UNIVERSITÀ DEGLI STUDI DI TRIESTE

Sede Amministrativa del Dottorato di Ricerca

XXIV CICLO

DOTTORATO DI RICERCA IN NEUROSCIENZE E SCIENZE COGNITIVE
INDIRIZZO NEUROBIOLOGIA

# Phage-display epitope library development for the biomarkers identification in autoimmune diseases of the Central Nervous System

Settore scientifico-disciplinare: BIO18

*Dottoranda*
Lorena Marson

*Coordinatore del collegio dei docenti*
Chiar.ma Prof. Laura Ballerini
Università degli studi di Trieste

*Supervisore-tutore*
Dott. Paolo Edomi
Università degli studi di Trieste

*Co-tutore*
Chiar.mo Prof. Roberto Marzari
Università degli studi di Trieste

Anno accademico 2010/2011

# Summary


## 1. Introduction

1.1 Immune System and Autoimmunity 1

1.2 Immune privilege of the Central Nervous System 4

1.3 Autoimmune neurological disorders 7

1.3.1 Autoantibody in Multiple Sclerosis 8

1.4 Biomarkers discovery 10

1.5 Display technologies 13

1.5.1 Bacterial surface display 15

1.5.2 Yeast display 20

1.5.3 Ribosome and mRNA display 24

1.5.4 Phage display 29

1.6 Phage display and epitope identification 33

1.7 cDNA library: Open Reading Frame (ORF) selection 35

1.8 Expression cDNA libraries: normalization 38

## 2. Aims 45

## 3. Materials and Methods

Bacterial strains 47

3.1 Plasmid pPE3 construction 47

3.2 Electrocompetent cells preparation 49

Genomic library

3.3 Preparation of genomic DNA from *Escherichia coli* One Shot® 49

3.4 Archeobacterial histones dimer production and purification 50

3.5 Genomic DNA digestion by nuclease and archeobacterial histones dimer protection 50
3.6 Genomic library construction 51
3.7 Library sequencing 52
Human Brain phage display library
3.8 Preparation of human brain cDNA 52
3.9 cDNA normalization 53
3.10 Testing the normalization process 55
3.10.1 Amplification of an high copy number gene: GAPDH 55
3.10.2 Normalized library construction 55
3.11 Normalized human brain cDNA digestion by nuclease and archeobacterial histones dimer protection 56
3.12 Normalized cDNA fragments library from human brain in pEP3 57
3.13 Subcloning of the ORF-enriched and normalized cDNA fragments library in phagmidic pDAN5 vector 58
3.14 Immunoglobulin purification from cerebrospinal fluid of Multiple Sclerosis patients 59
3.15 Selection of the ORF-enriched and normalized cDNA phage display library from human brain 59
3.16 Phage-ELISA 60
3.17 Secondary phage-ELISA 61

## 4. Results

4.1 Engineering of a new vectors to develop epitope libraries 63
Construction of ORF-selected library 66
4.2 DNA fragmentation system 67

4.3 Comparison of DNA protection using histones monomer and dimer 70

4.4 Preparation of genomic DNA and cloning 72

4.4.1 Genomic DNA library characterization and sequencing 74

4.5 Human Brain phage display library 76

First strategy 76

Second strategy 79

4.5.1 cDNA preparation and normalization 79

Test on DSN activity 80

Normalized library construction 82

4.5.2 Normalized cDNA fragmentation and HB library construction 85

4.5.3 Human Brain libraries subcloning in pDAN5 vector 88

Patients' data 90

4.6 Selections of Human Brain library 91

4.6.1 Pooling of MS CSF samples 92

4.7 HB selection against purified IgGs from CSF 93

4.8 Secondary ELISA on MS selected antigens 97

**5. Discussion** 99

**6. References** 109

# Riassunto

Mentre per alcune malattie autoimmuni del sistema nervoso vi sono chiare evidenze cliniche e sperimentali a sostegno della patogenesi, per la maggioranza l'autoimmunità è solo supposta, sebbene fortemente sostenuta dai caratteri infiammatori del danno tissutale, o da elementi paraclinici e strumentali, o dalla positiva risposta ai farmaci anti-infiammatori/immunosoppressori. Per individuare potenziali autoantigeni di malattie autoimmuni vengono utilizzati sia approcci proteomici, come l'elettroforesi 2-DE e la spettrometria di massa, che trascrittomici come l'analisi di microarray e le tecnologie di display library. La tecnologia del display è stata ottimizzata al fine di identificare proteine o peptidi (Bradbury et al. 2011; Lofblom 2011; Ullman et al. 2011; Beghetto and Gargano 2011
).

Il principale obiettivo del mio lavoro di ricerca è la messa a punto di un protocollo per la costruzione di librerie fagiche di frammenti di cDNA codificanti per frammenti ORF, e che quindi potrebbero corrispondere a potenziali epitopi. Questo tipo di librerie contengono, potenzialmente, tutto il repertorio ORF di una cellula o di un tessuto e possono quindi essere utilizzate nello studio di malattie autoimmuni al fine di identificate nuovi epitopi coinvolti nella risposta immunitaria, di fare un confronto tra lo stato patologico e quello sano o tra diverse condizioni patologiche (Puccetti and Lunardi 2010).

Per realizzare questo tipo di librerie è necessario disporre di un sistema in grado di selezionare i frammenti ORF, cioè codificanti. Abbiamo quindi messo a punto un complesso protocollo che prevede: (1) la normalizzazione del cDNA, (2) la sua frammentazione per ottenere peptidi di dimensioni opportune, e (3) l'arricchimento in frammenti realmente codificanti. Con questo sistema abbiamo realizzato una libreria di epitopi a partire da mRNA di cervello umano.

La normalizzazione risponde ad un tipico  problema nella costruzione di librerie di cDNA rappresentato dalla presenza differenziale dei messaggeri. A tale scopo abbiamo utilizzato l'enzima “Duplex Specific Nuclease” (DSN) che sfrutta la diversa cinetica di appaiamento del cDNA per attuare il processo di normalizzazione, cioè elimina dopo rinaturazione la frazione a doppio filamento formata dai trascritti più abbondanti. Per testare il protocollo di normalizzazione, è stata costruita una libreria normalizzate a partire da mRNA poli$A^+$ di cervello umano. Sono stati scelti

casualmente 96 cloni e sequenziati. L'analisi bioinformatica delle sequenze ha messo in luce che, tra i 72 geni mappati, il 90% sono geni unici. Questo risultato indica che il protocollo di normalizzazione ha funzionato.

Un altro aspetto da considerare nella costruzione di librarie è la lunghezza dei frammenti di cDNA da clonare. Le librerie in phage display di epitopi possono essere utilizzate sia per identificare epitopi lineari che conformazionali (Mackay and Rowley 2004). Per frammentare il DNA abbiamo deciso di utilizzare un sistema messo a punto nel laboratorio dove è stato svolto questo lavoro di Tesi (Azzoni e colleghi 2007) che si basa sulla protezione del DNA dall'azione di digestione dell'enzima *MNasiI* grazie al monomero istonico di archeobatteri ipertermofili da *Methanothermus fervidus* (HMF). Più precisamente, durante il mio dottorato, è stato introdotto l'utilizzo di una forma dimerica ricombinante, al fine di rendere più riproducibile e più controllabile la protezione del DNA. Confrontando i risultati ottenuti utilizzando la forma monomerica e quella dimerica, si è visto che le reazioni di digestione sono molto più controllabili e il recupero di DNA consente rese più alte. Variando il rapporto DNA:istoni, è inoltre possibile modulare la lunghezza dei frammenti di DNA che si possono ottenere. Il cDNA di cervello umano, precedentemente normalizzato, è stato sottoposto alla digestione con l'enzima *MNasiI* e protezione con istoni, e sono state ottenute due popolazioni di frammenti, da 60 pb a 300 pb e da 300 pb a 600 pb.

I frammenti sono stati clonati, separatamente, in un vettore, pEP3. Questo vettore è stato costruito durante il mio lavoro di Tesi, a partire dalla struttura di pEP2 (Bembich 2004), al fine di aumentare l'efficienza di clonaggio. Questo vettore consente di selezionare i frammenti ORF conferendo resistenza all'Ampicillina ai soli cloni che contengono frammenti in frame con il gene della β-lattamasi. Le due librerie normalizzate, frammentate e arricchite in ORF, clonate nel vettore pEP3, presentano rispettivamente una dimensione di $5.6x10^5$ e di $7.8x10^4$ cloni ORF.

Il protocollo, sia di frammentazione che di selezione delle sequenze ORF, è stato inizialmente testato costruendo una libreria di DNA genomico totale di *E. coli*. L'analisi di 93 sequenze casuali ha messo in luce che l'87% delle sequenze corrispondono a ORF. Il sistema è quindi in grado di selezionare in modo efficiente i cloni "in frame". Inoltre, la libreria presenta una copertura del genoma pari a 2.25x con il 90% di probabilità di includere tutte le sequenze.

Considerando le due popolazioni di frammenti normalizzati di cervello umano e clonati in pEP3, è più probabile che gli epitopi conformazionali siano più abbondati

nella libreria che comprende i frammenti da 300 pb a 600 pb. Questi frammenti sono stati sub clonati nel vettore fagmidico pDAN5, ottimizzato per la produzione di fagi. Considerando che in una cellula umana sono espressi da 10.000 a 15.000 trascritti (Jongeneel et al. 2003) e che la lunghezza media della parte codificante è di 1186 pb (Progetto MGC squadra 2004), una libreria di cDNA arricchito in ORF con una dimensione di $7.8x10^4$ cloni è sufficiente per fornire una copertura di 2.5 volte dei trascritti. Ciò corrisponde ad una probabilità del 92% di includere una particolare sequenza.

Utilizzando questa libreria è stato effettuato un primo biopanning, utilizzando tre pool, parzialmente sovrapposti, di IgG purificate dal liquor di pazienti con Sclerosi Multipla. I primi risultati ottenuti sono incoraggianti. Di 92 cloni testati da ogni selezione, circa il 50% ha mostrato un alto riconoscimento (D.O.>0.5) in test di phage ELISA. Tra i cloni più reattivi, alcuni sono stati testati in saggi di phage ELISA secondari mostrando alta specificità (98%) rispetto ad una bassa sensibilità (8%). Queste osservazioni, puramente speculative, confermano l'idea che la risposta autoimmune nella Sclerosi Multipla è talmente complessa da richiedere un'analisi diagnostica che comprende più marcatori contemporaneamente.

# Abstract

Only for some autoimmune diseases of the nervous system there are strong clinical evidences to support the autoimmune pathogenesis, whereas for the majority the autoimmunity is only supposed, although supported by inflammatory tissue damage, paraclinical and instrumental elements, or the positive response to anti inflammatory/ immunosuppressive drugs. To identify potential autoantigens in autoimmune diseases are use both proteomic approaches, such as 2-DE electrophoresis and mass spectrometry, and transcriptomic approaches such as microarray analysis and library display technologies. The display technology have been largely improved and successfully employed in affinity peptides or proteins identification and searching (Bradbury et al. 2011; Lofblom 2011; Ullman et al. 2011; Beghetto and Gargano 2011 ).

The principal aim of my PhD was the setting of a protocol for the creation of phage libraries to display cDNA fragments encoding real ORF sequences, that could correspond to potential epitopes. A similar phage display library contains all the potential ORF repertoire of a cell or tissue. This tool can be specially used in the study of autoimmune diseases to perform different kind of analysis, such as the identification of epitopes involved in pathological reaction, the comparison between healthy and pathological conditions, or between different pathological conditions (Puccetti and Lunardi 2010).

To create this kind of libraries, the development of a system for ORF fragment selection is essential. During my PhD, a complex protocol was developed. It provides for: (1) cDNA normalization, (2) cDNA fragmentation to obtain peptides with useful size, and (3) ORF enrichment to obtain really coding fragments.

The most common problem in the construction of cDNA libraries is represented by the relative abundance of the transcripts. For this reason, a normalization step was introduced using the "Duplex Specific nuclease" enzyme that cut the most abundant transcript, exploiting the different kinetics of annealing of the cDNA. To check the normalization protocol, a normalized library from Human Brain in pBluescript was constructed and 96 clones, randomly chosen, were sequenced. The bioinformatic analysis indicated that the normalization process can be considered successful, in fact, among the 72 mapped genes, 90% were unique.

One of the major issue to consider in library construction is the cDNA fragments size to clone. Epitope phage display libraries can be employed to identify both linear and conformational epitopes (Mackay and Rowley 2004). For this reason, we decided to adopt a system previously set up in my hosting laboratory (Azzoni and colleagues 2007) and based on the digestion with *MNase*I and protection with archeal histones monomer from *Methanothermus fervidus* (HMf). During my PhD we produce a recombinant form of a covalent dimer of histone, in order to make the DNA protection more reproducible. Our hypothesis was confirmed by the comparison between the use of the monomeric and the dimeric form: the digestion reactions are more controlled, and this affect also the yield of DNA after the post-reaction recovery. Further, varying the DNA:histones ratio it is possible to modulate the length of DNA fragments to be obtained. The normalized HB cDNA was submitted to *MNase*I digestion and fragments with useful size were obtained, in particular, two fragments population, one from 60 bp to 300 bp and the other from 300 bp to 600 bp.

The obtained fragments were cloned, separately, into a specific vector called pEP3, derivative form pEP2 (Bembich 2004) specifically modified to improve cloning efficiency. It allows the selection of ORF fragments, in fact only the clones containing an ORF fragment were able to survive in presence of Ampicillin whereas, clones with “out of frame” fragment were suppressed because of Ampicillin toxicity. The libraries showed, respectively, a total dimension of $5.6x10^5$ and $7.8x10^4$ ORF clones.

The procedure of DNA fragmentation and ORF selection was initially tested on the total genomic DNA of *E. coli*. The mapping of 93 randomly chosen sequences showed that 87% correspond to ORFs demonstrating the capability of the system of select efficiently “in frame” clones. The ORF genome representation obtained with a dimension of $8x10^4$ ORF clones can be estimated considering that the ORF sequences in *E. coli* genome amount for approximately to $4x10^6$ bp (4290 ORFs of 951 nt of medium length) (Blattner et al. 1997) and that the fragments have a medium length of 120 bp: with a 2.35x coverage there is the 90% of probability to include all sequences.

Considering the two HB fragments population, probably conformational epitopes can be more abundant in the “high size fragments” library. This library, with fragments between 300 bp and 600 bp, was sub-cloned into the phagemid pDAN5 for phage production. In a human cell are expressed from 10.000 to 15.000 transcripts (Jongeneel et al. 2003) and that the medium length of the coding part is 1186 bp (the MGC Project Team 2004), an ORF enriched cDNA library of $7.8x10^4$ clones is

sufficient to provide a 2.5-fold clone coverage of the transcripts present corresponding to a 92% of probability to include a particular sequence.

A first biopanning with this library was performed using three partially overlapped pools of purified IgGs from CSF of MS patients. The first results are encouraging. Of 92 clones tested for each selection, about 50% showed a high recognition (O.D.> 0.5) in Phage ELISA assays. Among the clones that exhibited the greatest recognition, some have been tested in secondary Phage ELISA. They showed high specificity (98%) compared with a low sensitivity (8%); these observations are purely speculative, but, in general, confirm the idea that the autoimmune response in MS is so complex to require a diagnostic analysis that includes multiple markers simultaneously.

## 1.1 Immune System and Autoimmunity

Vertebrate immune systems have evolved sophisticated genetic mechanisms to generate T-cell receptor and antibody repertories, which can be considered as “combinatorial libraries” of affinity molecules capable of distinguishing between self and non-self. This system protects vertebrates against environmental foreign agents, including microorganisms (bacteria, viruses, fungi and parasites), chemicals and allergens. Recent data highlights the delicate balance in higher mammals between robust immune defence against pathogens and autoimmunity (Graham et al, 2010). If this delicate balance fails, loss of tolerance to self antigens can occur. This condition represents the first step to develop an autoimmune reaction that leads to develop autoimmune diseases (Larman et al. 2011).

The immune system is divided into innate and adaptive system, although in reality they are highly integrated and interdependent. The innate immune system is philogenetically older and is designed for immediate engagement of pathogens by highly conserved pattern-recognition receptors, such as Toll-like receptors, coupled with a prompt defensive response by the cell. In contrast, the adaptive immune system consists primarily of T and B cells, which use a highly specialized receptor system selected somatically for antigen recognition (T-cell receptor and surface immunoglobulin, respectively) that can recognize millions of distinct foreign antigens. Another feature of the adaptive immune system is the formation of immunologic memory. These characteristics immediately raise the problem of selecting functional receptors that do not lead to uncontrolled self-reactivity.

The original idea of autoimmunity derives from Paul Ehrlichs’ realization that a functional immune system must have “horror autotoxicus”, which he conceived as having “certain contrivances” that would prevent immune attacks against the self (Silverstein 2005). Nowadays, it is known that self-reactive B and T cells are a normal component of the immune system, but

they are kept in check by a variety of mechanisms. Alteration at this "checkpoint" lead to autoimmunity (De Jager et al. 2009). Some are central mechanisms in the thymus and bone marrow that delete or disable self-reactive clones; others are peripheral and include specialized regulatory cells, such as regulatory T cells (Wing and Sakaguchi 2010).

