *et al.,* 2004). The strain specific targeting can be demonstrated clearly in sections immunostained with PrP-specific antibodies (Schulz-Schaeffer *et al.,* 2000; Thomzig *et al.,* 2004). As for vacuolation, there are clear and reproducible differences between TSE strains in the distribution and severity of these changes (Masters *et al.,* 1981; DeArmond *et al.,* 1993; Jeffrey *et al.,* 1994). As a consequence of the variation in the brain pathology some scrapie strains exhibit distinct behavioural and neurological symptoms during the course of the disease (Marsh *et al.,* 1975; Carp *et al.,* 1998; Thomzig *et al.,* 2004), which can be used to reveal strain diversity (Dell'Omo *et al.,* 2002).

1.1.5.3 Classification of sporadic CJD

Five different phenotypes of the disease have been defined- “Typical”, “Early onset”, “Long duration”, “Kuru plaques” and “Ataxic” (Gambetti *et al.,* 2003a). Although, they are well documented, there is still no systematic method for discriminative diagnostic of the disease variants (Parchi *et al.,* 1999; Hill *et al.,* 2003). Current molecular classification of the sCJD includes determination of the codon 129 state and the electrophoretic patterns of PrP^{Sc} (see, Table 1.I next page). The PrP gene at codon 129 shows common methionine/valine (M/V) polymorphism resulting in three PrP genotypes present in the human population: M and V homozygous, and M/V heterozygous. The PrP polymorphism at codon 129 is a major determinant of the sCJD variations (Parchi *et al.,* 1999; Hill *et al.,* 2003). Another important factor, which influences or reflects the sCJD disease characteristics are the individual properties of the PrP^{Sc} in PAGE. The Western blot pattern of PrP^{Sc} shows variations in the molecular size and/or glycosylation of the PrP^{Sc} pool isolated from diseased brain tissue from patients with different clinical propagation of the disease. The electrophoretic mobility of proteinase K resistant fragments of PrP^{Sc}, reveal the presence of two different PrP^{Sc} types, referred as type 1 and 2 according to the migration of the unglycosylated PrP band at 21 and 19 kDa , respectively (Parchi *et al.,* 1996; Parchi *et al.,* 1997).

Table 1.I Molecular classification of sCJD cases

PrP^{Sc} type	Codon 129 polymorphism		
	MM	**MV**	**VV**
1	Typical	Typical	Early onset
2	Long duration	Kuru plaques	Ataxic

(Gambetti *et al.*, 2003a)

However, currently there is a significant disparity in the literature regarding the variety of the distinct human PrP^{Sc} types. Several studies have revealed the existence of higher number of PrP^{Sc} types seen as differences in the migration patterns and glycosylation ratios of the three PrP bands, which could encode three, four or even six main variants and several subvariants (Aucouturier *et al.*, 1999; Wadsworth *et al.*, 1999; Gambetti *et al.*, 2003a; Hill *et al.*, 2003; Zanusso *et al.*, 2004). The considerable micro-heterogeneity of the PrP^{Sc} implicates the possibility of subclassification of the sCJD. The demonstration of the existence of different human prion strains require serial passage studies in experimental animals including transgenic mice expressing human PrP (either 129 M or V) (Hill *et al.*, 2003). Obviously the PrP^{Sc} typing and classification of sCJD cases is not a trivial task. It is complicated due to variability between gels, which hinders the accurate measurement of the molecular masses of the PrP^{Sc} bands. It should be noted that bound metal ions or variations in the pH affect the PrP^{Sc} conformation, thus influencing the PK cleaving site or the molecular size of the PrP27-30 molecules (Wadsworth *et al.*, 1999; Hill *et al.*, 2003; Notari *et al.*, 2004; Zanusso *et al.*, 2004).

1.1.5.4 Structural diversity of PrP^{Sc} from different TSEs strains

The possibility for the existence of multiple conformations of PrP^{Sc} was suggested after the observations that PrP27-30, associated with different strains of TSE, is cleaved at different N-terminal sites by PK. This was first documented with the Hyper (HY) and Drowsy (DY) strains of transmissible mink encephalopathy passaged in hamsters (Bessen *et al.*, 1992). In SDS-PAGE analysis, PK treated PrP^{Sc} from HY-infected hamsters appears to have a ~2 kDa greater molecular mass than the PK treated PrP^{Sc} from DY-infected hamsters. N-terminal sequencing revealed that PK cleaved additional residues from the N-terminus of DY PrP27-30 compared to HY PrP27-30. Since the DY and HY PrP27-30 molecules are derived from the same Syrian hamster PrP^{C} precursor, the conclusion was that the observed differences in cleavage by PK were due to strain-dependent differences in the conformation and/or ligand