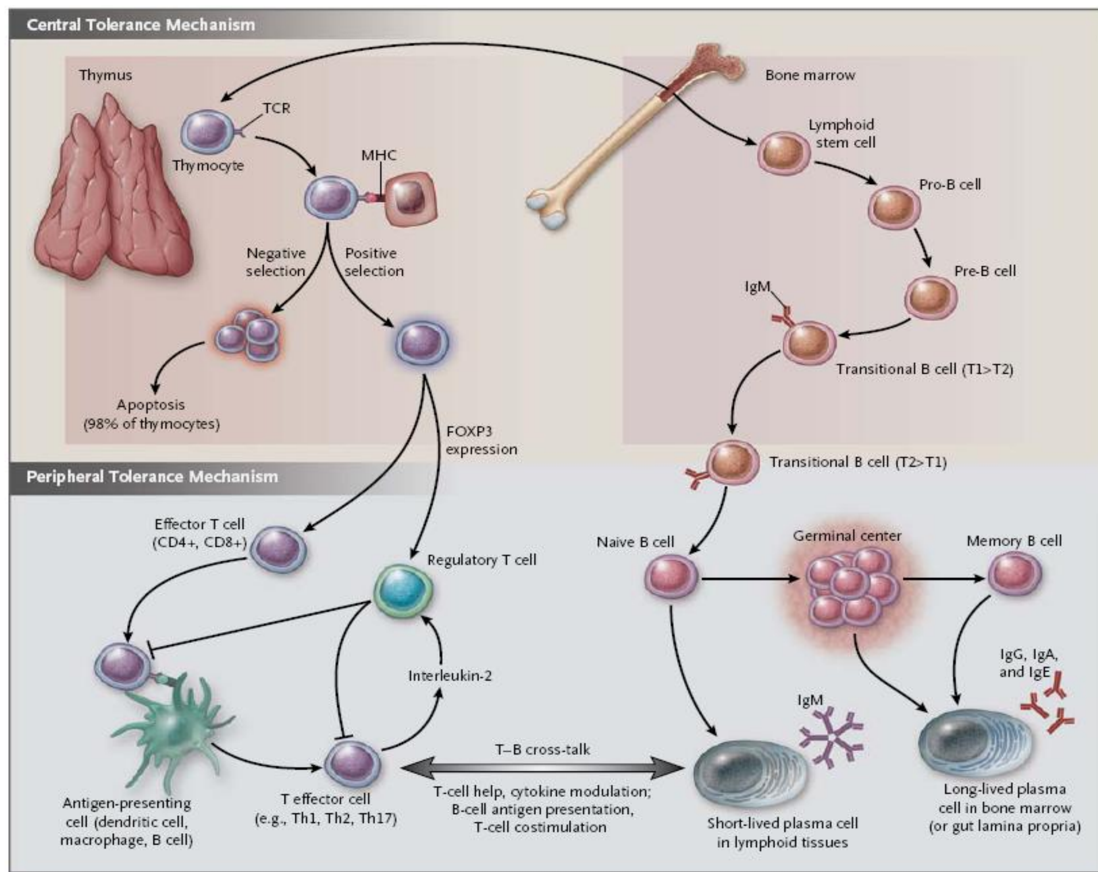



**Fig. 1.1** Central and Peripheral tolerance mechanisms in the adaptive Immune System. Selection against self reactivity in developing T cells occurs in thymus, where more than 98% of developing thymocytes die from apoptosis because of excessive reactivity to self-peptides bound to majority histocompatibility complex (MHC) molecules, followed by positive selection for functionally competent T cells (CD4+ and CD8+) that are released into the periphery. The expression of self-antigens in the thymus is genetically regulated by transcription factors, such as autoimmune regulator, or by genetic variation in self antigens themselves (*e.g.*, insulin). The production of peripheral regulatory T cells (Tregs) is also under genetic control, exemplified by the transcription factor FOXP3, the absence of which leads to severe autoimmunity. Alterations in genes affecting these various pathways may lead to quantitative as well as qualitative differences in the potential for self-reactivity of the repertoire of mature T-cell receptors (TCRs). An analogous process of selection against self-reactivity by B cells occurs in the bone marrow, where self reactivity is dramatically reduced as B cells transition out of the bone marrow into the peripheral B-cell population. Peripheral mechanisms for preventing self-reactivity also exist. In this context, Tregs play key a role in T cell, where genetic alterations in interleukin-2 pathways may influence the efficiency of Treg regulation. Multiple additional peripheral mechanisms contribute to keeping the immune response under

control during the activation of both B an T cells in the peripheral immune system, including extensive cross-talk between T cells and B cells, as well as interactions with the innate immune system (Cho and Gregersen 2011).

Several mendelian disorders directly corroborate the importance of these mechanisms. For example, mutation affecting the transcription factor autoimmune regulator, lead to a declining of selection against self-reactivity by T-cells in the thymus, giving rise to a rare, aggressive autoimmune disease, autoimmune polyendocrine syndrome 1 (Shikama et al. 2009). The autoimmune regulator controls the ectopic expression of self-antigens within the thymus (Guerau-de-Arellano et al. 2009) and thus is critical to the negative selection of T-cells reactive with these antigens.

In addition to this defect, in central tolerance, a loss of the FOXP3 transcription factor in the mendelian disorder IPEX (immune dysregulation, polyendocrinopathy, enteropathy, X-linked) (Bennett et al. 2001) causes aggressive autoimmunity as a result of defects in the function of regulatory T cells.

Analogous control mechanisms are active at numerous checkpoints during the B-cell formation within the immune system (von Boehmer and Melchers 2010). For example, pre-B cells in the bone marrow are highly autoreactive but become less autoreactive during differentiation into naive B cells in periphery, a process that is influenced by the gene encoding protein tyrosine phosphatise non-receptor type 22 and other genes associated with autoimmunity (Menard et al. 2011).

Overall, these processes of selection and regulation of T and B cells are controlled by cell-signalling events that are normally active within a range that may vary among people and among cell types, owing in large part to genetic diversity in the population. This leads to a general concept of immune responsiveness and regulation as a trait that exists on a continuum (a quantitative trait), setting thresholds for cell activation and response (Liston et al. 2005). Indeed, the original discovery of MHC (Major Histocompatibility Complex), which encodes HLA (Human Leukocyte Antigen), as a locus controlling immune responses was described as a quantitative trait. HLA-

regulated immune responses are generally high or low, as opposed to just absent or present, and responsiveness can vary among people.

After the induction, the autoimmune reaction is usually self sustained, leading to a chronic and definitive impairment of the target tissue. The damaging immune response can be organ-specific as well as systemic. When the response is targeted to an antigen expressed only in one cellular type, the immune aggression can bring to a complete and irreversible loss of function of the targeted tissue (as in type 1 or insulin dependent diabetes – IDDM) or to a hyperstimulation or inhibition of its function (as in Graves' hyperthyroidism and in myasthenia gravis). In other cases, the response seems to be directed against antigens which are not cellular type-specific, but widely expressed; in these cases the pathogenic events are multiple and complex, leading to impairment or destruction of several tissues at the same time (as in Systemic Lupus Erythematosus – SLE).

Today autoimmune diseases are estimated to afflict more than 5% of the population worldwide (Bright 2007) and for most of these diseases the etiology is still unknown. The identification of both B- and T-cell epitopes is a crucial step for the understanding the immune response mechanisms and their role in autoimmune diseases.

## 1.2 Immune privilege of the Central Nervous System

Numerous sites in the body possess varying degrees of immune privilege, including the brain, the anterior chamber of the eye, pregnant uterus, hair follicles and hamster cheek pouch. The advantage of an immune privilege for the tissue is that the damage generated during a normal immune response is attenuated and non-renewable tissues, *e.g.* brain, are protected (Forrester et al. 2008).

The Central Nervous System (CNS) is comprised of the brain and spinal cord, surrounded by three layers of meningeal membranes. The Blood Brain Barrier (BBB) is a feature of the cerebral vasculature, which restricts access of ions and other solutes present in the blood into the brain parenchyma. The anatomical structure of the BBB, as shown in figure1.2, comprises two cell

layers, which are separated by the perivascular space. One is formed by endothelial cells lining the brain capillaries and an underlying basement membrane, and the other is formed by astrocytic foot processes and their parenchymal basement membrane.

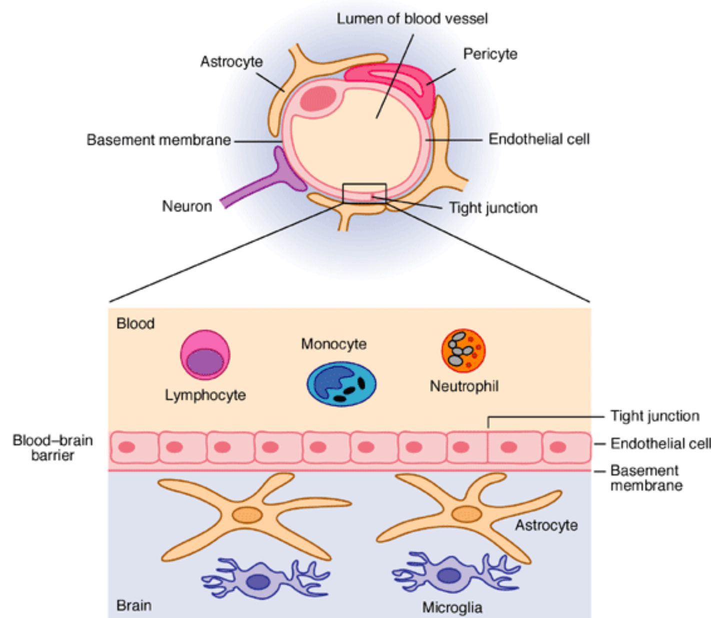



**Fig. 1.2** The Blood Brain Barrier (Expert Reviews in Molecular Medicine, 2003). As shown on the picture, BBB is created by the tight apposition of endothelial cells lining blood vessels in the brain, forming a barrier between the circulation and the brain parenchyma (astrocytes and microglia). Blood-borne immune cells such as lymphocytes, monocytes and neutrophils cannot penetrate this barrier. A thin basement membrane surrounds the endothelial cells and associated pericytes, and provides mechanical support.

Unlike other tissues, the endothelial cells of the BBB display no fenestration and are connected by tight junctions, which efficiently restrict the traffic of molecules and cells in and out of the brain. The cerebrospinal fluid (CSF) bathes the brain and it is produced from arterial blood by the choroid plexus. It flows from the ventricles of the brain into the subarachnoid space located between the arachnoid and the pial membrane and is eventually absorbed into the venous circulation. The CSF communicates with the interstitial fluid of the brain through the perivascular spaces. Due to the lack of tight junctions in the ependymal linings of the ventricles, small hydrophilic molecules as well as protein diffuse freely between the CSF and the brain interstitium (Ransohoff et al. 2003).

It is believed that several mechanisms are involved in immune privilege. First, the tight junction between vascular endothelial cells in the brain creates a BBB that retards extravasation of leukocytes into the brain. Second, the absence of lymphatic vessels prevents antigens from leaving the brain and reaching regional lymph nodes (Kaplan and Niederkorn 2007). Third, the immune responses cannot develop in the CNS because only few resident cells constitutively express MHC molecules in the steady state. Fourth, local tolerogenic mechanisms exist within the CNS (Cassan and Liblau 2007). It was shown that multiple cells in the CNS such as astrocytes, oligodendrocytes, microglia and the vascular endothelium express FasL (Choi and Benveniste 2004). It is believed that endothelial cells in the CNS reduce the risk for inflammation by expressing FasL, which limits extravasation of inflammatory cells (Walsh and Sata 1999). Additionally, CNS expression of $PGE_2$, TGF-β and galectin-9 is associated with functional silencing of incoming T lymphocytes (Khoury et al. 1992; Mannie et al. 1995; Zhu et al. 2005).

Presently, there are several lines of evidence that indicate that the immune privilege of the CNS is not absolute. First, access of T lymphocytes to CNS is limited and involves active transendothelial migratory process but is not completely forbidden (Cassan and Liblau 2007). It was already shown that the endothelial cells of BBB having only limited expression of endothelial P-selectin, E-selectin and VCAM-1 are not resistant to the development of immunopathology once inflammation within the organ itself has begun. Additionally, there is strong evidence that both naive $CD4^+$ and $CD8^+$ T lymphocytes are able to patrol non-lymphoid tissues, including the CNS (Cose et al. 2006). However, while native T lymphocytes can circulate in the CNS without triggering a deleterious response, activation, for example, of myelin-specific T cells is not always sufficient to allow self-reactive T lymphocytes to enter the CNS and additional signals are required (Cassan and Liblau 2007). Second, there is substantial, lymphatic drainage connecting the meninges and ventricular system, if not the brain parenchyma, directly through the cribiform plate to the deep cervical lymph nodes (Forrester et al. 2008). It was also

shown that antigens escape the CNS and accumulate in cervical lymph nodes where they induce a form of immune deviation (Wenkel et al. 2000). This process occurs not only in the brain but also in eyes and fetoplacental unit within the pregnant uterus. These sites contain unique fluids with suspected immunoinhibitory properties. Aqueous humor, which is normally present within the anterior chamber (AC) of the eye, has been shown to suppress antigen-driven T cell activation, and to contain significant amounts of transforming growth factor β-2 (TGF-β). Antigens injected into the AC of normal mice induce a deviant form of systemic immunity, termed anterior chamber-associated immune deviation (ACAID), which is characterized by a selective inability to display antigen-specific delayed hypersensitivity. The immune privileged states of the eye, the brain, and the fetoplacental unit share common features, and possess unique fluids with a similar capacity to force macrophages to present antigens in a "deviant" manner. This capacity is mediated, at least in part, by TGF-β. It is believed that brain-associated immune deviation contributes to the immune privilege of the brain wich reduces the risk for immune-mediated inflammation in the CNS (Wilbanks and Streilein 2005) . Third, antigen presentation may occur in the CNS. It was shown that oligodendrocytes and neurons exposed to proinflammatory environment express MHC I, whereas astrocytes and microglial cells express MHC II. So, while the CNS is not favourable for development of immune responses, under inflammatory conditions, T-cell mediated responses can develop within this tissue.

## 1.3 Autoimmune neurological disorders

Autoimmune reaction in the nervous system may concern all level of the neuraxis including brain and spinal cord (as in Multiple Sclerosis, neuromyelitis optica or Devic's disease, stiff-person syndrome and paraneoplastic neurological syndrome), dorsal root ganglia and peripheral nerves (in the case of Gullain Barrè syndrome - GBS and chronic demyelinating neuropathies), neuromuscular junction (as in myasthenia gravis) and muscles (as in the case of dermatomyositis). Emerging data from animal and human studies have renewed interest in the importance of B cells

in the pathophysiology of autoimmune neurological disorders (Dalakas 2006). The number of autoimmune diseases associated with the presence of autoantibodies directed against cells of the target tissue has been growing extensively over the past decade (Archelos et al. 2000; Leslie et al. 2001; Sherer et al. 2004; Rott and Mrowietz 2005). Autoimmune diseases are, in fact, the result of specific immune responses directed against “self” structures (Burnet 1963). All the above autoimmune diseases present an important involvement, more or less consistent, of clonally expanded B lymphocytes with intrathecal immunoglobulin synthesis, implicated in the pathogenesis of the neurological autoimmunity.

The presence of autoantibodies is a hallmark of many autoimmune diseases and has long been used for the diagnosis and classification of these diseases. Autoantibodies may exist years before the diagnosis and could be used for early prediction of the onset of the disease. The most widely used biomarkers are serum, and eventually CSF, immunoglobulin G (IgG) autoantibodies (Hu et al. 2011), that are detectable using a variety of different techniques, including Enzyme Linked Immunosorbent Assays (ELISAs), Western blot analysis, immunoprecipitation analysis, flow-based assays, and protein arrays (Robinson 2006).

### 1.3.1 Autoantibody in Multiple Sclerosis

Multiple Sclerosis (MS) is the most common neurological disease in young adults. Typically the first symptoms of MS occur between the ages of fifteen and fifty; females are affected twice as often as males (Alonso et al. 2008). The etiology of MS is still unknown, but many findings indicate a central role for the immune system in the disease pathogenesis, and both genes and environmental factors influence the risk of developing disease (Hemmer et al. 2006). This disease is characterized by discrete regions of CNS inflammation, lymphocyte infiltration, demyelination, axonal damage and the death of myelin-producing oligodendrocytes.

MS is a complex disease and its origin is still unknown, but the most diffuse opinion is that it derives from the contribution of multiple factors: a specific

genetic background, in presence of particular environment determinants (possibly including infectious agents) and disregulation of the immune response (as the lack of suppression of T and B autoreactive lymphocytes), can bring to the development of the autoimmune reaction. The targets of the autoimmune response in MS are believed to be cellular components of the CNS that are normally inaccessible to the immune system because of their location behind the BBB. In MS, the immune response presents both a cellular and a humoral component. Until some years ago, most studies had emphasized the role of T cells in the pathogenesis of MS (Krogsgaard et al. 2000). Especially in the last decade, several data have demonstrated a strong implication of B cells in the development of the disease (Oh et al. 2008; Racke 2008).

The first evidence of an association between MS and B-cell was highlighted in 1950 when intrathecal immunoglobulins synthesis in MS patients was observed (Kabat et al. 1950). There are some possible ways through which the lymphocytes B could enter into the CNS, reach the parenchyma and give rise to intrathecal immunoglobulins synthesis. Circulating B cells, after differentiation in the germinal center of peripheral lymphoid organs, are able to enrich the inflamed CNS as plasmablast (Odendahl et al. 2005) or as memory B cells. In this last case, memory B cell are able to differentiate into antibody- secreting cells. This differentiation occurs in response to antigen, outside or inside of follicle-like aggregates in the meninges. The antigen-driven B-cell activation inside of the CNS is an early event in the pathogenesis of MS, in fact, sequences analysis of rearranged immunoglobulin genes in CSF B cells indicate that this activation occurs in MS patients early after onset of the disease (Monson et al. 2005). The hypothesis of B-cell differentiation in follicle-like structures of the CNS is also supported by flow cytometry analysis, which detected B-cell differentiation stages in the CSF of patients with MS and other inflammatory neurological diseases (OIND) (Uccelli et al. 2005). Another alternative hypothesis is that memory B cells can differentiate to plasmablast in a bystander reaction with a T-cell help, but this activation can explain just a small part of immunoglobulins production in the inflamed CNS.

During the disease progression, a subset of plasmablast may develop to plasma cells in the inflamed CNS and under appropriate survival conditions (cytokines presence) can persist and produce IgGs in an antigen independent manner. This mechanism leads to the development of Oligoclonal Bands (OCBs), a key feature of MS (Manz et al. 2005).

Autoantibody profiling may serve different purposes including classification of individual patients and subsets of patients based on their 'autoantibody fingerprint', examination of epitope spreading and antibody isotype usage, discovery and characterization of candidate autoantigens, and tailoring antigen-specific therapy. Proteomics technologies, that are employed for large-scale study of expression, function and interactions of proteins (Geysen et al. 1984), enable profiling of autoantibody responses using biological fluids derived from patients with autoimmune disease. They provide a powerful tool to characterize autoreactive B-cell responses in diseases including Rheumatoid Arthritis, Multiple Sclerosis, Autoimmune Diabetes, and Systemic Lupus Erythematosus (Hueber et al. 2002).

## 1.4 Biomarkers discovery

In recent years it has been observed a vast expansion of the biomedical scientific literature in which the term "biomarker" is used. A biomarker is a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes or pharmacological responses to a therapeutic intervention (Lesko and Atkinson 2001; Rolan et al. 2003).

Biomarkers have an important influence on the clinical decision-making processes involved in diagnosis, assessment of disease activity, allocation of treatment, and determining prognosis. The clinical usefulness of a biomarker is dependent on demonstration of its validity. Ideally, biomarkers should provide information not available from currently available tests and should be tested as they would be used in clinical practice; however, potential biomarkers could be affected by many different clinical or patient variables, such as disease activity, therapeutic intervention, or the presence of

comorbidities. Validation studies might not include all the design features that are required to ensure that the biomarker is a true measure of the clinical process it is intended to reflect (Tektonidou and Ward 2011).

A large number of studies are directed to identify protein biomarkers for diagnosis and prediction in the clinical setting, disease severity, progression to disability, and response to therapy (O'Dell 2004; Scofield 2004). It is generally accepted that a single marker is unlikely to serve as a general diagnostic or prognostic tool for the diseases in which are observed heterogeneous genetic background and immunopathogenetic subtypes, various clinical disease courses, different and unpredictable therapeutic effects, as for example occurs in autoimmune diseases (Hueber and Robinson 2006). Therefore, the development of a panel of biomarkers, could be important for the understanding of pathogenesis, classification, diagnosis and therapeutic applications.