(i.e. metal ions) binding (Bessen *et al.,* 1992; Oesch *et al.,* 1994; Wadsworth *et al.,* 1999). The acquisition of a different structure by PrP^{Sc} could be a result of different posttranslational modifications-glycosylation, endosomal-lysosomal digestion of the newly synthesized PrP molecules in different cell types supporting different TSE infections, which leads to differences in the biochemical properties of PrP^{Sc} (Bessen *et al.,* 1992). The conversion of PrP^{C} to PrP^{Sc} is a dynamic process and, for unclear reasons the distribution of the spongiform change and PrP^{Sc} deposition varies among prion strains. An obvious explanation is that the transformation efficiency in any given brain region depends on favourable interactions between conformations of PrP^{C} and the prion strain being propagated within. By using specific monoclonal antibodies (Liu *et al.,* 2001; Beringue *et al.,* 2003) it was established, that PrP^{C} variants i.e. glycoforms, and truncated forms, are indeed structural conformers, and some of them are differentially expressed in different regions of the mouse brain. This suggests that variation in the expression of the PrP^{C} conformations in different brain regions may dictate the strain properties of PrP^{Sc}. However the variation of PrP^{Sc} conformation could be also due to self propagating activity of the TSE strain specific conformation (Caughey 2003).

The structural diversity of prion strains becomes “visible” by Fourier transform infrared (FT-IR) spectroscopy studies of PrP^{Sc} and PrP27-30 derived from different TSE strains adapted to Syrian hamsters. These studies showed that the secondary structure characteristically differs between various TSE agents and that the strain differences are “encoded” mainly in conformational variations of the β-sheet structure (Caughey *et al.,* 1998) as well as in other secondary structure components such as α-helices and turns (Thomzig *et al.,* 2004). Recently, related but subtle differences have been reported among spectra of various strains of mouse PrP^{Sc} (Silveira *et al.,* 2004). Evidence for multiple strain differences in PrP^{Sc} conformation has also been obtained with multispectral ultraviolet fluorescence spectroscopy (Rubenstein *et al.,* 1998). However, the most detailed investigation of the conformational diversity in PrP^{Sc} has been performed by highly sensitive conformation-dependent immunoassay developed by Safar and co-workers (Safar *et al.,* 1998). This immunoassay quantifies PrP isoforms by parallel analysis of denatured and native forms of the protein with specific antibodies. The ratio of antibody binding to denatured/native PrP showed a characteristic protein conformation for eight prion strains propagated in Syrian hamsters (Safar *et al.,* 1998).

1.1.5.5 FT-IR spectroscopy for TSEs strain differentiation

The investigation of the biological properties of TSE isolates in mice is currently the only reliable method of TSE agent strain typing (Bruce *et al.,* 1994; Begara-McGorum *et al.*, 2002; Thomzig *et al.,* 2004). However such a method is time- and cost-consuming and requires extremely large numbers of animals (Gretzschel *et al.,* 2005). Furthermore, not all scrapie strains are readily transferred to experimental animals. Due to selection, adaptation or variation during the passaging process, the strains finally characterized may not be representative of the original infective material (Begara-McGorum *et al.,* 2002). Moreover, because the incubation period, clinical representation, histopathology and some of the properties of PrP27-30 overlap when different strains are compared they cannot be used for definitive strain typing (Thomzig *et al.,* 2004).

To overcome these problems a method that recognizes unique strain specific differences is required for direct and reliable strain discrimination. The biochemical behaviour of PrP^{Sc} under denaturing conditions could provide a biochemical means for further characterization of different natural scrapie isolates (Madec *et al.,* 1997; Somerville *et al.,* 2002). Also the secondary structure of the PrP^{Sc} or PrP27-30 appears to be a strain specific characteristic of the TSE agent (Caughey *et al.,* 1998; Thomzig *et al.,* 2004). Therefore, methods based on the differentiation of structural and/or structural stability variations between purified disease-associated PrP isoforms, should provide sufficient information for the discrimination of TSEs strains. In contrast to X-ray crystallography and NMR spectroscopy, infrared (IR) spectroscopy can not reveal a 3D image of a protein structure, and is limited to the determination of the secondary structure. However using IR spectroscopy, the protein samples can be studied in aqueous solutions regardless of their state of aggregation (Stuard 1997). Moreover, the technique is very sensitive in detecting conformational changes (Barth *et al.,* 2002; Fabian *et al.,* 2002). It is therefore a very suitable tool for the investigation of the PrP27-30 secondary structure, stability and/or the monitoring of unfolding/refolding processes induced by chemical denaturants, high pressure or temperature. In comparison to circular dichroism spectroscopy (CD), which can also be applied for protein secondary structure analysis, the absorption features, reflecting conformational changes and secondary structure elements in the IR spectroscopy are very specific and without any interference from aromatic amino acids, buffers and salts (Jackson *et al.,* 1995; Stuard 1997). As CD, IR can monitor thermally and chemically induced alterations in proteins, but provides more detailed structural information about these

processes. In contrast to CD, IR could be used to investigate proteins in a solid state (Stuard 1997).

1.2 Aims

The main goal of the present work is the develop a fast, robust and specific method for TSEs strain differentiation, which can be easily applied for discrimination of strains in human prion diseases, for surveillance of strains with a potential of zoonotic transmission, such as BSE, scrapie and chronic wasting disease, or discrimination of TSE strains with public health impact such as BSE and natural scrapie in sheep under natural conditions.