Although remarkable progress toward understanding immune function has been made over the last four decades in term of the role of the major histocompatibility complex and the nature of lymphocyte antigen receptors that confer specificity to autoimmune responses, understanding of the dysregulation and autoimmune response specificity remains limited. For certain autoimmune diseases, including Sjögren's syndrome and Systemic Lupus Erythematosus, candidate autoantigens have been identified but their exact roles in the initiation, perpetuation, and pathophysiology are not well understood (Guggino et al. 2011; Ice et al. 2011). For other autoimmune diseases, including Rheumatoid Arthritis and Psoriasis, the target autoantigens remain unidentified or unaccepted despite extensive experimental efforts (Besgen et al. 2010; Oh et al. 2010).

Regarding the Multiple Sclerosis disease, the importance of identifying biological markers is continuously evolving, particularly because of the heterogeneity of immune response in MS patients (Reindl et al. 2006; Menge et al. 2007). Whereas several measures on conventional MRI enable clinicians to identify the disease and its stage, there are no accepted biological markers for the disease activity in MS (Polman and Killestein 2006;

Svejgaard 2008). Proteins of the myelin sheath have been identified as target of the immune response, among these, the most important and investigate are the Myelin Basic Protein (MBP) and the Myelin Oligodendrocyte Glycoprotein (MOG). Berger et al. (2003) have demonstrated that MS patients with clinically isolated syndrome (CIS) seropositive for anti-MOG and anti-MBP antibodies were more likely to suffer a relapse than seronegative patients. More recently, Rauer et al. (2006) reported that 31/45 CIS patients (69%) were seropositive for anti-MOG or anti-MBP, confirming the above mentioned data. A cell based assay that specifically measures antibodies directed against cell membrane expressed human MOG has been described (Lalive et al. 2006); in this study native MOG-specific IgGs were most frequently found in serum of CIS and RR-MS, only marginally in secondary progressive MS, and not at all in primary progressive MS. Instead, another study did not find any associations between the presence of anti-MOG and anti-MBP IgM and IgG antibodies, detected by Western blot analysis, and progression to clinically definite MS or a diagnosis of MS according to the McDonald criteria (Kuhle et al. 2007). On the other hand anti-myelin antibodies show a prognostic value according to Poser's criteria, but did not according to the McDonald's criteria (Tomassini et al. 2007). Therefore, the diagnostic value of serum antibodies against MOG and MBP, to predict a risk of progression to clinically definite MS in patients who have had a clinically isolated syndrome, is at the moment controversial. Apart from these well known myelin autoantigens, some non-myelin CNS antigens were investigated as potential biomarkers for MS. For example, recently, a greater prevalence of positive T-cell proliferative responses to NSE and arrestin in MS patients was reported (Forooghian et al. 2007). In general, also among non-myelin CNS protein there are no confirmed diagnostic markers for the diagnosis of MS. Recently the attention of the researchers has been focussed on the role of EBV (Lunemann et al. 2007) and several studies exist also on the potential diagnostic value of markers of viral origin (Jarius et al. 2009). Even for these markers, the data is absolutely not conclusive.

Conventionally, the study of autoimmune response has been conducted by analyzing the presence and/or concentration of individual antibodies in biological fluids. New proteomic techniques allow the simultaneous identification/measurement of different autoantibodies in sera of patients with autoimmune diseases. The possibility of simultaneously measuring a number of correlated analytes appears to be very interesting for analytical reasons (reduced volumes of biological samples, reagents and low costs), logistical/managerial reasons, and pathophysiological reasons (combination of markers in disease-oriented or organ-oriented profiling) (Plebani et al. 2009).

The ideal assay for detecting protein and their interactions should be sensitive, specific and reproducible. Among the several functional proteomic technologies, those more frequently applied to autoimmunity are: display technologies (phage-, bacterial-, yeast-, ribosome-, etc), two dimensional gel electrophoresis and mass spectrometry for autoantigen discovery; autoantigen microarray to characterize autoantibodies response; and antibodies microarray to profile cytokines and other biomolecules.

Until now peptide-based research has been important in attempts to identify autoantigens in MS (Alcaro and Papini 2006): both selecting on serum or CSF antibodies (Cortese et al. 2001) and on recombinant antibody from single cell (Yu et al. 2006) random peptide libraries have been always used. Very recently the use of a phage display library derived from MS brain plaques for a serological Ag selection was reported (Somers et al. 2008) but only one potential antigen was identified.

## 1.5 Display technologies

Display technology refers to a collection of methods for creating libraries of biomolecules that can be screened for specific properties. It has become a routine tool for enriching molecular diversity and producing novel types of protein (Li 2000). The ability to link a protein’s function to the gene encoding that protein using the so-called ‘display technologies’ has become an essential means to identify proteins with desired properties from large libraries and optimize their properties (Daugherty 2007).

The displayed protein, consequently, can be identified through a simple DNA sequencing reaction and a bioinformatics analysis. The most important characteristics of the display technologies are: (I) the possibility to produce a collection of billions of different particles displaying different polypeptides and (II) the selection and identification of clones with high reactivity for a specific target.

| **Technology (typical number of sequences screened per library)** | **Description** | **Strengths or weaknesses** |
|---|---|---|
| Bacterial display ($10^{8}$–$10^{9}$) | Proteins are displayed on the surface or cell envelope of *Escherichia coli* | Selects proteins that can be made in cells |
| | | Flow cytometry allows multiparameter, quantitative screening |
| mRNA display ($10^{15}$) | mRNA-protein fusions are synthesized through a puromycin linker; reverse-transcription PCR allows amplification after rounds of selections | Large libraries |
| | | Can screen proteins that would be toxic to cells |
| | | Works best with small proteins |
| | | Stringent conditions required |
| Phage display ($10^{11}$) | DNA libraries encoding displayed proteins and required phage genes are put into bacteria, which produce the library attached to the phage surface | Robust and quick |
| | | Smaller libraries than cell-free systems |
| Ribosome display ($10^{15}$) | DNA libraries encode the displayed proteins as a fusion to a sequence that tethers both mRNA and protein on stalled ribosomes; reverse-transcription PCR allows amplification after rounds of selections | Large libraries |
| | | Can screen proteins that would be toxic to cells |
| | | Requires stringent conditions and stable proteins |
| Yeast display ($10^{8}$–$10^{10}$) | Gene libraries code for the target protein fused to a yeast surface protein | Flow cytometry allows multiparameter, quantitative screening |
| | | Selects proteins that can be made in eukaryotic cells |

**Fig. 1.3** Display technologies link protein to gene by putting proteins on the outer surface of viral particles or cells, or by physically linking mRNA to protein in cell-free systems (Baker 2011).

Display methods can be classified into two groups: (I) cell display methods, including bacterial and yeast display, and (II) in-vitro display technologies including ribosome and mRNA display. Among the display technologies, there is also phage display.

## 1.5.1 Bacterial surface display

In bacterial display naturally occurring surface proteins are used as carriers for foreign molecules to be displayed on the surfaces of bacteria.

Bacterial surface display entails the presentation of recombinant proteins or peptides on the surface of bacterial cells. *Escherichia coli* is the most frequently used bacterial host for surface display and a variety of *E. coli* display systems have been described that primarily promote the surface exposure of peptides and small proteins (van Bloois et al. 2010). Gram-positive bacteria seem to be more suitable for applicative purpose like whole-cell catalysts and whole-cell adsorbents because of the rigid structure of their cell walls; *Bacillus* and *Staphylococcus* strains have been used most often.

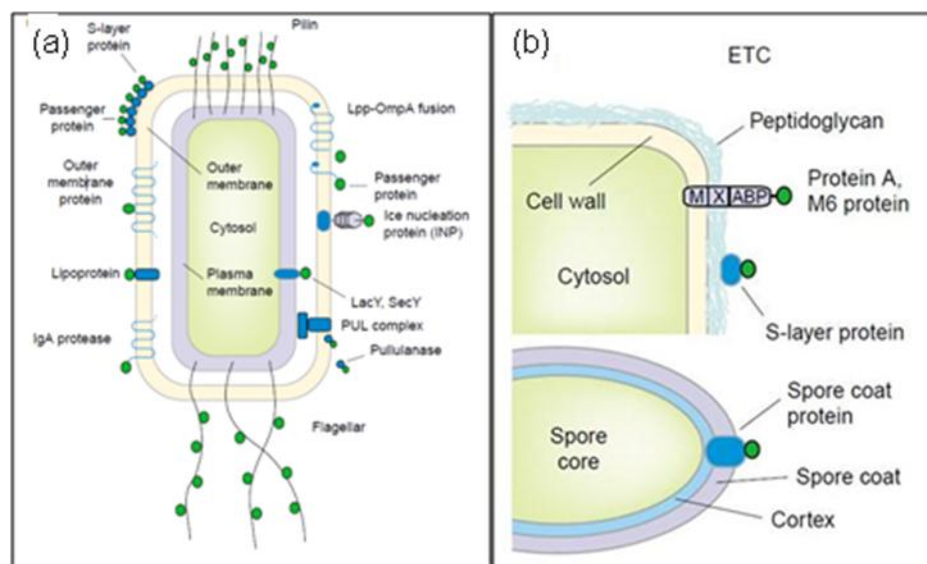



**Fig. 1.4** Cell surface display system in Gram-negative (a) and Gram-positive (b) bacteria; green circles represent heterologous passenger proteins.

*E. coli* was chosen thanks to rapid growth rate, ease of genetic and physical manipulation and its suitability for making large libraries. A wide variety of different scaffolds, or carrier proteins, have been utilized to present peptides

and proteins on the outer surface of *E. coli* (Lee et al. 2003). For “surface display” the scaffold must be capable of transporting the desired passenger protein to the external surface of *E. coli*. The passenger’s size, folding efficiency, and disulfide content can strongly influence its ability to be secreted across the outer membrane and become localized on the cell surface. Unfortunately, differences in the host strain and expression conditions, surface localization methods, and the passengers themselves make the comparison of the passenger limitations for each scaffold problematic (Veiga et al. 2004).

To enable effective affinity-based screening against protein targets, the scaffolds should be monomeric to reduce the likelihood that multiple receptor–ligand interactions, or avidity effects, obscure the true affinity of a 1:1 stoichiometric complex. Scaffolds should be randomly distributed and spatially separated on the outer membrane of bacteria, to avoid local clusters of receptors that mediate avidity effects, i.e. supradditive effects observed upon dimerization or multimerization of monomers. Additionally, despite the cross linked architecture of the *E. coli* outer membrane, some outer membrane proteins such as LamB may be capable of lateral diffusion within the membrane (Gibbs et al. 2004), leading to avidity effects for multivalent targets.

The surface display of passenger proteins on *E. coli* can be achieved by genetic fusion with various ‘scaffold’ proteins targeted to the outer membrane as well as those assembled into flagella and fimbrial structures. Bacterial display scaffolds can be divided into three groups: which allow N-terminal, C-terminal, and insertional fusions.

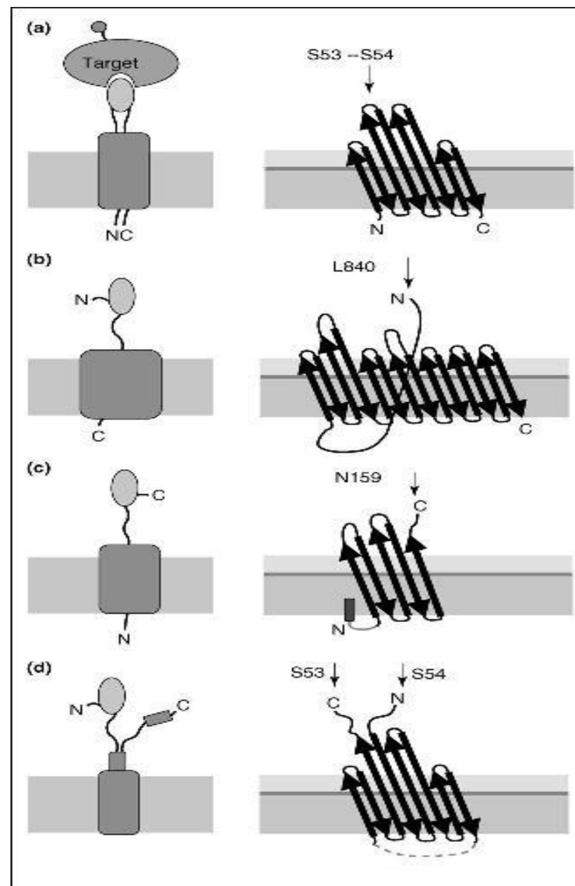



**Fig. 1.5 Representative** bacterial display scaffolds and their topologies. (a) Insertion scaffolds (e.g. OmpX), (b) N-terminal display scaffolds (e.g. AIDA-I autotransporter), (c) C-terminal display scaffolds (e.g. LppOmpA), and (d) combination of N-terminal and C-terminal display using circularly permuted OmpX (CPX). Arrows indicate permissive insertion or fusion site for display (Daugherty 2007).

Among these approaches, display of peptides on bacterial flagella using the 'FliTrx' system has been used most often because its commercial availability (Lu et al. 2003). With FliTrx, peptides are presented as constrained insertions within the active site loop of *E. coli* thioredoxin, which is in turn inserted into a surface-exposed region of the abundant, repeating flagellar protein FliC (Lu et al. 1995). Insertion libraries such as those used by FliTrx system and those created in outer membrane proteins (e.g. OmpA, OmpC, OmpX, and FhuA) are well suited for mapping antibody and protein-binding epitopes, and selecting initial low affinity binders toward challenging targets for subsequent affinity maturation.

Display of passenger polypeptides as N-terminal fusions with a surface-exposed N-terminus of the display scaffold can be accomplished via fusion to

autotransporter proteins. Autotransporters used for library screening include the IgA protease from *Neisseria gonorrhoeae*, *E. coli* AIDA-I (Maurer et al. 1997), or EstA from *Pseudomonas aeruginosa* (Yang et al. 2004). Although autotransporters are thought to translocate unfolded passengers, other studies suggest that autotransporters can also translocate various folded passengers (Veiga et al. 2004).

Display via the scaffold's C-terminus may be beneficial to enhance the diversity of peptide libraries since stop codons arising from common randomization schemes and nonintended errors (primer deletions or PCR errors) can yield functional binders without truncating the carrier protein. C-terminal display libraries have been generated and screened using intimins (EaeA), invasins, and the LppOmpA vector. For efficient C-terminal display of some proteins via EaeA is required the maintenance of the passenger in an unfolded conformation for export (Adams et al. 2005). The ice nucleation protein (INP) scaffold (Jung et al. 1998) might also enable screening of C-terminal display libraries for binders, for example enzyme libraries.

A scaffold presenting both N-terminal and C-terminal on the cell surface was engineered by circular permutation of the smallest member of the outer membrane protein family, OmpX (Rice et al. 2006). The circularly permuted OmpX (CPX) scaffold enables normalization of protein display levels by fluorescence labeling of a C-terminal affinity tag. Alternatively, the adjacent termini could be used to present heterodimeric proteins.

Affinity-based screening of cell surface display libraries, like bacterial and yeast display library, generally requires use of FACS, since use of magnetic selection (MACS) alone or panning processes such as that used with the FLiTrx system lead to avidity interactions that interfere with affinity screening.

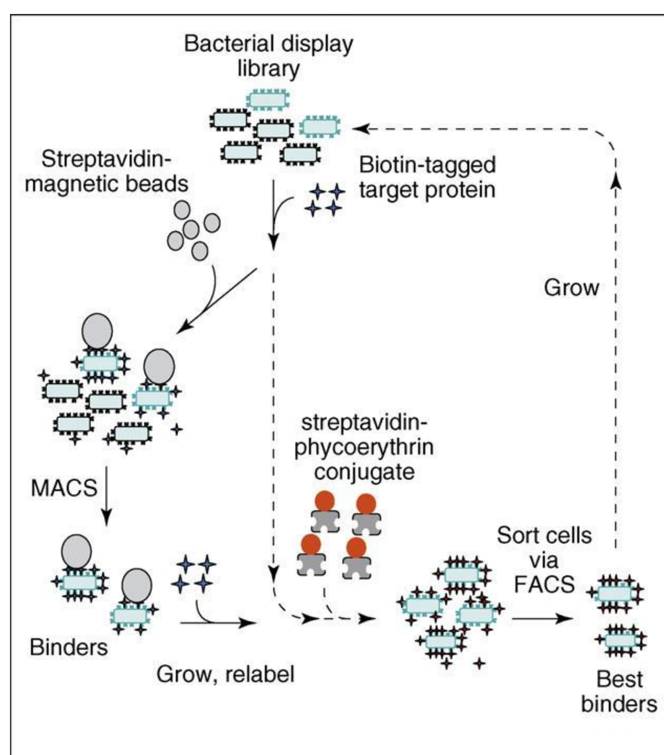



**Fig. 1.6** Bacterial display library screening. Typical combined selection and screening strategy for large libraries using biotinylated target proteins for sequential magnetic separation (MACS) with streptavidin-functionalized magnetic particles followed by fluorescence-activated cell sorter (FACS) of the enriched population for fine affinity resolution (Daugherty 2007).

Determination of an antibody's binding specificity using peptide libraries, i.e. epitope mapping, has been demonstrated using several bacterial display scaffolds using both selections and screening via FACS. To demonstrate the multispecificity capability of antibodies, FliTrx has been used to identify peptide mimitopes for an antihapten IgE (James et al. 2003). Linear peptides derived from screening possessed barely detectable affinity, while peptides constrained within thioredoxin possessed dissociation constants of roughly 10 mM. Sequential MACS and FACS has been used to screen a large library of $5x10^{10}$ random 15-mer insertions into OmpA against the anti-T7 tag antibody (Bessette et al. 2004) yielding a six-residue consensus sequence. Using a similar library constructed as insertions within OmpX, the epitopes of two monoclonal antibodies were mapped by performing two cycles of library enrichment in a dime-sized microfluidic device (Bessette et al. 2007).

Bacterial display was also used to identify the dominant specificities of the circulating antibody repertoire. Such serum 'antibody fingerprinting' studies

can provide insights into mechanisms of pathogenesis, as well as provide reagents that could potentially be used to improve diagnostic tests.

## 1.5.2 Yeast display

The yeast *Saccharomyces cerevisiae* is very useful as a host cell in genetic engineering because it is generally recognized as safe and this feature allows its use in food and pharmaceutical applications (Schreuder et al. 1993; Murai et al. 1997; Ueda and Tanaka 2000; Matsumoto et al. 2002). Yeast is able to glycosylate heterologous eukaryotic proteins and has the advantage of high density cultivation in inexpensive medium. In addition, the yeast has a potential to display not self eukaryotic proteins, and can display different kinds of protein on the same cell surface, named "co-display". Moreover, a flow cytometer is applicable for yeast cells in the case of high-throughput screenings.

The yeast *S.cerevisiae* cell is surrounded by a hard cell wall (Fig. 1.7), about 200 nm thick that consists of *β*-linked glucans and mannoproteins, and exists outside of the plasma membrane. The cell wall consists of an internal skeletal layer of glucan, composed of *β*-1,3- and *β*-1,6-linked glucose and a fibrillar or brush-like outer layer composed predominantly of mannoproteins. There are two types of mannoprotein in the thick cell wall. One is loosely bound to the cell wall with non-covalent bonds, and is extractable with sodium dodecylsulfate (SDS). The other type is extractable by digestion of the cell wall with *β*-1,3- or *β*-1,6 glucanase.