The framework of this project included two main parts: first, evaluation of the potential of the Fourier-transform infrared spectroscopy technique to discriminate TSE strains adapted to Syrian hamsters, and second, discrimination of strains within human prion diseases.

The first part of the project was intended to adapt the existing PrP27-30 extraction and purification procedure, developed in the literature, to a small amount of starting brain material. To purify and investigate by various FT-IR techniques PrP27-30 from four hamster adapted TSE agents (263K, ME7-H, 22A-H and BSE-H). To find specific spectroscopic indicators that could provide an objective discrimination between PrP27-30 from different TSEs.

The second part of the project was intended to apply the developed discrimination technique to a set of human brain samples including sCJD cases and controls from diseased and healthy individuals in order to distinguish (i) the cases of prion disease from the controls, and (ii) to discriminate the variants causing different forms of sCJD.

1.3 Fourier-transform infrared spectroscopy

1.3.1 Infrared spectroscopy

The electromagnetic spectrum includes all varieties of electromagnetic radiation, each of which could be considered as a wave or particle propagating with the speed of light. These waves differ from each other in the length and frequency, and hence in their energy. The frequency (ν) is the number of wave cycles per second. The wavelength (λ) is the length of one complete wave cycle. Wavelength and frequency are inversely related:

$$\nu = \frac{c}{\lambda} \text{ and } \lambda = \frac{c}{\nu} \qquad \textbf{Eq. 1.1}$$

where c is the speed of light, 2.997925×10^{8} ms^{-1}

and the energy of the light is related to the wavelength as follows:

$$E = h\nu = \frac{hc}{\lambda} \qquad \textbf{Eq. 1.2}$$

where h is the Planck's constant, 6.6×10^{-34} Js

Infrared (IR) radiation is the part of the electromagnetic spectrum between the visible light and microwaves regions, with wavelengths between 700 nm and 1 mm. The wavelength in IR spectroscopy is generally given in wavenumbers ($\bar{\nu}$). The wavenumber is the inverse of the wavelength:

$$\bar{\nu} = \frac{1}{\lambda} \qquad \textbf{Eq. 1.3}$$

$\bar{\nu}$ is given in cm^{-1}

hence:

$$E = hc\bar{\nu} \qquad \textbf{Eq. 1.4}$$

In IR spectroscopy the sample is exposed to infrared radiation, which is absorbed by the molecules, when a vibrational mode in the molecule produces an oscillating electric field with the same frequency as the frequency of the incident IR light. The number of fundamental vibrational modes in a particular molecule depends on the degrees of motional freedom. A molecule consisting of N atoms has a total of $3N$ degrees of motional freedom. In nonlinear

molecules, 3 of these degrees are translational, 3 rotational and the remaining correspond to the internal movements, also called normal modes of vibrations. In a linear molecule, 3 degrees are translational and 2 degrees are rotational since no detectable energy is involved in rotations around the molecular axis. Or, the normal modes of vibrations for linear and nonlinear molecules are 3*N*-5 or 3*N*-6, respectively. For any diatomic molecule only one kind of vibration is possible, which corresponds to stretching movements along the chemical bond, this accounts for one degree of vibrational freedom. Nonlinear triatomic molecules have three modes, two of which correspond to stretching and one to bending motions. A linear triatomic molecule has four vibrational modes. Molecules with more than three atoms have an increasing number of vibrational degrees of freedom, which include torsion and movements of fragments in relation to each other. However, not all calculated vibrations can always be observed since (i) vibrations that do not change the dipole moment are IR inactive, (ii) when two different vibrations have the same frequency, the vibration modes degenerate, (iii) an absorption frequency based on multiple fundamental frequency, or overtone, can overlay a normal vibrational mode if both share similar absorption frequencies and intensities. Additionally, in a molecule two oscillating bonds can share a common atom. When this happens, the vibrations of the two bonds are coupled. As one bond contracts, the other can either contract or expand. In general, the absorption bands of coupled bonds occur at different frequencies.

The simplest model for the description of vibrations between two atoms is the harmonic oscillator, a system composed of two finite masses joined by an ideal spring. The potential energy in this system is given by the equation:

$$E = \tfrac{1}{2} k x^2 = h\nu \qquad \textbf{Eq. 1.5}$$

where k is the force constant of the spring x is the displacement of the spring and ν is the frequency of vibration

The allowed transition levels for the harmonic oscillator are given by:

$$E = (n + \tfrac{1}{2}) h\nu \qquad \textbf{Eq. 1.6}$$

where n is the quantum number (0, 1, 2, 3, …)

The lowest energy level is E_0=1/2 hν. It depends on the strength of the bond and, inversely, on the mass of the atoms and corresponds to a type of molecule vibration that is present even at absolute zero. The next highest level is E_1=3/2 hν. According to the selection

rules, only transitions to adjacent energy levels are allowed, therefore the molecule will absorb an amount of energy equal to hν. The constant difference between neighbouring energy levels is a typical feature of a harmonic oscillator. Its absorption spectrum would contain a single line for each vibrational mode.

The harmonic oscillator model is often a sufficient description for the E_0 energy level but is limited for higher energy levels. However, real molecules cannot be ideal harmonic oscillators because of the bond dissociation and the fact that the chemical bond cannot compress beyond a certain level. In fact real molecules are anharmonic oscillators. In this model the potential energy increases more strongly with compression, and less with stretching of the chemical bond. The model that considers these energy changes and the fact that diatomic molecules can dissociate is the Morse oscillator potential. According to this model the overall energy level expression is:

$$E_n = (n + 1/2)h\nu - a(n + 1/2)^2 h\nu \qquad \textbf{Eq. 1.7}$$

where *a* is the anharmonicity constant, which is in the order of 10^{-3}.

As the vibrational quantum number *n* increases, the energy levels get closer and closer together (Figure 1.4). The dissociation energy level is reached as soon as the energy difference becomes zero.

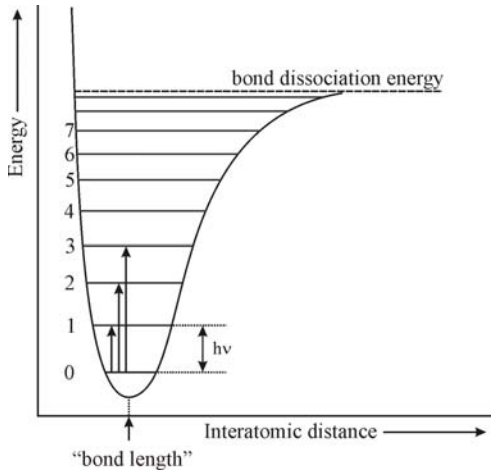

Figure 1.4 Morse oscillation potential

Molecules behave as harmonic oscillators only at low vibrational quantum numbers. Deviations from the harmonic behaviour occur at high vibrational quantum numbers. This

leads to the following two effects: (i) at higher vibrational quantum numbers the energy difference ΔE corresponding to Δn = ±1 is no longer constant. As a result transitions with Δn = ±2, ±3… are allowed, where the transition probability for Δn= ±1 is significantly larger than for Δn= ±2, which in turn is significantly larger than that for Δn= ±3 and so on. The spectrum of the anharmonic oscillator shows a normal mode (n=0 → n=1) and significantly weaker overtone vibrations (n=0 → n=2, n=0 → n=3,…) at frequencies slightly below the double or triple values and with intensities much smaller than the fundamental vibration; (ii) different vibrations in a molecule can interact to give absorption peaks with frequencies that are approximately sums or differences of their fundamental frequencies. However, the intensities of overtone, combination and difference peaks are low.

1.3.2 FT-IR spectrometer

Infrared spectrometers are commercially available since the 1940s. At that time the instruments employed prisms as dispersive elements. In the mid 1950s diffraction gratings have been introduced into dispersive machines. The most significant advance in IR spectroscopy, however, came along with the introduction of Fourier-transform (FT) spectrometers. The basic components of an FT-IR spectrometer are shown in

Figure 1.5. The radiation emitted by the IR source goes trough an interferometer to the sample before reaching the detector. After amplification of the signal, the data are converted in digital form by an analog-to-digital converter and then transferred to a computer for Fourier transformation to be carried out.

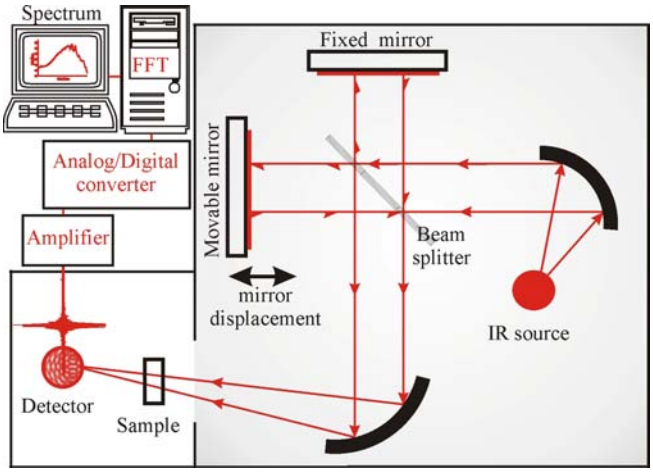

Figure 1.5 Components of an FT-IR spectrometer

The essential component of an FT-IR spectrometer is the optical hardware producing the interfering light, called interferometer. The most common interferometer used is the Michelson's interferometer (shadowed panel, Figure 1.5), which consists of two plane mirrors perpendicular to the optical axis. One of the mirrors can move along the optical axis. Infrared light emitted by a Globar or Nernst source is directed to a device called beam splitter, which allows half of the light to pass trough, while it reflects the other one. The reflected part of the beam travels to the fixed mirror through distance *L*, reflects there and hits the beam splitter again after a total pathlength of *2L*. The same happens with the transmitted part of the beam. However, as the reflecting mirror for this interferometer is not fixed at the same position *L* but can be moved back and forth by distances *x*, the total pathlength is accordingly *2(L+x)*.