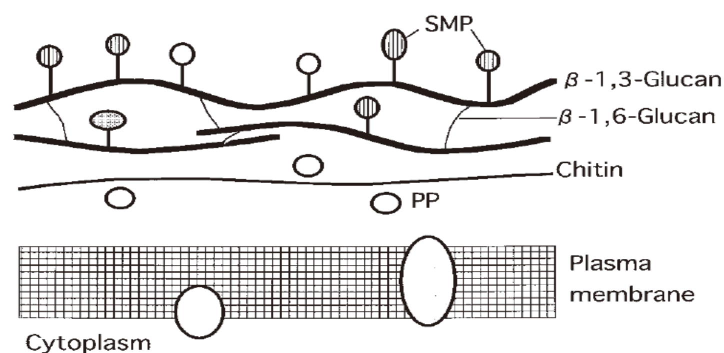



**Fig. 1.7** Architecture of the yeast cell wall. SMP, glucanase-extractable surface-layer mannoprotein; PP, SDS-extractable periplasmic protein (Shibasaki et al. 2009).

A number of heterologous proteins varying in size between 0.93 and 136 kDa have been successfully displayed on the yeast cell surface. In many cases, $10^4$–$10^5$ molecules were displayed on each cell (Nakamura et al. 2001; Shibasaki et al. 2001).

| Protein and peptides | Promoter | Secretion signal sequence | Moleculer mass (kDa) |
|---|---|---|---|
| α-Agglutinin system | | | |
| *Rhizopus oryzae* glucoamylase | GAPDH | Glucoamylase | 62 |
| *Bacillus stearothermophilus* α-amylase | GAPDH | MFα1 | 59 |
| *Aspergillus aculeatus* β-glucosidase | GAPDH | Glucoamylase | 136 |
| *Trichoderma reesei* endoglucanase | GAPDH | Glucoamylase | 42 |
| *Rhizopus oryzae* lipase | GAPDH | MFα1 | 30 |
| *Aequorea victoria* GFP | GAPDH | Glucoamylase | 27 |
| BFP | *UPR-ICL* | Glucoamylase | 27 |
| ECFP | *PHO5* | Glucoamylase | 27 |
| EYFP | *MEP2* | Glucoamylase | 27 |
| Apoaequorin | GAPDH | Glucoamylase | 21 |
| Hexa-His | GAPDH | Glucoamylase | 0.93 |
| ZZ | *UPR-ICL* | Glucoamylase | 14 |
| Fab fragment of antibody | GAPDH | Glucoamylase | 50 |
| a-Agglutinin system | | | |
| Single-chain antibody | Gal1 | Aga2 | 27–30 |
| Single-chain T-cell receptor | Gal1 | Aga2 | 26–28 |
| C-terminus region of Flo1p | | | |
| *Rhizopus oryzae* glucoamylase | GAPDH | MFα1 | 62 |
| EGFP | *UPR-ICL* | MFα1 | 27 |
| Flocculation functional domain of Flo1p | | | |
| *Rhizopus oryzae* lipase | *UPR-ICL* | Flo1p | 30 |
| EGFP | *UPR-ICL* | Flo1p | 27 |

**Fig. 1.8** Example of yeast cell-surface display of heterologous proteins (Kondo and Ueda 2004).

Many glucanase-extractable proteins on the yeast cell surface, for example, agglutinin (Agα1 and Aga1) and flocculin Flo1, Sed1, Cwp1, Cwp2, Tip1, and Tir1/Srp1, have glycosylphosphatidylinositol (GPI) anchors which play important roles in the surface expression of cell surface proteins and are essential for the viability of the yeast (Hardwick et al. 1992; Watari et al. 1994; van der Vaart et al. 1995).

There are two types of cell-surface system: (I) the GPI system (figure 1.9C) which contains a GPI-attachment signal (in the C-terminal region of Flo1p), and the C-terminus of the target protein fused to the anchor and (II) a second

system (figure 1.9D) where the N-terminus of the target protein is fused to the Flo1p flocculation functional domain (Matsumoto et al. 2002).

GPI anchors have a structure that is very well preserved across a range of different organisms. The core structure of the yeast GPI anchor is similar to that found in other eukaryotes, although the lipid composition varies from yeast to yeast. The glycophospholipid moieties are covalently attached to the C-termini of proteins and their primary function allow the stable association of proteins with the membrane. GPI-anchored proteins contain hydrophobic peptides at their C-termini. The localization of GPI-anchored proteins on the cell surface occurs through the secretory pathway of *S.cerevisiae*. Because of the covalently linked lipid anchor, the protein remains membrane-bound, exposed initially on the luminal side of the ER; the protein is then transported from the ER to the Golgi apparatus and from there to the plasma membrane in membrane-enclosed vesicles. Fusion of the Golgi-derived secretory vesicles with the plasma membrane releases the secreted proteins to the cell exterior. Post-translational proteolytic modification of the precursors of secretory peptides occurs late in the secretory pathway (Schekman 1992). GPI-anchored proteins are further transported to the outside of the plasma membrane through the general secretory pathway in a GPI-anchored form, then released from the plasma membrane by a phosphatidylinositol-specific phospholipase C (PI-PLC) and transferred to the outermost surface of the cell wall, where anchorage is accomplished by the addition of β1,6-glucan to the GPI anchor remnant in a manner dependent on the prior addition of a GPI anchor (Lu et al. 1995).

Agglutinin and flocculin are the typical GPI anchor protein used in yeast cell-surface display systems.

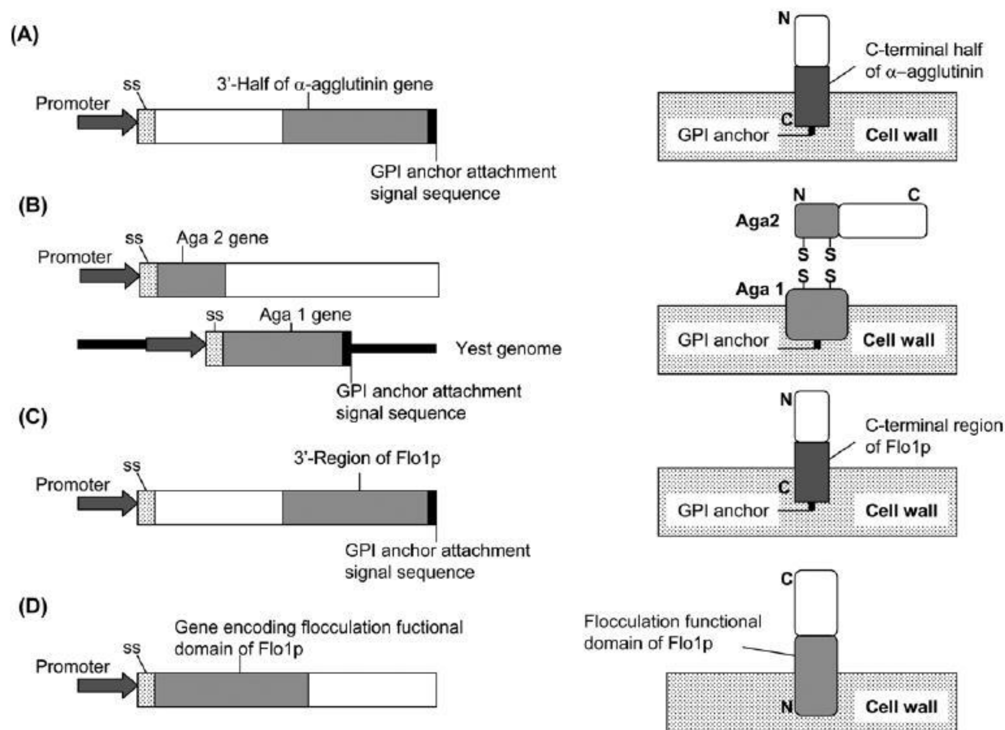



**Fig. 1.9** Yeast cell surface display system using **A** α-agglutinin, **B** a-agglutinin, **C** C-terminus region of Flo1p, and **D** N-terminus flocculation function domain of Flo1p (Kondo and Ueda 2004).

Among the glucanase-extractable mannoproteins on the cell surface of *S. cerevisiae*, the mating-type-specific agglutinins, which mediate the direct cell-cell adhesion between cells of the opposite mating type during mating and represent minor cell-wall components, are assumed to be located on the outermost surface (Lipke and Kurjan 1992). Mating type a and α cells express a-agglutinin and α-agglutinin, respectively.

α-Agglutinin has a predicted length of 650 amino acids before processing. As shown in Fig.8A fusion to the C-terminal half of α-agglutinin (320 amino acid residues), which contains a GPI-anchor attachment signal at the C-terminal end, is used to anchor heterologous proteins on the yeast surface, since these proteins are covalently linked with glucan.

Considering a-agglutinin, the secretion-type protein Aga2p, the binding subunit linked by S-S to the core protein Aga1p, is used (figure 1.9B). The Aga2p fusion protein and Aga1p associate within the secretory pathway, are exported to the cell surface and covalently linked to the cell wall.

Flocculin Flo1p is a lectin-like cell-wall protein of *S. cerevisiae*, composed of several domains: secretion signal, flocculation functional domain (near the N-terminus, recognizes and adheres non-covalently to cell-wall components causing reversible aggregation of cells into flocs), GPI-anchor attachment signal, and membrane-anchoring domain.

Yeast can be used to display binding proteins as adsorbents in environmental processes for example to recover heavy-metal ions: an histidine oligopeptide (Hexa-His) with the ability to chelate divalent heavy metal ions ($Cu^2$+, $Ni^2$+, etc.) has been displayed on the yeast cell surface to enhance adsorption (Kuroda et al. 2002). Another application is the purification of antibodies, in fact using the C-terminal half of α-agglutinin can be obtained a yeast strain displaying the ZZ domain derived from *Staphylococcus aureus* (SPA), which binds the Fc part of immunoglobulin G from various species, including human and rabbit. This yeast cells displaying ZZ have been successfully used for purification of IgG from serum (Nakamura et al. 2001; Shimojyo et al. 2003). Yeast cell-surface display systems can be used for the display of single-chain antibody (scFv) and for the development of antibodies with enhanced affinity and stability to improve the use of autoantibodies as therapeutics. Large combinatorial scFv libraries can be screened and, together with fluorescence-activated cell sorting (FACS), allow rapid quantitative isolation of rare clones with the desired characteristics (Feldhaus et al. 2003).

### 1.5.3 Ribosome and mRNA display

In this two techniques, DNA molecules are first transcribed in vitro into mRNA, and then translated into proteins by a stoichiometric number of ribosomes, so the library size is determined by the number of different full-length protein molecules coupled to their coding mRNA. The most frequently used types of libraries have been based on peptides and antibodies. Peptides are advantaged by small size and simple structure, which made them easy to randomize synthetically. Antibodies have the advantage of a high-diversity natural repertoire that can be used in artificial selection systems.

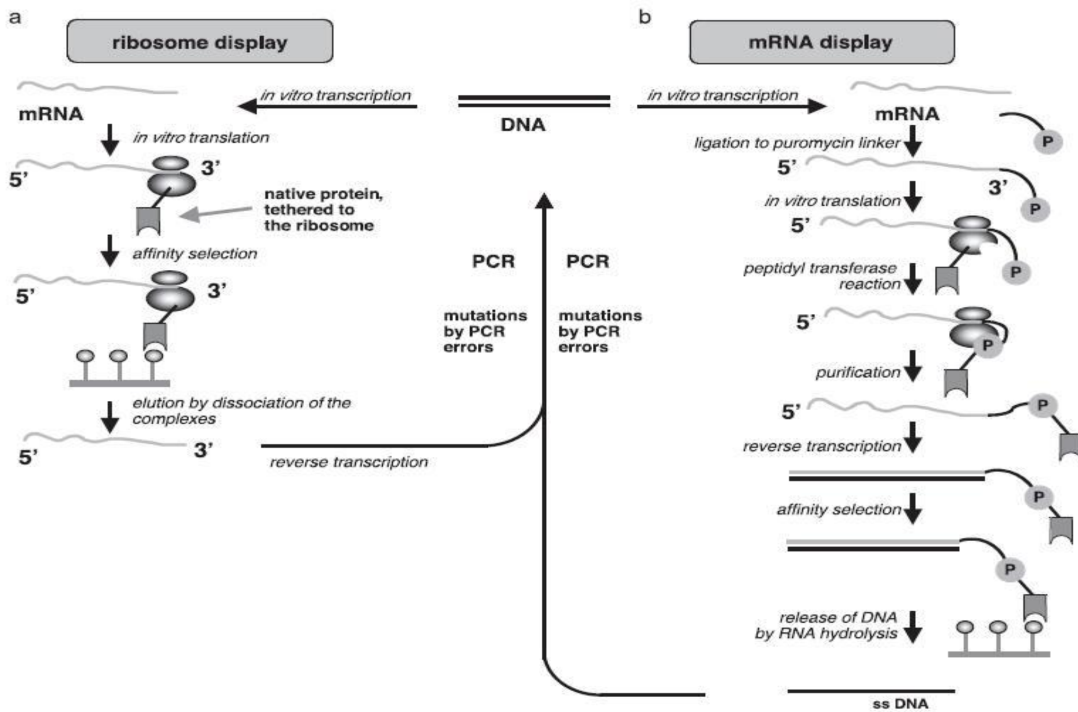



**Fig. 1.10** Comparison of two different strategies of in-vitro display (a) ribosome display and (b) mRNA display (Lipovsek and Pluckthun 2004).

In **ribosome display**, the translated protein remains connected to the ribosome and to its encoding mRNA; the final product is a protein-ribosome-mRNA complex (PRM) that can be used in a selection process to bind an immobilized ligand. If mRNA molecule has no stop codon, the ribosome will run to the end of the mRNA molecule. The corresponding polypeptide emerges from the ribosome while its end is still in the ribosomial tunnel, and its last amino acid is still connected to the peptidyl-tRNA. In nature the release of polypeptide from ribosome is due to release factors (RF) that bind the STOP codon. Ribosome that translate mRNA without stop codon will be trapped in a form composed by the mRNA (the genotype) and the protein, fold in its correct structure (the phenotype).

In figure 1.10a the typical cycles of ribosome display are shown. A large DNA library, coding for binding proteins genetically fused to a spacer sequence without stop codon, is transcribed in vitro into mRNA. The resulting modified mRNA is used as template for in vitro translation. This spacer sequence, when translated, is still attached to the peptidyl-tRNA. The complex PRM can be stabilized and used for selection against immobilized target (e.g. in

biotinylated form) and the ribosomial complexes are captured (e.g. by streptavidin-coated magnetic beads), washed to remove library component that bind non-specifically or weakly. The enriched library can be recovered using EDTA to destabilize the ribosomal complexes, or using a competitive elution with the target molecule. The DNA template for the next round of selection is provided with reverse transcription reaction followed by PCR (RT/PCR). Some errors, at this stage, can be useful to increase the diversity centered around enriched sequences (Zahnd et al. 2004).

Ribosome display has different applications. It was used to identify the main antigenic polypeptides of *Staphylococcus aureus* (Weichhart et al. 2003). A cDNA library, enriched in the correct reading frame, was selected against anti-staphylococcal antisera and a number of ORFs has been identified. This encodes those polypeptides that give rise to a major fraction of the human antibody response and could result in interesting vaccine candidates. For this kind of application, the high library size obtainable with ribosome display was an important advantage. As a first validation of the ribosome display technology, a model system of two antibody scFv fragments was used. A $10^9$-fold enrichment was achieved by five cycles of ribosome display (Lipovsek and Pluckthun 2004). All selected scFvs acquired genetic mutation during the cycles. Starting from this immune library, it has been demonstrated that scFv antibody fragments could be selected and evolved using the ribosome display system. In fact in a study of Knappik et al. they observed that using as starting material a very large synthetic antibody scFv library HUCAL-1 of $2x10^9$ independent members, after six rounds of ribosome display selection, some mutation were introduced (Hanes et al. 2000; Knappik et al. 2000). This procedure mimics the process of somatic hypermutation of antibodies during secondary immunization. The resulting products of selection were different families of closely related sequences, which stem from a common progenitor that evolved during ribosome display. These mutations can improve the affinity to the antigen significantly.

In antibodies the binding surface is limited by the size of the scaffold. Natural repeated proteins, such as ankyrin or leucine-rich repeat protein have been

used to generate a combinatorial library of target-binding polypeptides, based on the modularity of this class of proteins (Binz et al. 2003; Forrer et al. 2003; Kohl et al. 2003). Through the consensus design of self-compatible modules that display variable surface residues, followed by their random assembly into repeat domains a very large interaction surface is created. If we start from a diversity of $7x10^7$ by randomly assembling the modules the theoretical diversity is potentiated to $(7x10^7)^n$ where n is the number of repeats. Since the designed libraries contain no cysteines, these molecules may be suitable for intracellular or proteomics applications. These protein libraries are highly valuable sources for novel binding molecules, which may be suitable for a whole range biotechnological, biomedical, and potentially therapeutic applications.

In **mRNA display** the linkage between mRNA and the protein is achieved by binding the two macromolecules through a small adaptor molecule (figure 1.11), typically puromycin. Puromycin is an antibiotic derived from *Streptomyces alboniger* bacterium; it is an analogue of the 3' end of a tyrosyl-tRNA one part of its structure mimics a molecule of adenosine, the other part mimics a molecule of tyrosine. Puromycin is a nucleotide-amino acid chimera and ultimately forms the nucleic acid-protein junction in the mRNA displayed protein.

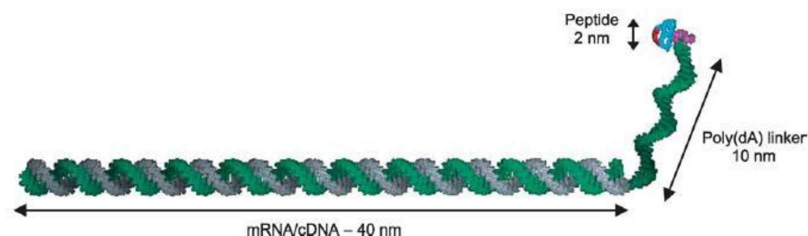



**Fig. 1.11** Model of mRNA-peptide fusion, the unit of selection in mRNA display. This figure is drawn to scale *Ecballium elaterium* trypsin inhibitor, EETI-II, a disulfide-constrained, 28-residue peptide and for its encoding nucleic acid. Both a molecule of mRNA (gray) and the encoded peptide or protein (teal and red) form covalent bonds to an adaptor molecule, e.g., puromycin (purple). A strand of complementary DNA (green) is reverse-transcribed from the mRNA and hybridizes to its template (Lipovsek and Pluckthun 2004).

In figure 1.10b the typical mRNA-display selection cycle is shown. The synthesis of an mRNA display library (free of stop codons) starts from the synthesis of a DNA library. Usually, each member of this DNA library has a T7

RNA polymerase transcription site and a ribosomal binding site at the 5' end. The T7 promoter region allows large-scale in vitro T7 transcription of the DNA library into an mRNA library. All mRNA templates used for mRNA display technology have puromycin at their 3' end. The resulting mRNA is used as template for in vitro translation. After translation, the ribosome proceeds to form a peptide bind between the adaptor molecule and the C-terminal amino acids residue in the nascent polypeptide chain (figure 1.11). The resulting mRNA-protein fusion is purified away from ribosomes and other components (for most applications, a DNA chain complementary to the protein-bonded mRNA is introduced to stabilize the nucleic acid component and to facilitate the recovery of genetic information after the selection). The mRNA/cDNA-protein library is selected using a specific target and the complexes are captured using affinity chromatography or immunoprecipitation of the target. To recover the enriched library, the complexes are hydrolyzed and cDNA strand is released. The recovered cDNA, amplified by PCR, provides the DNA template for another round of selection. Error-prone PCR can be used to increase the diversity centered around enriched sequences (Xu et al. 2002). To select protein with high affinity for the target, 4-10 rounds of selection are required.

mRNA display has been used to select high affinity reagents from engineered libraries of linear peptides, constrained peptides (Baggio et al. 2002), single-domain antibody mimics (Xu et al. 2002), variable heavy domains of antibodies and single-chain antibodies. In a study of Baggio et al. it has been demonstrated that mRNA display can be used routinely to identify linear epitopes of existing antibodies, in fact a peptide library with 27 randomized positions yielded a family of sequences that bound the anti-c-Myc antibody. The selected clones contained sequences that were either homologous or identical to a 10-residue stretch of the 32-residue c-Myc antigen (Baggio et al. 2002).

## 1.5.4 Phage display

Since its introduction by G. Smith in 1985 (Smith 1985), phage display technology has demonstrated to be effective for producing large libraries of polypeptides and efficiently isolating molecules with a given function (Azzazy and Highsmith 2002). Also, it has been employed for selecting antigens (Beghetto et al. 2003; Beghetto et al. 2006; Beghetto et al. 2009), characterizing epitopes (Deroo and Muller 2001; De Paolis et al. 2007) and antibodies (Winter et al. 1994), for investigating protein-protein interaction (Cesareni 1992; Tong et al. 2002) and for enhancing affinity in protein-ligand interaction (Li et al. 2003).

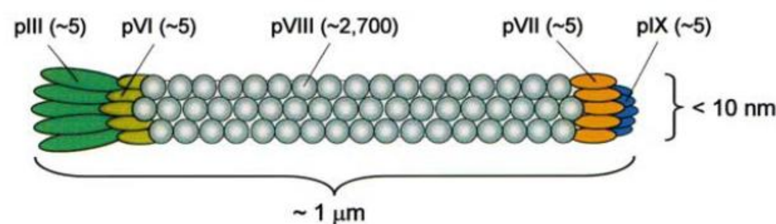



**Fig. 1.12** Schematic representation of M13 bacteriophage (Willats 2002).

Phage display system can be classified into two categories: non-lytic phage display and lytic phage display.

Lytic phage display includes lambda phage and T7 phage (Danner and Belasco 2001; Zhang et al. 2005). In this system, foreign cDNA library is directly inserted into lambda or T7 phage genome and expressed as capsid fusion proteins. A peculiar characteristic of lytic phage display is that it is not necessary for the proteins display on the surface of lambda or T7 phage to be secreted through the host bacterial membrane (Kruger and Schroeder 1981). On the contrary, this is a crucial step in filamentous phage assembly. Lambda is a temperate bacteriophage of *E. coli*, characterized by a double-stranded DNA genome of 48,502 bp in length. Inside the bacteriophage head, the viral genome is packaged as a unique double-stranded linear molecule with two single-stranded protruding terminals of 12 nucleotides (cohesive ends). When lambda infects the bacterium, the linear DNA molecule is injected into the

host, rapidly forming a circular molecule that serve as a transcription template during the early uncommitted phase of infection.

The lambda genome is organized into three functionally related gene clusters: (I) the left-hand region, including genes responsible for packaging and assembling the DNA genome into bacteriophage head; (II) the central region, including genes involved in establishment and maintenance of lysogeny, and genes not essential for lytic growth (useful for cloning of DNA inserts); (III) the right-hand region, containing genes which are essential for DNA replication and lysis of infected bacteria. During the lysogenic state, the lambda genome is stably integrated into the bacterial chromosome and is replicated as a part of the host genome and transmitted to the bacterial progeny. During the lytic cycle, the circular DNA directs the synthesis of proteins required for viral replication, assembly of bacteriophage particles and cell lysis. The lambda genome is replicated bi-directionally by a "rolling circle" mechanism, producing a linear concatemeric substrate for DNA packaging. The viral particle is constituted by an icosahedral head of 415 and 405-420 copies of the major capsid proteins gpE and gpD, respectively, and by a flexible helical tail, consisting of 32 disks each containing six subunits of the major tail protein gpV (Dokland and Murialdo 1993).

Lambda bacteriophage has been demonstrated as being a useful system to display complex cDNA libraries, and both gpV and gpD proteins have been used as fusion partners. The pioneering vector for displaying foreign proteins onto bacteriophage lambda employed the gpV tail protein s fusion partner. In 1994, Maruyama and co-workers engineered a lambda vector (λfoo) which allows the expression of foreign polypeptides as fusion to the C-terminus of a truncated gpV protein by replacing the last 70 amino acids of the tail protein (Maruyama et al. 1994). These fusion constructs have been used efficiently for library panning but showed same limitations: (I) low display level (i.e., few fusion products per phage particle) and (II) low yield of phage recovery after affinity purification. To increase the display level, protein gpD has been explored as fusion partner of foreign proteins. It can tolerate both amino- and carboxyl- terminal insertions of peptides and is accessible for ligand

interactions without interfering in phage replication and assembly. It has been reported that foreign polypeptides can be displayed on gpD at very high density copies and may interfere with phage morphogenesis. This is commonly displayed using a two-gene system, where both wild-type gpD and recombinant D-fusion genes are co-packaged into lambda head.

The first example towards the use of lambda display for antigen discovery was the screening of a human hepatitis C virus (HCV) cDNA library with monoclonal antibodies or sera from HCV-infected individuals, where foreign cDNA inserts were cloned into lambda genome by tagged random-primed elongation (Santini et al. 1998). The results demonstrate that several different protein domains displayed on lambda gpD could be incorporated into viable particles and they were accessible for interaction and affinity-selection with specific antibodies. Most of the selection schemes for screening lambda libraries utilize a protocol which had been originally developed for filamentous phage applications.

Non–lytic phage display systems use vectors derived from filamentous phages (M13, f1 and fd) (Paschke 2006). They infect Gram-negative *E. coli* bacteria using pili as receptor (male *E. coli*). They consist of a single-stranded (ss) DNA that is enclosed in a protein coat. The entire genome of the phage consists of 11 genes grouped on the DNA according to their function: DNA replication, capsid packaging and encoding proteins involved in phage membrane assembly. A viable phage consists of about 2700 copies of gene 8 protein (g8p or pVIII, a 50 aa residue protein that is also known as the major capsid protein), 3 to 5 copies of the gene III (g3)-encoded adsorption protein (g3p or pIII, a 406 aa protein that is one of the three minor coat proteins of the filamentous phage) and other 5 copies of the gene VI (g6 or pVI) on its tip. On the opposite tip, it expresses 5 copies of the gene VII (gVII or pVII) and 5 copies of the gene IX (gIX or pIX) (Azzazy and Highsmith 2002; Willats 2002; Beghetto 2011).

The most popular phage display strategy is to fuse the DNA encoding for a library polypeptides to the N-terminus of phage gene III capsid protein (pIII). In literature is described protein display also on other capsid protein, such as

pVI, pVII, pVIII and pIX (Kehoe and Kay 2005). The result is the creation of a chimeric coat protein, with the gene encoding for this protein packaged within the phage.

The phage libraries consist of a pool of soluble molecules exposed on the phage surface; from this pool can be isolated specific phages on the basis of the binding proprieties of the displayed polypeptide towards a protein of interest. The screening of phage libraries is achieved by consecutive cycles of selection and amplification (biopanning) where specific phages can be selected from a background of billions of other phage displaying polypeptides which are unable to bind the specific target and are washed away. Only specific phages are eluted and amplified through successive rounds of selection (Azzazy and Highsmith 2002).

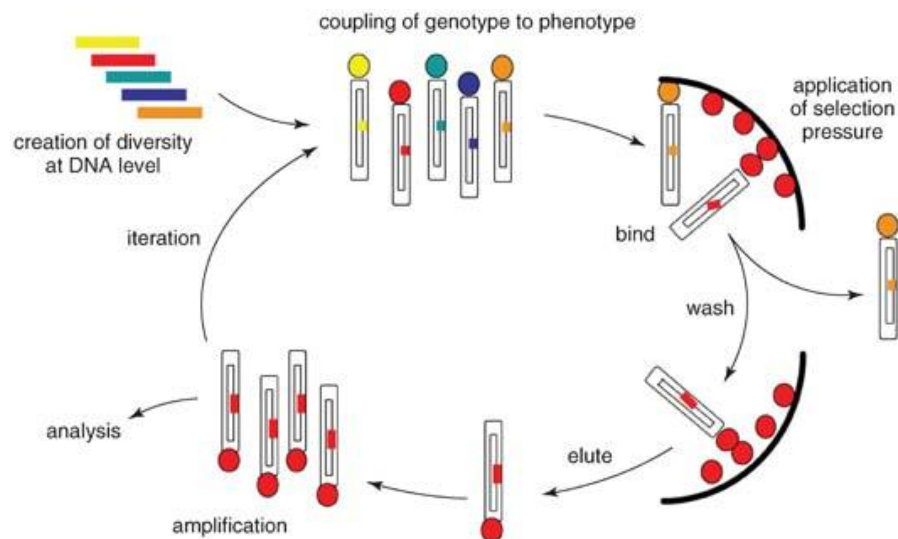



**Fig. 1.13** The phage display selection cycle (biopanning). It starts with the creation of diversity at the DNA level; this is translated into phenotypic diversity by the display of different polypeptides on the surface of the phage. The application of selection pressure, manifested by cycles of binding to a ligand, washing, and elution, allows the selection of phage displaying polypeptides which have high binding affinity to the ligand of interest (Bradbury 2010).

Most filamentous phage display systems use phagemids, which are hybrid of phage and plasmid vector, expressing only capsid fusion protein with a packaging signal and required a “helper phage” to provide wild-type pIII and other phage proteins to “rescue” the assembly of phagemids as phage particles with the display exogenous protein. The resulting phage particles may incorporate either pIII derived from the helper phage or the polypeptide-

pIII fusion protein, encoded by the phagemid. The ratios of polypeptide-pIII fusion protein: wild type pIII may range between 1:9 and 1:1000 depending on the type of phagemid, growth conditions, the nature of the polypeptide fused to pIII, and proteolytic cleavage of polypeptide-pIII fusions (Azzazy and Highsmith 2002; Willats 2002). Phagemids have specific characteristics: (I) gene III, (II) appropriate multiple cloning site upstream gene III, (III) antibiotic resistance gene, (IV) origin of replication of both the M13 phage and *E. coli* (Rhyner et al. 2002). Major advantages of phagemid vectors include smaller size and ease of cloning, compared with cloning in phage.

The filamentous phages appear to be suitable vectors for cloning for many reasons: (I) the ability of DNA replication, (II) assembly, and (III) especially because these characteristics are not influenced by the size of the DNA (Rhyner et al. 2002).

## 1.6 Phage display and epitope identification

The phage display technology has been used for many aims, but the most used application is the isolation of recombinant antibodies with a high specificity (Azzazy and Highsmith 2002). In this regard, it was used to generate human monoclonal antibodies or humanized mouse antibodies, isolate human antibodies from patients exposed to certain viral pathogens, elucidate the specificity of autoimmune antibodies.

To construct an antibodies library using phage display, genes encoding variable heavy ($V_H$) and variable light ($V_L$) chains of antibodies, which are responsible for their specificity, are prepared by reverse transcription of mRNA obtained from B-lymphocytes and after assembling they are cloned in a phagmidic vector. The obtained recombinant antibody is called single chain variable fragment (scFv). Antibody phage display libraries can be used to identify possible autoantigens. In this case the peptides exposed on the surface of phage correspond to the variable part of the antibody (scFv). This kind of library reflects the individual antibody profile. The library can be screened with specific antigens for which the reactivity is already known, to identify autoreactive antibodies.

Another application of the phage display technology is the creation of antigen library. This kind of library can be performed with two main approaches.

A first method involves the construction of phage libraries using random peptide ("random peptide phage libraries"); in this case the consensus sequences selected by binding proteins are compared with sequences in the database to identify the isolated protein. These libraries can be screened with sera from autoimmune patients, especially to identify potential immunogenic targets that evoke a specific antibody response. After a few cycles of selection, you can isolate and amplify only those phage displaying a peptide recognized by the serum immunoglobulin and, in the same time, their coding sequence (Azzazy and Highsmith 2002).

A second method involves the creation of gene/cDNA expression libraries, obtained by reverse transcription of mRNA extracted from the interest tissue. A first example of this type of expression libraries is the study of Systemic Lupus Erythematosus (SLE) in order to identify autoantigens involved in these autoimmune diseases. In this case the mRNA used to create the library was obtained from patients fibroblasts and it was possible to identify potential new -or already known- antigens involved in different autoimmune diseases (Kemp et al. 2002).

In direct comparison between the two methods, epitopes are more accurately identified using the gene fragment approach (Matthews et al. 2002).

Phage display with cDNA library is rare and not so efficient, in fact, among more than 4,000 literature citations related to phage display, only a few (about 5%) deal with cDNA libraries (Li and Caberoy 2010). One important and critical point is the possible reading frame shifts in the cDNA repertories fused to the N-terminus of filamentous phage pIII. In fact cDNA with unpredictable reading frames and stop codons interfere with pIII expression, resulting in only about 6% of identified clone encoding real proteins (Faix et al. 2004; Paschke 2006; Larman et al. 2011). Majority of identified non-open reading frames (non-ORFs) encoding non natural short peptides have minimal implications in

protein networks, for this reason an important point is the possibility of discriminating between ORF and non ORF sequences.

## 1.7 cDNA library: Open Reading Frame (ORF) selection

Gene fragment libraries require cloning system which allows the selection of ORFs fragments. In general, in expression libraries, where the multiple cloning site is upstream gIII/VIII (for example) to obtain a fusion protein, only one in 18 clones will be functional due to the difficulty of cloning in frame with g3p/8: one clone in two will be in the correct orientation, one in three will start in the correct frame, and one in three will finish in the correct frame.

| | |
|---|---|
| 1/2 correct orientation | ATG GTT CAG CCC CGT ATC CGT |
| | ACG GAT ACG GGG CTG AAC CAT |
| 1/3 start correctly | ATG GTT CAG CCC CGT ATC CGT |
| | TG GTT CAG CCC CGT ATC CGT |
| | G GTT CAG CCC CGT ATC CGT |
| 1/3 finish correctly | ATG GTT CAG CCC CGT ATC CGT |
| | ATG GTT CAG CCC CGT ATC CG |
| | ATG GTT CAG CCC CGT ATC C |
| 1/18 completely correct | ATG GTT CAG CCC CGT ATC CGT |

**Fig. 1.14** Only one in 18 clones will be functional, in fact, one clone in three will start correctly, one clone in three will end correctly, and one clone in two will have the correct orientation.

Several strategies have been developed to solve this problem, in fact, a rate of non-ORFs insert may be tolerable starting with DNA from a single gene or a small gene rich genome for which representation may be obtained with small libraries. Whereas using DNA more complex, for example, from a tissue, the possibility to select ORF insert is indispensable to obtain large representative cDNA libraries.

One strategy is to display polypeptides at the C-terminus of pIII, pVI and pVIII (Paschke 2006; Jestin 2008) because there is no need for DNA clones to terminate correctly. Crameri and Suter generated phagmid pJuFo, in which c-Jun leucin zipper domain was displayed in frame on the N-terminus of pIII. The cDNA library was fused to the C-terminus of c-Fos leucine zipper domain and secreted with a PelB signal sequence at the N-terminus of c-Fos. Both leucine zipper domains were flanked by cysteine residues. The Fos-library

fusion proteins were captured by displayed c-Jun domain with the formation of heterodimer and disulfide bonds (Crameri and Suter 1993). Only less than 10% of clones identified from a conventional C-terminal cDNA library of T7 phage display were ORFs, but, at least, avoided the problem of stop codons (Kalnina et al. 2008).

Another strategy is the creation of ORF cDNA libraries. The principle is simple: non-ORFs cDNA have a high frequency of stop codons (about 96% of 200-bp non-ORF cDNAs have at least one stop codon) (Garufi et al. 2005). In ORF phage display a C-terminal selection tag or marker is expressed only with ORF cDNA inserts. Tag- or marker-based selection eliminates non-ORFs and generates ORF libraries. The most common strategy are: (I) C-terminal phage capsid selection, (II) C-terminal Ampicillin selection and (III) C-terminal biotin tag.

- <u>C-terminal phage capsid selection</u>: if cDNA library proteins were fused to the N-terminus proteins of phage capsid, non-ORF phage clones with stop codons would not express the fusion capsid, thus they would be eliminated during phage panning. This strategy is not successful, in fact, only about 6% of identified clones encode ORFs (Faix et al. 2004). This data could be explained by speculating that the helper phage carrying a predominant wild-type pIII gene to supply other proteins for the rescue of the phagemid assembly allows pIII expression, as result, a high number of clones carrying not ORF fragments were detectable. To solve this problem, a new type of phage packaging system of hyperphage was developed. Approximately, 60% of cDNA library phages generated with the hyperphage had ORF inserts (Hust et al. 2006). However, phage panning with this ORF cDNA library was not reported, resulting in the efficiency of this system to identify ORF phage clones to remain unknown.

- <u>C-terminal Ampicillin selection</u>: the general idea is that DNA encoding an ORF permits transcription of an antibiotic resistance gene (downstream), whereas DNA containing a stop codon does not. This concept was originally described by Seehaus et al. in 1992 with a plasmid in which an antibody library was cloned upstream of a β-lactamase gene. Thus only genes in frame

with β-lactamase gene were able to survive in presence of Ampicillin. Zacchi et al. in 2003 demonstrated a similar strategy with a paghemid, wherein cDNA inserts were cloned upstream of β-lactamase gene (flanked by two homologous *lox* sites) fused to pIII. The β-lactamase gene was removed by Cre recombinase-mediated recombination after Ampicillin selection of the clones. The removal of the β-lactamase gene was necessary for the display of exogenous polypeptides at pIII N-terminus (Zacchi et al. 2003). In a study of Faix et al. in 2004, they analyzed sequences from Ampicillin resistant bacteria, they observed that the library had about 87% of ORF clones.