Thus, when the two halves of the beam recombine again on the beam splitter, they a have path length difference of *2x*.The partial beams are spatially coherent and will interfere when they recombine. The movable mirror is the most crucial component of the interferometer. It has to be accurately aligned and capable of scanning two distances so that the path difference corresponds to a known value. FT-IR spectrometers use the interference pattern of a helium neon laser to control the change in the optical path difference with wavenumber accuracy of 0.01 cm^{-1}. The beam leaving the interferometer is passed through the sample and is finally focused on the detector. The quantity measured by the detector is the intensity *I(x)* of the recombined IR beams as a function of the moving mirror displacement *x*, so-called interferogram. A frequently used detector is a pyroelectric device incorporating deuterium tryglycine sulfate (DTGS). For more sensitive experiments, mercury cadmium telluride (MCT) detectors cooled to liquid nitrogen temperature are used.

1.3.3 Spectra processing

The interferogram is registered as an analog signal by the detector, and must be digitized prior to FT transformation into a conventional single beam spectrum. This transformation involves integration over a finite displacement. As a consequence, the apparent line shape of a spectral line may be present as a main band which has a series of negative and positive side lobes with diminishing amplitudes. The process of apodisation is the removal of these side lobes by multiplying the interferogram by a suitable function, e.g. so called Blackman-Harris-3-term function, which causes the intensity of the interferogram to fall smoothly to zero at its ends. The number of data points per wavenumber in the spectrum can be increased by adding zeros to the end of the interferogram before performing FT. This increases

the number of data points in the spectrum, Zero filling is equivalent to a very efficient data point interpolation. Every point of an interferogram contains information about all frequencies present in the spectrum. Because this frequency information depends on the distance the mirror has moved from the interferogram centerburst, or so-called zero path difference point, sampling on either side of the centerburst provides the same frequency information. Theoretically, the interferogram should be perfectly symmetrical but in reality, the sample intervals are not exactly the same on each side of the maxima, corresponding to zero path difference. Phase correction is then required, and this correction procedure ensures that the sample intervals are the same on each side of the first interval and should thus correspond to a zero path difference.

After amplification, the analogue signal is converted in digital form by an analog-to-digital converter. The IR data acquisition yields the digitalized interferogram *I(x)* that is being converted into a spectrum by means of a mathematical operation called Fourier-transformation (FT). The essential equations that relate the intensity measured by the detector *I(x)*, to the spectral power density at a particular wavenumber, given by $B(\bar{\nu})$, are as follows:

$$I(x) = \int_{-\infty}^{+\infty} B(\bar{\nu}) \cos 2\pi \bar{\nu} x \cdot d\bar{\nu} \qquad \textbf{Eq. 1.8}$$

where $B(\bar{\nu})$ is the Fourier coefficient.

$$B(\bar{\nu}) = \int_{-\infty}^{+\infty} I(x) \cos 2\pi\, \bar{\nu} x . dx \qquad \textbf{Eq. 1.9}$$

These two equations are interconvertable and are known as a Fourier-transform pair. The first shows the variation in the power density as a function of the difference in pathlength, which is an interference pattern. The second shows the variation in intensity as a function of wavenumber. In practice, the spectrum is calculated from the interferogram by an algorithm described by Cooley and Tukey, called fast Fourier-transform (FFT) algorithm, which reduces the number of complex calculations. The relation between the band intensity and the concentration is given by the Beer's low. For a single analyte in nonabsorbing solvent this equation gives the absorbance at any wavenumber as:

$$A(\bar{\nu}) = -\log T(\bar{\nu}) = a(\bar{\nu}) bc \qquad \textbf{Eq. 1.10}$$

where $T(\bar{\nu})$ and $a(\bar{\nu})$ are the transmittance and absorptivity at $\bar{\nu}$, *b* is the pathlength, and *c*- the sample concentration, respectively.

and the transmission *T* is:

$$T = I \diagup I_0 \qquad \textbf{Eq. 1.11}$$

The light entering the sample (I_0) is determined from an interferogram measured without sample in the optical path, which yields the so called single beam background spectrum, and the intensity of the light transmitted by the sample (I) collected as an interferogram with a sample in the optical path, respectively, which yields the single beam sample spectrum. Transmission and absorption spectra are then calculated by means of equations Eq. 1.11 and Eq. 1.10, respectively.

1.3.4 Advantages of the FT-IR spectroscopy

In principle, the interferometer has several basic advantages over the classical dispersive instruments. In dispersive spectrometers the spectrum is measured directly by recording the intensity at different monochromator settings. In FT-IR, all frequencies emanated from the IR source impinge simultaneously on the detector. This accounts for the so called multiplex advantage. A complete spectrum can be collected very rapidly. Indeed, the measuring time in FT-IR depends on the time needed to move the mirror over a distance proportional to the desired resolution. As there is no grating or filter changes, there are no discontinuities in the spectrum.

The so-called Jacquinot or throughput advantage arises from the fact that the circular apertures used in FT-IR spectrometers have a larger area than the linear slits used in grating spectrometers, thus enabling a higher throughput of radiation. For the same resolution, the energy throughput in an interferometer can be higher than in a dispersive spectrometer, where it is restricted by the slits. In combination with the Multiplex Advantage, this leads to one of the most important features of an FT-IR spectrometer: the ability to achieve the same signal-to-noise ratio as a dispersive instrument in a much shorter time.