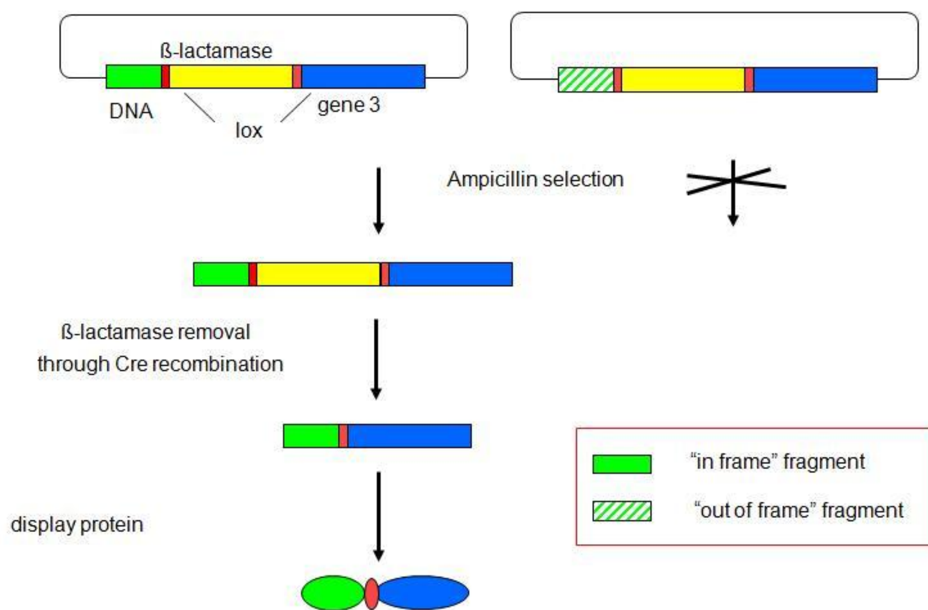



**Fig. 1.15** Selection of ORF fragments. Random fragments are cloned upstream of a β-lactamase gene. Those fragments that are ORFs permit read through into the β-lactamase gene and confer Ampicillin resistance. Those that are out of frame, or contain stop codons, do not survive. After selection on Ampicillin, the β-lactamase gene can be removed by passage through bacteria expressing Cre recombinase. The selected ORF can then be displayed on phage (Zacchi et al. 2003).

- <u>C-terminal biotin tag</u>: Ansuini et al. in 2002 generated ORF phage display cDNA library in lambda phage fused to the C-terminus of capsid protein, followed by the biotin tag. If a cDNA insert is an ORF, the C-terminal tag is expressed and biotinylated by BirA (biotin holoenzyme synthetase endogenously present in *E. coli*). In this way, only the ORF phage clones are labeled and enriched by binding to immobilized streptavidin. After selection, about 79% of clones randomly chosen were in correct reading frames.

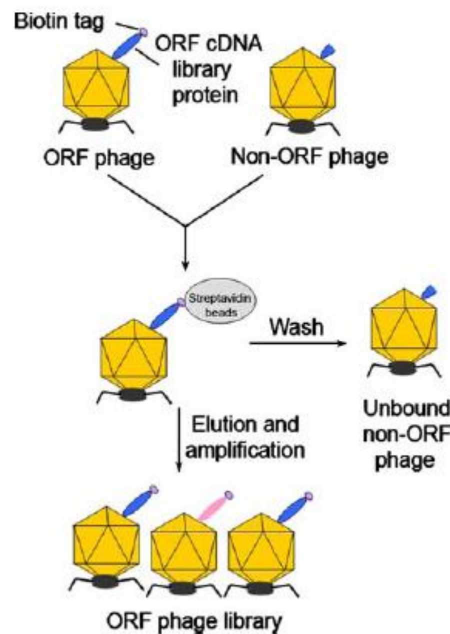



**Fig. 1.16** ORF phage display cDNA library with a C-terminal biotin tag. If the library protein is an ORF, biotin tag is expressed and biotinylated, otherwise, the biotin tag is not expressed. Biotinylated ORF phage can be enriched by binding to streptavidin to generate ORF cDNA library (Li and Caberoy 2010).

## 1.8 Expression cDNA libraries: normalization

One well recognized obstacle in the construction of expression cDNA libraries is the differential presence of the transcripts that reflects their level of expression in the target tissue. The differential abundance of various transcripts in any particular cell type was estimated as follows: 10-20 abundant genes (several thousand mRNA copies per cell) account for at least 20% of the cellular mRNA mass, several hundred genes of medium abundance ( several hundred mRNA copies per cell) constitute 40-60% of the total cellular mRNA, and finally several thousand rare genes (less than ten copies of mRNA per cell) represent about 20-40% of total mRNA mass (Carninci et al. 2000). This kind of distribution may markedly vary between different tissues, and the presence of numerous highly expressed genes may further unbalance this distribution. Random sequencing of clones from standard cDNA libraries is inadequate for discovering rare transcripts, owing to the repeated occurrence of intermediately and highly abundant cDNA that result very redundant. To decrease the abundance of clones representing abundant transcripts, a step of cDNA libraries normalization may significantly

increase the efficiency of random sequencing. Methods to decrease the prevalence of highly abundant transcripts and to equalize mRNA concentrations in a cDNA library are designated “cDNA normalization”. The use of normalized cDNA libraries greatly increases the efficiency of identification of rare transcripts (Zhulidov et al. 2005). Several cDNA normalization methods have been developed since 1990. A number of this methods have been optimized for the normalization of full-length enriched cDNA and used in various applications, including transcriptome analysis and functional screening of cDNA libraries (Bogdanova et al. 2008).

There are two main approaches for library normalization: (I) the reassociation of denatured double-stranded cDNA and (II) the hybridization of first strand cDNA with its complementary RNA. The normalization process is based on the different kinetics of hybridization of double strand nucleic acid: the most abundant cDNA pair up quicker and can be removed from the pool of cDNA. In this way the pool can be enriched with under-represented sequences. Ideally, each sequence in a normalized library should be represented in a comparable number of times.

These libraries present several advantages, especially if the main purpose is to discover rare genes or epitopes, but also disadvantages since it can produce artifacts, and because the indication of the level of gene expression was lost, whereas in some cases may represent an important information.

Briefly, the majority of the methods used for the cDNA normalization involves the following steps: (I) from synthesis or denaturation of double-stranded cDNA is obtained the single-strand cDNA; (II) the single-strand is submitted to hybridization in the presence of its complementary strand (cDNA or RNA) and normalization process. The normalized sequences are (III) separated from the not normalized double-strand sequences and (IV) cloned to obtain a normalized cDNA library.

The major differences between the normalization methods used consist essentially in the separation procedure of the fraction of normalized single-stranded cDNA sequences (Zhulidov et al. 2005). The technique proposed to

achieve this goal include separation of single-strand and double-strand fraction using hydroxylapatite columns (Ko 1990; Patanjali et al. 1991; Soares et al. 1994), or magnetic beads (Sasaki et al. 1994; Carninci et al. 2000), digestion of the double-strand fraction by restriction endonucleases (Coche and Dewez 1994) and amplification of the single strand fraction using suppression PCR (Luk'ianov et al. 1994; Luk'ianov et al. 1996; Diatchenko et al. 1999). Some disadvantages of normalization process of full-length cDNAs are the following:

- The suppression PCR is only applicable to short cDNA fragments, since the suppression is not effective for long molecules;
- Procedures based on restriction endonucleases to digest the double-strand fraction result in loss of transcript that form secondary structures (Patanjali et al. 1991);
- Solid matrix-based methods are generally not efficient enough because the kinetics of hybridization of nucleic acids immobilized on a solid phase are slower than those in solution (Zhulidov et al. 2004).

The first method proposed for full-length cDNA normalization is the “normalization and subtraction of Cap-Trapper selected cDNA” (Carninci et al. 2000). In this method, “Cap-trapper-prepared first-strand cDNA” is equalized during reassociation in the presence of biotinylated mRNA from the same source. Biotinylated RNA/cDNA hybrids are removed using magnetic beads. The multiple step of this method are shown in figure 1.17.

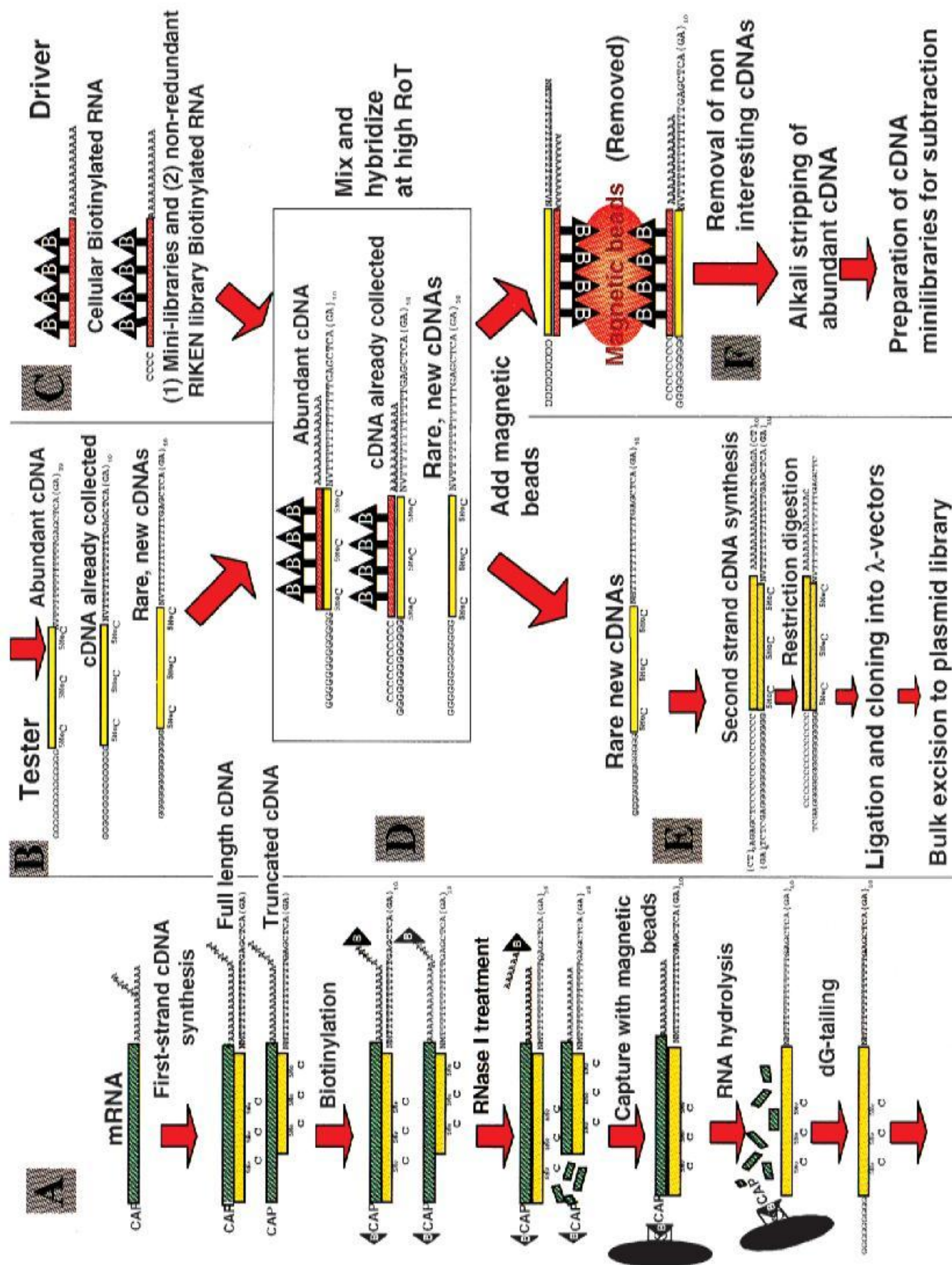



**Fig. 1.17** Schematic representation of the normalized-subtraction cDNA preparation protocol. It consists of the following main steps: (A) synthesis of first strand cDNA from mRNA poly (A) + using "T-primer"; Biotinylation of the cap structure of RNA; Degradation of RNA (with RNase I treatment); Separation of the DNA-RNA hybrids that contain single-strand cDNA complete sequence, using paramagnetic beads. (B) Tester cDNAs and (C) normalizing driver and subtracting drivers underwent a process of (D) hybridization. (E) Rare cDNAs are used for second-strand cDNA preparation while (F) abundant cDNAs are removed (Carninci et al. 2000).

This method includes several steps, involving the physical separation of the target cDNA fraction, and requires a large quantity of poly(A)+ RNA (Zhulidov et al. 2005).

Another method that allows the normalization of the cDNA, is based on the use of an enzyme called "Duplex Specific Nucleases" (DSN) (Shagin et al. 2002). This method does not include physical separation steps and is based on selective hydrolysis of the double strand-DNA fraction formed by abundant transcripts using DSN. DSN is an enzyme purified from hepatopancreas of Kamchatka crab (Shagin et al. 2002); it shows a strong preference for cleaving double-stranded DNA and DNA in DNA-RNA hybrid duplexes, compared with single-stranded DNA and RNA. Moreover, the cleavage rate of short, perfectly matched DNA duplexes by this enzyme is considerably higher than that for non-perfectly matched duplexes of the same length. DSN finds use in various applications to isolate single-stranded DNA from complex nucleic acids, for example in cDNA normalization method (Zhulidov et al. 2004; Zhulidov et al. 2005; Bogdanova et al. 2008), for quantitative telomeric overhang determination (Zhao et al. 2008), and for SNP detection (Shagin et al. 2002).

DSN normalization is performed before cloning cDNA and includes cDNA denaturation followed by re-hybridization of denatured double strand cDNA. For each specific transcript, the hybridization rate is proportional to the square of the transcript concentration because nucleic acid hybridization is a second order chemical reaction. Therefore, abundant transcripts convert to the double strand form more effectively than those that are less common, and the single strand cDNA fraction is equalized. After re-hybridization, ds-cDNA is hydrolyzed by DSN, whereas the ss-cDNA fraction remains unchanged. This fraction can be amplified and used for construction of a normalized cDNA library or immediate high-throughput sequencing.

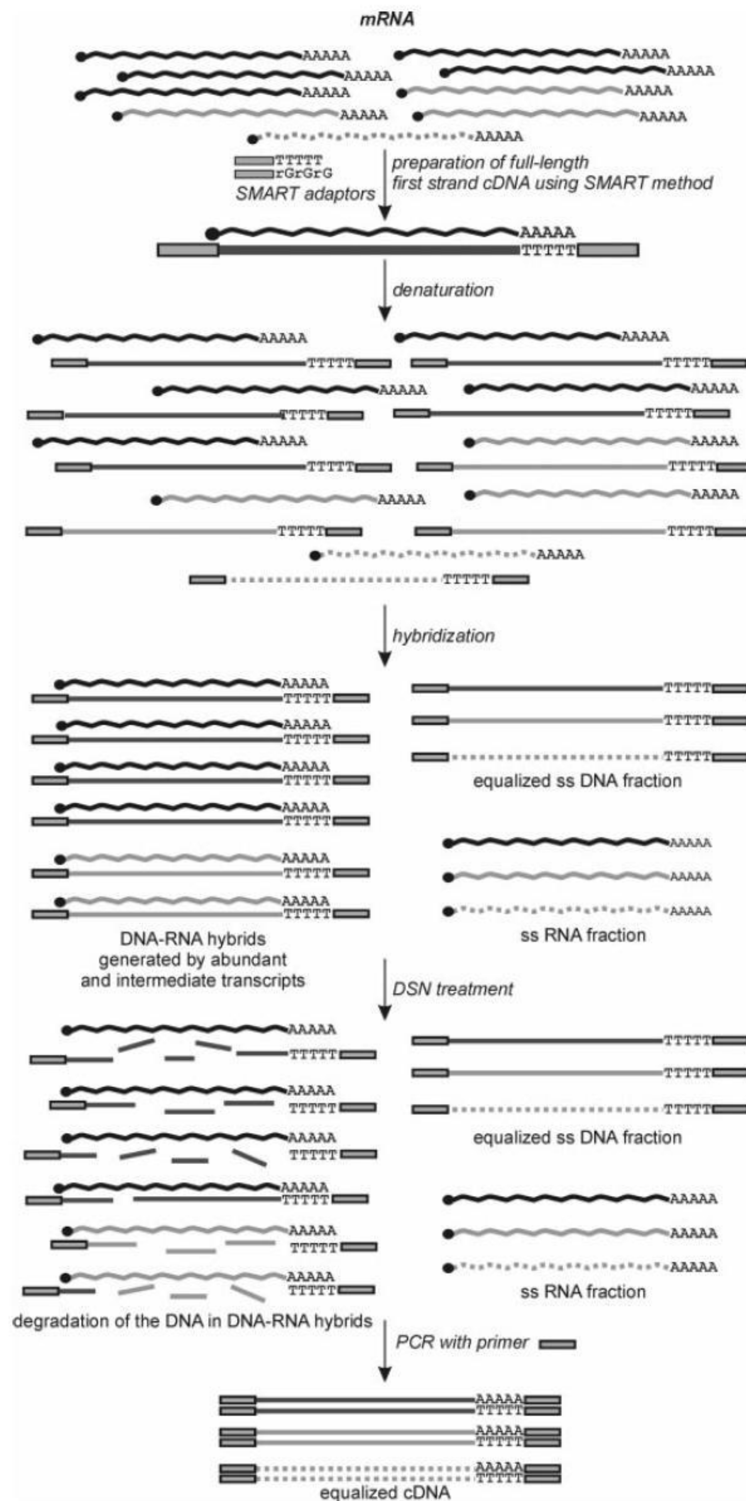



**Fig. 1.18** Schematic representation of DSN normalization process starting with mRNA amplified following SMART$^{TM}$ protocol. Black lines are abundant transcripts; gray lines are intermediate transcripts and dotted lines are rare transcripts (Zhulidov et al. 2004).

DSN normalization has become a standard methodology because of its simplicity, applicability to total RNA and availability. The efficacy of this

method was largely demonstrated not only in full-length enriched normalized cDNA library but also in sequencing projects, employing the new parallel sequencers such as the 454 Life Sciences GS 20 model using different model organisms like plants, insects, mollusks, nematodes and fish (Qin et al. 2010; Yi et al. 2011).

Autoimmune diseases are estimated to afflict more than 5% of the population worldwide (Bright 2007) and for the majority the etiology is still unknown.

To identify potential autoantigens of autoimmune diseases, transcriptomics (microarray analysis) or proteomic approaches (electrophoresis 2-DE and mass spectrometry) and display technologies can be used. The display technologies have the peculiar characteristic of associating phenotype to genotype. Every particle that displays a certain protein (phenotype) leads its own genetic information (genotype). So, the particle selection, on the basis of the binding properties of the polypeptide displayed on the particle surface, simultaneously results in the isolation of the gene encoding the selected polypeptide. In the case of "phage display" the expression of polypeptide on the surface of phage particles was obtained by cloning the exogenous gene in fusion with a gene encoding for one of the phage coat proteins (g8p or g3p). The resulting phage can be selected for its ability to bind a specific target protein.

The production of phage display libraries, from cDNA, allows to obtain the ORFs repertoire of a cell or tissue, including the different forms derived by alternative splicing. In particular, it is possible to create libraries of cDNA fragments coding for peptides that are really expressed and could therefore correspond to potential epitopes. To achieve this kind of libraries we need a system able to couple the production of high quality libraries and ORF selection.

In this context the purpose of this PhD project is:

- The development of an innovative protocol, for the construction of ORF-enriched and normalized cDNA fragments phage display libraries from human brain, that comprises:
    a) cDNA normalization;

b) normalized cDNA fragmentation to obtain peptide with appropriate size;
c) enrichment in ORFs fragments by engineering of an innovative vector.