An additional strength of FT-IR spectroscopy is the speed advantage. The mirror has the ability to very rapidly move at short distances, and thus, makes it possible to obtain spectra on a millisecond timescale. However, the main advantage of rapid scanning instruments is the ability to increase the signal-to-noise ratio by signal averaging (Fellgett advantage). FT-IR spectrometers control the change in the optical path difference with high wavenumber accuracy (Connes advantage). The frequency scale of an interferometer is derived from a helium neon laser that acts as an internal reference for each scan. The frequency of this laser is known and is

very stable. As a result, the wavenumber calibration of interferometers is much more accurate and has much better long term stability than the calibration of dispersive instruments.

1.3.5 FT-IR spectroscopy of proteins

1.3.5.1 Amide modes

The IR infrared spectra of proteins exhibit absorption bands associated with their characteristic amide group, which is common for all molecules of this type (Figure 1.6). The characteristic bands of the amide groups of protein chains are similar to the absorption bands exhibited by secondary amides in general, and are assigned as amide bands. There are nine such bands, called amide A, amide B and amide I-VII, in order of decreasing frequency (see Table 1.II)

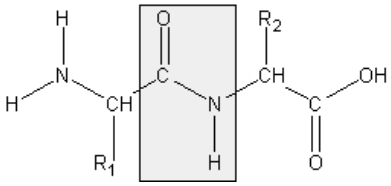

Figure 1.6 Amide group of proteins

The amide I band, which occurs between 1700 and 1600 cm^{-1}, represents mostly C=O stretching vibrations of the amide group. The stretching vibrations depend on the length and the strength of the bond. The exact frequency of this vibration depends also on the nature of the hydrogen bonding involving the C=O and N-H groups, the hydrogen bonds lower the frequency of the stretching vibrations, and the characteristic coupling between transition dipoles. The latter leads to characteristic splitting effects of the amide I band. The magnitude of this splitting depends on the orientation and distance of interacting dipoles and thus provides information about geometrical arrangements of peptide groups in a polypeptide chain. Proteins generally possess a variety of different conformations contained in different fragments. As a consequence, the observed amide I band is usually a complex composite, consisting of a number of overlapping component bands representing helices, β-sheet structures, turns, loops and random coils. Hydrogen-deuterium (H/D) substitution also leads to small band shifts of the amide I components (I′). The scale of the shift depends mainly on the type of the secondary structure.

The amide II band represents mainly the N-H bending, with some C-N stretching. As with the amide I band, it is possible to split the amide II band into components that depend on the secondary structure of the protein. The position of the amide II band is sensitive to deuteration. The N-deuteration converts the mode to predominantly C-N stretching vibrations and shifts from around 1550 cm^{-1} to a frequency of 1450 cm^{-1} when deuterated. The N-D bending vibration has a considerably lower frequency than the N-H bending vibration and thus no longer couples with the C-N stretching vibration. The amide II band of the deuterated proteins (II′) overlaps with the H-O-D bending vibrations. The latter event makes it difficult to extract structural information from the amide II′ band. However, the remainder of the amide II band at 1550 cm^{-1} may provide information on the accessibility of solvent to the peptide backbone.

Table 1.II Characteristic amide IR bands of proteins

Designation	Approximate frequency (cm^{-1})	Vibrational modes
A	~3300	N-H stretching in resonance with first amide II overtone
B	~3100	
I	1695-1610	80% C=O stretching; 10% C-N stretching; 10% N-H bending
II	1575-1480	60% N-H bending; 40% C-N stretching
III	1320-1220	30% C-N stretching; N-H bending; 10% C=O stretching; 10% O=C-N bending; mixed with other modes
IV	765-625	40% O=C-N bending; 60% other modes
V	640-800	Out-of-plane N-H bending
VI	605-535	Out-of-plane C=O bending
VII	~200	Skeletal torsion

(Fabian *et al.*, 2002)

Secondary amides in trans-configuration exhibit two characteristic bands at ~3300 cm^{-1} and 3100 cm^{-1}, while those in the cis-configuration show two bands at ~3200 and 3100 cm^{-1}. The band between 3300-3200 cm^{-1} is the amide A band and is due to the stretching mode of the N-H bond, which is engaged in hydrogen bonding. Its frequency depends on the strength of the hydrogen bond. The amide A band is usually part of a Fermi resonance coupled with the second component absorbing weakly between 3100 and 3030 cm^{-1} (amide B). In polypeptide helices, the NH stretching vibration is resonant with an overtone of the amide II vibration, in β-sheets with an amide II combination mode.

1.3.5.2 Analysis of the secondary structure of proteins

The amide I (I′) band of proteins and peptides consists of a series of overlapping component bands which occur as a result of the secondary structures present in such molecules. As a consequence, the individual component bands that represent different structural elements, such as α-helices, β-sheets, turns, loops and irregular structures are not resolved and are difficult to identify in the broad amide I band contours of the experimentally obtained spectra. Resolution enhancement of the amide I band allows identifying of the various structures present in a protein. Fourier self-deconvolution and derivative spectroscopy are used to identify the band components and to determine their frequency positions. Both methods do not increase the instrumental resolution, but are mathematical procedures that yield narrower component bands. The second derivative spectrum gives a negative peak for every band maximum or shoulder in the absorption spectrum. However, the derivatization process does not preserve the integrated areas of individual components. Additionally these techniques greatly amplify those features in the spectra originating from random noise or uncompensated sharp water vapour band components. Thus the resolution enhancement should be performed on spectra with a high signal-to-noise ratio and after complete elimination of water vapour bands.

The use of resolution enhancement spectra for estimation of the secondary structure components in proteins is an established technique (Susi *et al.*, 1983; Byler *et al.*, 1986; Dong *et al.*, 1990; Jackson *et al.*, 1995; Fabian *et al.*, 2002). The resolved band components are assigned to α-helix, β-sheet and turns (see Table 1.III).

Table 1.III Amide I(I′) frequencies for proteins secondary structures

Conformation	Frequency range cm^{-1}		
	Calculated	Experimentally obtained	
		H_2O	D_2O
Aggregated strands	1621-1627	1618-1628	1610-1620
β-sheet	1628-1640	1625-1640	1620-1635
Random coil	1641-1647	1652-1660	1640-1650
α-helix	1651-1657	1648-1660	1650-1658
Turns and bends	1651-1657	1660-1697	1655-1685
High frequency β-sheet	1658-1696	1675-1695	1680-1695

(Stuard 1997; Tamm *et al.*, 1997; Fabian *et al.*, 2002)

The assignment of the band components to particular secondary structure of proteins is based on theoretical calculations and on empirical spectral-structural correlations established for model polypeptides and proteins of known three-dimensional structure. However, such

secondary structure assignments are by no means exact and for certain proteins band components may appear outside the calculated coordinates (Tamm *et al.*, 1997). For example in highly solvent-exposed α-helices the amide I band can shift to 1630-1645 wavenumbers due to additional hydrogen bonding of the solvent accessible C=O groups to water (Surewicz *et al.*, 1993; Fabian *et al.*, 2002).

1.3.5.3 IR Absorption of biological molecules present in the prion aggregates

The absorption bands of the cell derived chemical compounds associated with the prion rods as lipids, carbohydrates and nucleic acids (see1.1.4.4), do not coincide with the amide modes of particular interest such as the secondary structure sensitive amide I neither with the sensitive to deuteration amide II or amide A regions (Table 1.IV).

Table 1.IV IR absorptions of cell derived chemical compounds co-purified with PrP27-30 aggregates

Chemical compound	Absorption band (cm^{-1})	Assignment
lipids	**~3030**	=C-H stretching in alkenes
lipids, proteins, carbohydrates, nucleic acids	**2955 and 2872**	asymmetric and symmetric C-H stretching vibration from CH_3 groups
lipids, proteins, carbohydrates, nucleic acids	**2923 and 2853**	asymmetric and symmetric C-H stretching vibration from CH_2 groups
phospholipids, thymine, uracil	**1721** **A (1500-1390)**	C=O stretching vibration (Esters)
	~1465	-CH_2 stretching
lipids, proteins, nucleic acids	**~1450**	-CH_3 asymmetric bending
proteins, lipids	**~1400**	C=O stretching vibration from $–COO^-$ groups -CH_3 symmetric bending
fatty acids, amino acid side chains	**~1390**	
	B (1250-900)	
proteins, lipids nucleic acids, phospholipids. polysaccharides	**~1250-1220**	Asymmetric P=O stretching in PO_2^- C-O-P- and C-O-C- stretching vibrations
polysaccharides	**~1200-900**	C-O-P- and C-O-C- stretching vibrations, ring vibrations
lipids	**~1170**	Asymmetric CO-O-C- stretching vibrations
nucleic acids, phospholipids	**~1080**	symmetric P=O stretching in PO_2^- CO-O-C symmetric stretching vibrations
lipids	**~1070**	symmetric CO-O-C stretching vibrations
nucleic acid backbone, carbohydrates, lipids	**~1050**	different overlaid bands from C-O stretching vibrations
proteins, lipids, nucleic acids	**700**	-CH_2 rocking

(Fabian *et al.*, 2000; Kneipp *et al.*, 2000; Fabian *et al.*, 2003)

However, their absorption components should be considered, too, since they could influence the protein structure and stability and the relative proportions between the protein and non-protein constituents might contribute to the strain characteristics of the prion rods.

1.3.5.4 Investigation of the structural stability of proteins by FT-IR

The amide I band of an aggregated protein is dominated by an intensive component at around 1620 cm^{-1} and a minor component at the high-frequency edge of the amide I region due to formation of intermolecular β-sheets. In contrast, an unfolded protein with a random coil structure exhibits only a broad amide I band centered at approximately 1654 cm^{-1} (in H_2O) or 1645 cm^{-1} (in D_2O). Thus the amide band of unfolded or aggregated proteins can be distinguished easily from the amide I band. Therefore, IR spectroscopy is well suited for protein stability studies. Folding and unfolding of proteins can be initiated in various ways: chemically, thermally and by changing of the pressure. The changes in the amide I bandwidth, of the wavenumber of maximum absorption, or of the absorbance at appropriate wavenumbers can be used to determine specific secondary structure transitions.