The prototype of inflammatory autoimmune diseases of the Central Nervous System (CNS) is the Multiple Sclerosis (MS), a pathology that is characterized by inflammatory phase followed by chronic neurodegenerative evolution. The MS has been chosen to test the system with the aim of finding new candidate biomarkers for the disease.

The final goal is the biopanning of the ORF-enriched and normalized cDNA fragments phage display libraries from human brain with immunoglobulins from samples (sera or liquor) of Multiple Sclerosis patients.

## Bacterial strains

The bacterial strains used, were:

- *Escherichia coli* One Shot® MAX Efficiency® DH5α™-T1$^R$ Competent Cells (Invitrogen): F- φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F)U169 *rec*A1 *end*A1 *hsd*R17(rk-, mk+) *pho*A *sup*E44*thi*-1 *gyr*A96 *rel*A1 *ton*A for cDNA human brain library construction;
- *Escherichia coli* DH5αF' (Gibco BRL): F´/*endA1 hsd17* ($r_K^- m_K^+$) *supE44 thi-1 recA1 gyrA* ($Nal^r$) *relA1*Δ (*lacZYA-argF*) *U169 deoR* (F80*dlacD*-(*lacZ*)*M15*), for phage propagation;
- *Escherichia coli* XL1-Blue: F´:: Tn*10 proA$^+$B$^+$ lac1$^q$* Δ *(lacZ) M15/recA1 endA1 gyrA96* ($Nal^r$) *thi hsdR17* ($r_K^-$ $m_K^+$) *glnV44 relA1 lac*, for the expression of recombinant proteins.

## 3.1 Plasmid pPE3 construction

The plasmid pEP3 derives from pPAO2 (Zacchi et al. 2003) specifically modified, respect to other pEP version, to improve the cloning efficiency. Compare to pPAO3, the phagemid pEP1 has a new polylinker sequence with the "Strep-tag II" and the *Spe*I and *Pst*I restriction sites. The subsequent version, pEP2, was modified to perform an oriented cloning. The last version, pEP3, derivative from pEP2, lacks of gene III, for the recombinant phages production.

pEP3 was generated as follows: pEP2 was digested with 8 units of *Kas*I enzyme (*NEB*) and with 10 units of *Not*I HF enzyme (*NEB*), in presence of buffer 4 1X. Both enzymes were incubate at 37 °C for 2 hours and at 65 °C for 20 minutes. Digested pEP2 was excised from agarose gel 1%, to remove β-lactamase and gene III, and purified using the *Nucleospin Extract II kit* (*Macherey-Nagel GmbH & Co.KG*). At the same time, β-lactamase gene was obtained by amplifying on the commercial vector pBlueScript (*Invitrogen*) with

specific primers designed using DNA sequences available on BLAST database. To 1 µL of pBlueScript mini preparation were added 49 µL of the following mix: 10 µL buffer I 10X, 4 µL dNTPs 2.5 mM, 2.5 µL oligo BLApblu1 10 µM, 2.5 µL oligo BLApblu2 10 µM, 0.25 µL Taq *AccuPrime* (*Invitrogen*, 2.5 U/ µL) and 29.75 µL sterile $H_2O$. The amplification program was: 95 °C for 3 minutes (initial denaturation); 30 cycles at 95 °C for 30 seconds, 55 °C for 30 seconds, 72 °C for 45 seconds; 72 °C for 5 minutes (final extension). The PCR product was controlled on agarose gel 1%. PCR amplified β-lactamase was successively digested with 8 units of *Kas*I enzyme (*NEB*) and with 10 units of *Not*I HF enzyme (*NEB*), in presence of buffer 4 1X. Both enzymes were incubate at 37 °C for 2 hours and at 65 °C for 20 minutes. pEP2, cut *Kas*I-*Not*I HF and treated with Antarctic Phosphatase (*NEB*, 5 U/µL), was mixed with amplified β-lactamase gene, cut *Kas*I-*Not*I HF, at a molar ratio of 1:5 and ligated at 16°C overnight (O/N) in presence of 20 units of T4 DNA ligase (*NEB*). The ligation mixture was extracted with an equal volume of 50:50 (v/v) phenol: chloroform followed by ethanol precipitation in presence of 0.3 M sodium acetate pH 5.2 and glycogen and the resulting DNA pellet was resuspended in 10 µL of sterile water. The purified ligation was used to transform 40 µL of electrocompetent DH5αF´ cells. The electrophoration parameters were 2000V for 5.2-6.2 msec. After the electric impulse, was added 500 µL of SOC media to the electrophorated cells which were then grown at 37°C for 1 hour. The transformation mixture was plated on 2xYT plates containing 25 µg/mL of Chloramphenicol and incubated at 37ºC overnight. Some resulting colonies were randomly picked and sequenced using the primer pEPSEQ.

BLApblu1: 5’-GGGTGCGGCCGCTCTCACCCAGAAACGCTGGTGA-3’

BLApblu2: 5’-GTGAGGCGCCTCACCAATGCTTAATCAGTGAG-3’

pEPSEQ: 5´-CCGCTGGATTGTTATTACTC-3´

2xYT media: 16 g bacto-trypton, 10 g bacto-yeast extract, 5 g NaCl (for 1 litre of liquid media) + 15 g bacto-agar (for 1 liter of solid media). Sterilized in autoclave.

SOB media: 20 g bacto-trypton, 5 g bacto-yeast extract, 0.5 g NaCl, 1 M KCl 2.5 mL. Sterilized in autoclave and add 5 M $MgCl_2$ 2 mL, 1M $MgSO_4$ 20 mL.

SOC media: SOB + 1% glucose.

Chloramfenicol: 34 mg/mL in ethanol 100% and sterilized by filtration.

## 3.2 Electrocompetent cells preparation

A dilution 1/100 of DH5αF' grown overnight were inoculated in 200 mL of 2xYT and was grown at 37°C to O.D.$_{600nm}$ = 0.8-1. The bacterial culture was stopped for, at least, 20 minutes on ice and then was centrifuged at 4200 rpm 10 °C for 12 minutes. The pellets were washed 4 times using an equal volume of 10% glycerol and once using half volume. For each washing the pellet was resuspended and was centrifuged at 4200 rpm 10 °C for 12 minutes. In the penultimate step the pellet was resuspended and centrifuged at 4000 rpm 10 °C for 12 minutes. The final pellet was resuspended in 250 μL of 10% glycerol and divided in aliquots of 40 μL stored at -80°C.

## Genomic library

## 3.3 Preparation of genomic DNA from *Escherichia coli* One Shot®

*Escherichia coli* One Shot® was grown overnight at 37 °C in 5 mL of 2xYT. The bacterial culture was centrifuged at 3000 rpm 8 °C for 10 minutes. The pellet was resuspended in 250 μL of STE solution, after first was added lysozyme in a final concentration of 10 mg/mL and incubated for 10 minutes at 37 °C, and then 5 μL of RNase A (10 mg/mL) and incubated for 10 minutes at 56 °C. 250 μL of TNS solution was added and incubate for 30 minutes at 60 °C. The genomic DNA was extracted with an equal volume of phenol: chloroform: isoamyl alcohol (25:24:1) followed by ethanol precipitation and the resulting DNA pellet was resuspended in 500 μL of sterile water. The genomic DNA was analyzed by agarose gel electrophoresis.

STE solution: saccharose 8%, TrisHCl 50 mM, EDTA 50 mM pH8 in $H_2O$ mQ and sterilized by filtration.

TNS solution: TrisHCl 50 mM, NaCl 100 mM, SDS 2% in $H_2O$ mQ and sterilized by filtration.

## 3.4 Archeobacterial histones dimer production and purification

pPR2 is the expression vector in which the gene encoding for HMf (Archaeal histone from *Methanothermus fervidus*) dimer has been cloned.

For the histones dimer protein expression, the clone of *E. coli* XL1-Blue carrying the recombinant *hmf* gene was grown in 1 L of 2xYT with Ampicillin 100 µg/mL; at $O.D._{600\ nm}$=0.5 it was induced with IPTG 0.5 mM (final concentration) overnight at 30°C. Protein purification was performed as described in *Strep*-tag II purification protocol IBA (using all the solution without EDTA).

Ampicillin: 50 mg/mL in $H_2O$ mQ and sterilized by filtration.

IPTG 1 M: 2.38 g IPTG in 10 mL $H_2O$ mQ and sterilized by filtration.

## 3.5 Genomic DNA digestion by nuclease and archeobacterial histones dimer protection

Genomic DNA was incubated with the purified archeobacterial histones, 1:2 ratio, in presence of 100 mM NaCl, 1 mM $CaCl_2$, 50 mM Tris Acetate pH 8.8 to a final volume of 100 µL. The sample was incubated at 37 °C for 20 minutes, than the digestion was performed by adding 0.005 units of *MNase*I (Micrococcal nuclease from *Staphylococcus aureus*, *Sigma* #N3755) at the same temperature for 2 minutes and was stopped adding 10 mM EDTA pH 8. The DNA was purified from the reaction mixture by phenol/chloroform extraction followed by ethanol precipitation in presence of 0.3 M sodium acetate pH 5.2 and glycogen. The DNA pellet was resuspended in 10 µL of sterile water.

During the sample digestion, same parameters could be changed in order to obtain DNA fragments of a useful size: DNA/HMf ratio, DNA quantity, digestion incubation time, *MNase*I amount and DNA sample. Other parameters, e.g. pH and ionic strength, of incubation reaction were standard, because previously optimized. The fragmented DNA was analyzed by agarose gel 2.5% electrophoresis to examine fragments size distribution.

## 3.6 Genomic library construction

The DNA fragment ends were repaired and blunt-ended using the *End-It DNA End-Repair kit* (*Epicentre Biotechnologies*) and purified using the *Nucleospin Extract II kit* (*Macherey-Nagel GmbH & Co.KG*). EcoLPE60 linkers (Eco60LPE1 and Eco60LPE2) were ligated to the blunt end fragments in a reaction mixture containing EcoLPE60 linkers (in equimolar ratio with DNA, pre-incubated at 95 °C for 2 minutes, 65 °C for 10 minutes, and room temperature for 45 minutes to allow their annealing), 1X T4 DNA ligase buffer, DTT 0.01 M and 20 units of T4 DNA ligase (*NEB*). The reaction mix was carried out at 16°C overnight. To remove unligated linkers or eventually linkers dimers and to create specific sticky ends for cloning in pEP3, the O/N treated DNA was digested in a reaction mixture containing 10 units of *Eco*RI HF enzyme (*NEB*) and 10 units of *Asc*I enzyme (*NEB*) in presence of buffer 4 1X (*NEB*). The reaction was incubate at 37 °C for 3 hours and at 65 °C for 20 minutes. The genomic digested DNA fragments were separated by electrophoresis in 1.5% TAE 1X agarose gel. DNA fragments in the appropriate range (60-300 bp) were excised from the gel and purified using the *Nucleospin Extract II kit* (*Macherey-Nagel GmbH & Co.KG*). The purified DNA fragments were cloned in pEP3 vector cut *Eco*RI and treated with Antarctic Phosphatase (*NEB*) in a reaction with a molar ratio vector/DNA of 1:10. The reaction mixture was purified with the *Nucleospin Extract II kit* (*Macherey-Nagel GmbH & Co.KG*). Tenfold 5 µL of purified ligation reaction were used to transform 40 µL of DH5α One Shot (*Invitrogen*) electrocompetent cells. The electrophoration parameters were 2000V for 5.2-6.2 msec. After the electric impulse, was added 500 µL of SOC media to the electrophorated cells which were then grown at 37°C for 2 hours. The transformation mixture was plated on 2xYT plates containing 25 µg/mL of Chloramphenicol and 12 µg/mL of Ampicillin and incubated at 37 ºC overnight. The resulting colonies were randomly picked and analyzed for inserts presence by PCR with primers pEPSEQ and pEPBACK which anneal external to the cloning site. PCR products were analyzed by agarose gel electrophoresis.

The genomic DNA library was collected in five aliquots of 1 mL and stored at -80°C.

Eco60LPE1: 5’-GTGTGAATTCGGCG-3’
Eco60LPE2: 5’-CGCCGAATTCACAC-3’
pEPSEQ: 5´-CCGCTGGATTGTTATTACTC-3’
pEPBACK: 5´-AGTACTATCCAGGCCCAGCAG-3´

## 3.7 Library sequencing

96 clones, randomly chosen were sequenced using the primer pEPSEQ.

## Human Brain phage display library

## 3.8 Preparation of human brain cDNA

A cDNA library was constructed from 1 µg of Human Brain (HB) Poly $A^+$ RNA (*Clontech*; Cat.N. # 636102). For the cDNA synthesis SMARTer$^{TM}$ PCR cDNA synthesis (*Clontech*) and MINT cDNA synthesis (*Evrogen*) protocol were followed. The first strand synthesis of the cDNA was performed using 1 µg of Human Brain (HB) Poly $A^+$ RNA, 1.5 µL primer SMART24oI 15 µM, 1 µL primer SMART2A3 10 µM and 1.5 µL RNase free sterile $H_2O$. The reaction was incubate at 72°C for 2 minutes and in ice for 2 minutes. To this reaction were added 4.5 µL of the following mix: 2 µL first strand buffer 5X, 1 µL DTT 20 mM, 1 µL dNTPs 10 mM and 20 units of RNase OUT (*Invitrogen*) and incubated at 25 °C for 10 minutes. Finally 10 units of SuperScript$^R$ reverse transcriptase III (*Invitrogen*) were added and incubated at 25 °C for 1 hour. The first strand sample reaction was subjected to the second strand synthesis and amplification. To 2 µL of cDNA first strand were added 98 µL of the following mix: 10 µL 10X PCR Advantage 2 buffer, 2 µL dNTPs 10 mM, 2.4 µL primer SMARTPCR 10 µM, 2 µL 50X Advantage 2 polymerase mix (*Clontech*) and 81.6 µL RNase free sterile $H_2O$. The amplification program was: 95 °C for 1 minutes (initial denaturation); 15 cycles at 95 °C for 15 seconds, 65 °C for 30 seconds, 68 °C for 6 minutes; 72°C for 5 minutes (final extension). A 2 µL of the amplified cDNA were analyzed by agarose gel 1.5% electrophoresis

and all the rest of sample was purified using the *Nucleospin Extract II kit* (*Macherey-Nagel GmbH & Co.KG*).

SMART2AoI: 5'-AAGCAGTGGTATCAACGCAGAGTACGCGGG-3'
SMART2A3: 5'-AAGCAGTGGTATCAACGCAGAGTAC(T)30VN-3'
SMARTPCR: 5´-AAGCAGTGGTATCAACGCAGAGT-3'

## 3.9 cDNA normalization

Duplex-specific nuclease activity (DSN, *Evrogen*, 1 U/µL) was checked before use, in according to TRIMMER (*Evrogen*) protocol.

To 1200 ng of double strand human brain cDNA were added 4 µL hybridization buffer 4X and RNase free sterile $H_2O$ to final volume of 16 µL. The mixture was divided into four sterile PCR tubes and incubate at 98 °C for 2 minutes and at 68 °C for 5 hours and 30 minutes. 5 µL of preheated at 68 °C DSN master buffer were added to each tube containing hybridized cDNA and incubate at 68 °C for 10 minutes. To the reactions tube was added 1 µL DSN enzyme as specified in the table below and incubate at 68 °C for 25 minutes.

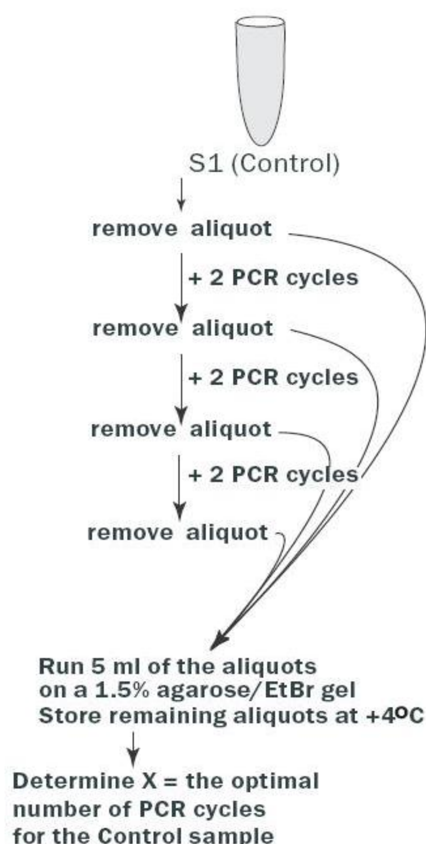
| Component/Tube | Tube 1 | Tube 2 | Tube 3 | Tube 4 |
|---|---|---|---|---|
| DSN enzyme in storage buffer | 1 µL | _ | _ | _ |
| 1:3 DSN dilution | _ | 1 µL | _ | _ |
| 1:2 DSN dilution | _ | _ | 1 µL | _ |
| DSN storage buffer | _ | _ | _ | 1 µL |

DSN was inactivated with 10 µL of DSN stop solution and incubated at 68 °C for 5 minutes. The tubes were extracted from the thermal cycler and placed on ice for 10 minutes and 20 µL of sterile water were added for each tubes.

To determine the optimal amplification cycles number, the DSN untreated sample was PCR amplified. To 1 µL of DSN untreated sample were added 49 µL of the following mix: 10 µL buffer CG 5X, 1 µL dNTPs 10 mM, 2.5 µL primer SMARTPCR 10 µM, 1.5 µL $MgCl_2$ 50 mM, 0.5 µL Taq *Phusion Hot Start* (*Finnzymes*, 2 U/ µL) and 31 µL sterile $H_2O$. The amplification program

was: 98 °C for 3 minutes (initial denaturation); from 7 to 21 cycles at 98 °C for 30 seconds, 55 °C for 30 seconds, 72 °C for 45 seconds; 72°C for 5 minutes (final extension). A 5 µL of the amplified cDNA, at different amplification cycles, were analyzed by agarose gel 1.5% electrophoresis.

To determine "X", where X is the optimal number of cycles required for amplification of each of the control tubes, was chosen the optimal number of PCR cycles ensures that the double strand cDNA was remained in the exponential phase of amplification. When the yield of PCR products stops increasing with every additional cycles, the reaction has reached its plateau.




After calculating the optimal cycles number X, it was calculate the number N, where N=X-7 (number of starting amplification cycles). The number of additional cycles was obtained by summing the number N to a number of cycles equal to 9. The total number of cycles for the amplification of DSN treated sample was so calculated: number of additional cycles + number of starting amplification cycles.

The DSN treated samples were PCR amplified. To 1 µL of DSN treated samples were added 49 µL of the following mix: 10 µL buffer CG 5X, 1 µL dNTPs 10 mM, 2.5 µL primer SMARTPCR 10 µM, 1.5 µL $MgCl_2$ 50 mM, 0.5 µL Taq *Phusion Hot Start* (*Finnzymes*, 2 U/ µL) and 31 µL sterile $H_2O$. The amplification program was: 98 °C for 3 minutes (initial denaturation); 22 cycles at 98 °C for 30 seconds, 55 °C for 30 seconds, 72 °C for 45 seconds; 72°C for 5 minutes (final extension). A 5 µL of the amplified cDNA were analyzed by agarose gel 1.5% electrophoresis.