Another way to test the stability or flexibility of proteins is to monitor the H/D exchange level, investigating samples in D_2O environment. Amide and side chain hydrogen exchange rates depend on pH, temperature, and the protein environment. Groups exposed to the solvent exchange faster. The hydrogen atoms in a structured region of a protein exchange more slowly than the hydrogen atoms in an unstructured part. This is due to the hydrogen bonds, low solvent accessibility, and steric hindrance. Protected amide hydrogen protons can be regarded as “closed” to exchange. A transition to an “open” state is required to enable exchange with solvent. H/D exchange experiments can be used to compare the flexibility of different proteins (Backmann *et al.,* 1996; Barth *et al.,* 2002; Fabian *et al.,* 2002).

1.4 Cluster analysis

Cluster analysis is a multivariate data analysis procedure for classification of similar objects into different groups. In FT-IR spectroscopy it is used to arrange a set of spectra into clusters, based on the spectral distance, or the degree of similarity among the spectra. The aim is to establish a set of clusters such, that spectra within a cluster are more similar to each other than they are to spectra in other clusters. Hierarchical clustering builds a hierarchy of clusters from the individual elements and is represented graphically as a dendrogram.

The two key steps in cluster analysis are calculating and comparing the distances between the objects and grouping these objects according to their similarity:

- First, the spectral distances between all spectra are calculated
- The two spectra showing the greatest similarity (the spectra with the smallest distance) are merged into a cluster
- The distances between this cluster and all other spectra are calculated
- The two objects (spectrum/spectrum or spectrum/cluster) with the smallest distance again are merged into a new cluster
- The distances between this new cluster and all other objects (spectra, clusters) are calculated
- The two objects (spectrum/spectrum or spectrum/cluster or cluster/cluster) are merged into a new cluster
  This procedure is repeated until only one large cluster remains

For the calculation of the spectral distances, the so-called “Scaling to First Range algorithm” was used in the presented work. This algorithm has been elaborated for the identification of microorganisms on the basis of their FT-IR spectra (Helm *et al.,* 1991). First, the spectral differences are calculated as Pearson’s correlation coefficients *r*. This coefficient defines the correlation between two spectra as:

$$r = \sum a_n(k) \,.\, b_n(k) \qquad \textbf{Eq. 1.12}$$

where a_n and b_n are the normalized spectral intensities.

The correlation coefficient is calculated separately for each frequency range. The sum covers all data points *k* of a defined frequency range. The normalization yields:

$$r = \frac{\sum_{i=1}^{k} (a_{n,i} - a_m).(b_{n,i} - b_m)}{\sqrt{\sum_{i=1}^{k} (a_{n,i} - a_m)^2} . \sqrt{\sum_{i=1}^{k} (b_{n,i} - b_m)^2}} \qquad \textbf{Eq. 1.13}$$

where a_m and b_m are the mean spectral intensities, and *a(k)* and *b(k)* are the original spectral intensities

It can range from -1 to +1. Indicating negative (-) or positive (+) relationship between the spectra. The correlation coefficient is then transformed to spectral differences *D* according to the equation:

$$D = (1 - r).1000 \qquad \textbf{Eq. 1.14}$$

The spectral distance can vary between 0 (identical spectra) and 2000 (inverse spectra).

The Scaling to First Range algorithm calculates the spectral difference separately for each frequency range. Then an average value is calculated, for which each frequency range can be weighted differently:

$$D = \frac{\sum w_i . D_i}{\sum w_i} \qquad \textbf{Eq. 1.15}$$

For *n* number of spectra, a cross correlation of the distances between the spectra can be compared in a symmetrical matrix:

$$\begin{pmatrix} d_{11} & d_{12} & \dots & d_{1n} \\ d_{21} & d_{22} & \dots & d_{2n} \\ d_{31} & d_{32} & \dots & d_{3n} \\ \vdots & \vdots & \dots & \vdots \\ d_{n1} & d_{n2} & \dots & d_{1n} \end{pmatrix}$$

From equation Eq. 1.13 it follows that d_{ij} is the spectral distance between the spectra *i* and *j*, $d_{ij} = d_{ji}$ and $d_{ij}=0$ when $i=j$.

For clustering, Ward's Technique is used, since it produces homogeneous spectral groups. Neighbouring groups are merged in respect to the smallest growth in the heterogeneity factor (*H*). Thus, instead of the spectral distance, the growth in heterogeneity *H* is monitored according to the following formula:

$$H_{r,i} = D_{r,i} = \frac{(n_p + n_i).D_{p,i} + (n_q + n_i).D_{q,i} - n_i.D_{q,i}}{n + n_i} \qquad \textbf{Eq. 1.16}$$

n(i) is the number of spectra merged in the object *i*