Hybridization buffer 4X: Hepes 200 mM pH 7.5, NaCl 2M in $H_2O$ mQ and sterilized by filtration.

DSN master buffer 2X: TrisHCl 100 mM pH 8.0, $MgCl_2$ 10 mM, DTT 2 mM in $H_2O$ mQ and sterilized by filtration.

DSN storage buffer: TrisHCl 50 mM pH 8.0 in $H_2O$ mQ and sterilized by filtration.

DSN stop solution: EDTA 5 mM in $H_2O$ mQ and sterilized by filtration.

## 3.10 Testing the normalization process

### 3.10.1 Amplification of an high copy number gene: GAPDH

To evaluate the normalization process, the DSN treated and untreated samples were amplified using specific primers for an high copy number gene like GAPDH. To 0.5 µL of DSN treated samples (or untreated sample) were added 24.5 µL of the following mix: 2.5 µL buffer 10X, 0.5 µL dNTPs 10 mM, 1.25 µL primer GAPDH forward 10 µM, 1.25 µL primer GAPDH reverse 10 µM, 0.75 µL $MgCl_2$ 50 mM, 0.25 µL Taq *DNA Polymerase* (*Invitrogen*, 5 U/µL) and 18 µL sterile $H_2O$. The amplification program was: 95 °C for 5 minutes (initial denaturation); 30 cycles at 95 °C for 30 seconds, 61 °C for 30 seconds, 72 °C for 45 seconds; 72°C for 10 minutes (final extension). A 5 µL of the amplified cDNA were analyzed by agarose gel 1.5% electrophoresis.

GAPDH forward: 5'-GCAGGGGGGACGCAAAAGGG-3'
GAPDH reverse: 5'-TGCCAGCCCCAGCGTCAAAG-3'

### 3.10.2 Normalized library construction

The amplified DSN 1:3 treated sample was purified using the *Nucleospin Extract II kit* (*Macherey-Nagel GmbH & Co.KG*) and digested with 20 units of *Rsa*I enzyme (*Promega*) in presence of buffer C 1X and BSA 1X. The reaction was incubate at 37 °C for 1 hour and at 65 °C for 15 minutes and then purified using the *Nucleospin Extract II kit* (*Macherey-Nagel GmbH & Co.KG*). EcoLPE60 linkers (Eco60LPE1 and Eco60LPE2) were ligated to the blunt end fragments in a reaction mixture containing EcoLPE60 linkers (in 1:3 ratio with DNA, pre-incubated at 95 °C for 2 minutes, 65 °C for 10 minutes, and room temperature for 45 minutes to allow their annealing), 1X T4 DNA ligase buffer, DTT 0.01 M and 200 units of T4 DNA ligase (*NEB*). The reaction mix was carried out at 16°C overnight. To remove unligated linkers or eventually linkers dimers and to create specific sticky ends for cloning in the commercial

vector pBlueScript, the O/N treated DNA was digested in a reaction mixture containing 10 units of *Eco*RI HF enzyme (*NEB*) and 10 units of *Asc*I enzyme (*NEB*) in presence of buffer 4 1X (*NEB*). The reaction was incubate at 37 °C for 3 hours and at 65 °C for 20 minutes.

The digested normalized cDNA was separated by electrophoresis in 1.5% TAE 1X agarose gel. cDNA from 200 bp to 800 bp was excised from the gel and purified using the *Nucleospin Extract II kit* (*Macherey-Nagel GmbH & Co.KG*). The purified cDNA was cloned in pBlueScript vector (cut *Eco*RI and treated with 5 units of Antarctic Phosphatase (*NEB*)) in a reaction with a molar ratio vector/DNA of 1:5. The reaction mixture was purified with the *Nucleospin Extract II kit* (*Macherey-Nagel GmbH & Co.KG*). The purified ligation was divided in 4 aliquots, each one was used to transform 40 µL of DH5α One Shot (*Invitrogen*) electrocompetent cells. After the electric impulse, was added 500 µL of SOC media to the electrophorated cells and they were grown at 37°C for 2 hours. The transformation mixture was plated on 2xYT plates containing 100 µg/mL of Ampicillin and incubated at 37 ºC overnight. The resulting colonies were randomly picked and analyzed for inserts presence by PCR with oligo T3 and oligo T7. PCR products were analyzed by agarose gel electrophoresis.

T3: 5'-AATTAACCCTCACTAAAGGG-3'
T7: 5'-GTAATACGACTCACTATAGGGC-3'

**Sequencing**

96 clones, randomly chosen were sequenced using the primer T7.

## 3.11 Normalized human brain cDNA digestion by nuclease and archeobacterial histones dimer protection

Normalized cDNA from human brain was incubated with archeobacterial histones dimer , 1:2 ratio, in presence of 100 mM NaCl, 1 mM $CaCl_2$, 50 mM Tris Acetate pH 8.8 to a final volume of 100 µL. The sample was incubated at 37 °C for 20 minutes. The digestion was performed by adding 0.005 units of *MNase*I (*Sigma* #N3755) at the same temperature for 5 minutes and was stopped adding 10 mM EDTA pH 8. The DNA was purified from the reaction

mixture by phenol/chloroform extraction followed by ethanol precipitation in presence of 0.3 M sodium acetate pH 5.2 and glycogen. The DNA pellet was resuspended in 10 µL of sterile water. The fragmented DNA was analyzed by agarose gel 2.5% electrophoresis to examine fragments size distribution.

## 3.12 Normalized cDNA fragments library from human brain in pEP3

It was performed as previously described: cDNA fragments were end repaired using the *End-It DNA End-Repair kit* (*Epicentre Biotechnologies*) and purified using the *Nucleospin Extract II kit* (*Macherey-Nagel GmbH & Co.KG*). EcoLPE60 linkers (Eco60LPE1 and Eco60LPE2) were ligated to the blunt end fragments in a reaction mixture containing EcoLPE60 linkers (in 1:2 ratio with DNA, pre-incubated at 95 °C for 2 minutes, 65 °C for 10 minutes, and room temperature 45 minutes to allow their annealing), 1X T4 DNA ligase buffer, DTT 0.01 M and 20 units of T4 DNA ligase (*NEB*). The reaction mix was carried out at 16°C overnight. To remove unligated linkers or dimers and to create specific sticky ends for cloning in pEP3, the O/N treated DNA was digested in a reaction mixture containing 10 units of *Eco*RI HF enzyme (*NEB*) and 10 units of *Asc*I enzyme (*NEB*) in presence of buffer 4 1X (*NEB*). The reaction was incubate at 37 °C for 3 hours and at 65 °C for 20 minutes. The genomic digested DNA fragments were separated by electrophoresis in 1.5% TAE 1X agarose gel. DNA fragments in the appropriate range (60-300 bp and 300-600 bp) were excised from the gel and purified using the *Nucleospin Extract II kit* (*Macherey-Nagel GmbH & Co.KG*). The purified DNA fragments were cloned in pEP3 vector cut *Eco*RI and treated with Antarctic Phosphatase (*NEB*) in a reaction with a molar ratio vector/DNA of 1:10. The reaction mixture was purified with the *Nucleospin Extract II kit* (*Macherey-Nagel GmbH & Co.KG*). Tenfold 3 µL of purified ligation reaction were used to transform 40 µL of DH5α One Shot (*Invitrogen*) electrocompetent cells. After the electric impulse, were added 500 µL of SOC media to the electrophorated cells which were then grown at 37°C for 2 hours. The transformation mixture was plated on ten 2xYT plates containing 25 µg/mL of Chloramphenicol and 12 µg/mL of Ampicillin and incubated at 37 ºC overnight. The normalized

cDNA fragments libraries (60-300 bp and 300-600 bp) were collected and grown into 50 mL of 2xYT added with 25 µg/mL of Chloramphenicol at 37 ºC for 3 hours. The plasmid DNA libraries (300-600 bp) was recovered with the *Nucleospin Plasmid kit* (*Macherey-Nagel GmbH & Co.KG*).

## 3.13 Subcloning of the ORF-enriched and normalized cDNA fragments library in phagmidic pDAN5 vector

Plasmids DNA from the HB 300-600 pb library were digested at 50 °C for 1 hour and at 80 °C for 20 minutes with 8 units of *Bss*HII enzyme (*NEB*) in presence of buffer 3 1X and purified using the *Nucleospin Extract II kit* (*Macherey-Nagel GmbH & Co.KG*). To recover ORF-enriched cDNA fragments, linearized plasmids were digested at 37 °C for 2 hours and at 65 °C for 20 minutes with 10 units of *Not*I-HF enzyme (*NEB*) in presence of buffer 4 1X and BSA 1X. The digestion products were separated by electrophoresis in 1.5% TAE 1X agarose gel and DNA fragments in the appropriate range (300-600 bp) were excised from the gel and purified using the *Nucleospin Extract II kit* (*Macherey-Nagel GmbH & Co.KG*). The purified DNA fragments from 300 to 600 bp were cloned in pDAN5 vector cut *Bss*HII-*NotI*-HF and treated with Antarctic Phosphatase (*NEB*) in a reaction with a molar ratio vector/DNA of 1:10, 1X T4 DNA ligase buffer, DTT 0.01 M and 20 units of T4 DNA ligase (*NEB*). The reaction mix was carried out at 16°C overnight and purified with the *Nucleospin Extract II kit* (*Macherey-Nagel GmbH & Co.KG*). Tenfold 3 µL of purified ligation reaction were used to transform 40 µL of DH5α One Shot (*Invitrogen*) electrocompetent cells. After the electric impulse, were added 500 µL of SOC media to the electrophorated cells which were then grown at 37°C for 2 hours. The transformation mixture was plated on ten 2xYT plates containing 100 µg/mL of Ampicillin and incubated at 37 ºC overnight. The ORF-enriched and normalized cDNA fragment library from human brain was collected in aliquots of 50 µL and stored at -80 °C.

**Sequencing**

17 clones, randomly chosen were sequenced using the primer pELBS.

pELBS: 5'-CAGGAAACAGCTATGAC-3'

## 3.14 Immunoglobulin purification from cerebrospinal fluid of Multiple Sclerosis patients

To identify the most reactive cerebrospinal fluid samples from patients with Multiple Sclerosis, an ELISA assay was performed.

Ninety-six-well plates (*Costar*) were coated overnight at 4°C with 100 µL of monoclonal anti human IgG- Fc specific antibody (*Sigma*, I2136), 10 µg/mL in PBS 1X. The plate was blocked with 120 µL of 2% BSA in PBS 1X for 1 hour at Room Temperature (RT). Then, wells were incubated with CSF diluted 1:2 in BSA for 1 h at RT and washed 3 times with PBS 1X-Tween 0.1% and 3 times with PBS 1X. A volume of 100 µL of peroxidase conjugate anti human IgG- Fc specific antibody (*Sigma*, A0170), diluted 1:3000 in 2% BSA-PBS 1X, was incubated for 1 h at RT. After washing the wells 3 times with PBS 1X-Tween 0.1% and 3 times with PBS 1X, 100 µL/well of 3,3´,5,5´-tetramethylbenzidine (TMB, *Invitrogen*) was added and the signal development was stopped with 100 µL/well 1 M HCl. The plates were read at 450 nm. The 25 most reactive CSF, divided into two populations, were used for the Immunoglobulin G purification. The purification was performed using *Montage$^R$ antibody purification kit* and spin columns with PROSEP$^R$-G Media (*Millipore*) in according with the manufacturer protocol. The purified IgGs were analyzed by SDS/PAGE 10% gel electrophoresis.

## 3.15 Selection of the ORF-enriched and normalized cDNA phage display library from human brain

Three groups of CSF samples were used to perform HB phage display library selections:

- **Pool 1**: Purified IgGs from CSF Multiple Sclerosis samples with $O.D._{450nm}>2.2$ in ELISA assay (8 woman and 3 men);
- **Pool 2**: Purified IgGs from CSF Multiple Sclerosis samples with $2>O.D._{450nm}>2.2$ in ELISA assay (8 woman and 6 men);
- **Pool 3**: Purified IgGs from pool 1 and 2.

For the selection of the human brain library with purified IgGs from CSF, two panning were performed. For each panning, immunotubes (*Nunc*, Roskilde, Denmark) were coated with 500 µL of purified IgGs from CSF at a concentration of 10 µg/mL in PBS 1X at 4°C over-night. After washing immunotubes once with PBS 1X, they were blocked with BSA 2% in PBS 1X for 1 hour at room temperature. The immunotubes were washed once with PBS 1X and PEG-purified phages (about $10^{10}$) were added. They were prepared from human brain library, diluted in an equal volume of 4% BSA-PBS 2X, incubated for 30 minutes at RT and then were added in the immunotubes for 30 minutes on a rotating platform, followed by 90 minutes of standing at RT. At the first panning, were performed 10 washes with PBS 1X-Tween 0.5% and 10 with PBS 1X. Bounded phages were eluted by adding 500 µL of DH5αF' *E. coli* cells at $O.D._{600nm}$= 0.5. The cells were infected with output phages for 45 minutes at 37 °C. Three dilutions (1:500, 1:5000, 1:50000) were plated on 2xYT agar plates containing Ampicillin 100 µg/mL, all the rest was infected with M13K07 helper phage (at a MOI of 20:1) and was grown overnight at 30 °C in 10 mL of 2xYT with Ampicillin 100 µg/mL and Kanamycin 25 µg/mL. For the second panning, the washes were 20 with PBS 1X-Tween 0.1% and 20 with PBS 1X. Input and output phages from each round of selection were titrated to determine the enrichment of specific clones, and the ratio of output/input phage was calculated. After selection, single clones were randomly picked and analyzed for inserts presence by PCR with primers pELBS and pGENE3 and were screened for their reactivity by phage-ELISA.

pELBS: 5'-CAGGAAACAGCTATGAC-3'
pGENE3: 5'-GTCGTCTTTCCAGACGTTAG-3'

## 3.16 Phage-ELISA

Individual selected clones were grown in 2xYT with Ampicillin 100 µg/mL and glucose 1% in 96-well round-bottomed plates (*Sarstedt*). When they enriched $O.D._{600}$= 0.5, each clones was infected with M13K07 helper phage at a MOI of 20:1 at 37 ºC for 45 minutes and was left in 2xYT with Ampicillin 100 µg/mL and Kanamycin 25 µg/mL at 30 ºC overnight to allow the production of

phages. Ninety-six-well flat bottomed plates (*Costar*) were coated overnight at 4 °C with 100 µL of goat anti human IgG-Fc specific (*Sigma*, I2136), 10 µg/mL in PBS 1X and blocked with 120 µL of 2% BSA-PBS 1X for 1 hour at RT. Then, wells were incubated with purified IgGs form CSF (diluted 1:3 in BSA 2%-PBS 1X) for 1 hour at RT and subsequently washed 3 times with PBS 1X-Tween 0.1% and 3 times with PBS 1X. Fifty microliters of each supernatant containing the phages of the individual selected clones were recovered, diluted in an equal volume of 4% BSA-PBS 2X, added into wells and incubated for 90 minutes a RT. After 3 washes with PBS 1X-Tween 0.1% and 3 with PBS 1X, 100 µL of a peroxidase conjugated anti-M13 monoclonal antibody (*Amersham/Pharmacia/Biotech*), diluted 1:3000 in 2% BSA-PBS 1X was incubated for 1 hour at RT. After washing the plates 3 times with PBS-Tween 0.1% and 3 times with PBS 1X, 100 µL/well of 3,3´,5,5´-tetramethylbenzidine (TMB, *Invitrogen*) was added and the color development stopped with 100 µL/well 1 M HCl. The plates were read at 450 nm.

**Sequencing**

The most reactive clones, showing high $O.D._{450\ nm}$ value, were analyzed for inserts presence by PCR with primers pELBS and pGENE3 and sequenced using the primer pGENE3.

## 3.17 Secondary phage-ELISA

Ninety-six-well flat-bottomed plates (*Costar*) were coated overnight at 4°C with 100 µL of 1:1000 sera samples, in PBS 1X, washed once with PBS 1X and blocked with 120 µL of 2% BSA- PBS 1X for 1 hour at RT. Then, wells were incubated with $10^{10}$ single PEG-precipitated Human Brain clone phagemids diluted in final 2% BSA-PBS 1X, for 90 minutes at RT. After 3 washes with PBS-Tween 0.1% and 3 with PBS 1X, 100 µL of peroxidase conjugated anti-M13 monoclonal antibody (*Amersham/Pharmacia/Biotech*), diluted 1:3000 in 2% BSA- PBS 1X was incubated for 1 hour at RT. After washing the plates 3 times with PBS-Tween 0.1% and 3 times with PBS 1X, 100 µL/well of 3,3´,5,5´-tetramethylbenzidine (TMB, *Invitrogen*) was added

and the color development stopped with 100 µL/well 1 M HCl. The plates were read at 450 nm.

## 4.1 Engineering of a new vectors to develop epitope libraries

In my hosting laboratory, a well characterized vector for the ORF (Open Reading Frame) selection, pPAO2 (Zacchi et al. 2003) was previously modified in two subsequent version (Bembich 2008) called pEP1 and pEP2. The last one, pEP2, was used as backbone for the engineering of pEP3 during my PhD.

pPAO2 vector allowed the ORFs selection by expressing polypeptides in fusion with the β-lactamase enzyme. Only the clones containing an ORF fragment expressed a functional β-lactamase are able to survive in presence of Ampicillin whereas, clones with "out of frame" fragment were suppressed because of Ampicillin toxicity. A limitation in the use of this vector was the characteristic of the cloning system used by pPAO2 vector that provides the T4 DNA Polymerase treatment. This methods could lead to an excessive cDNA degradation with consequent low cloning efficiencies (Bembich 2004). To overcome this limitation, starting from pPAO2, two distinct kind of vector versions, called respectively pEP1 and pEP2 were previously developed. In **pEP1** was introduced a new polylinker, containing additional sites for a restriction mediated cloning, that conferred some additional properties: (I) a *Spe*I restriction site for the generation of cohesive end for cloning, (II) a short tag of 8 amino acids (WSHPQFEK), called "Strep-tag II" (Korndorfer and Skerra 2002) upstream the *Spe*I cloning site for affinity purification of the phages that specifically carrying recombinant pIII and (III) two restriction sites (*Nhe*I and *Pst*I) for the direct ORF inserts subcloning in specific expression vector for Strep-tag II. **pEP2** was created, starting from pEP1, to increase from 1/18 to 1/9 the ratio of "in frame" inserts. This aim has been achieved by adding in the polylinker *Eco*RI and *Hind*III sites and using *Hind*III random primer (*Novagen*) for cDNA synthesis and oriented cloning (Bembich 2008). The polylinker is "out of frame" with β-lactamase and it presents a stop codon (TAG) in correspondence of *Spe*I site. pEP2 was designed to select only

those inserts that are in right orientation and 3n+2, while inserts 3n or 3n+1 introduce stop codon TAG immediately downstream of the cloning sites.

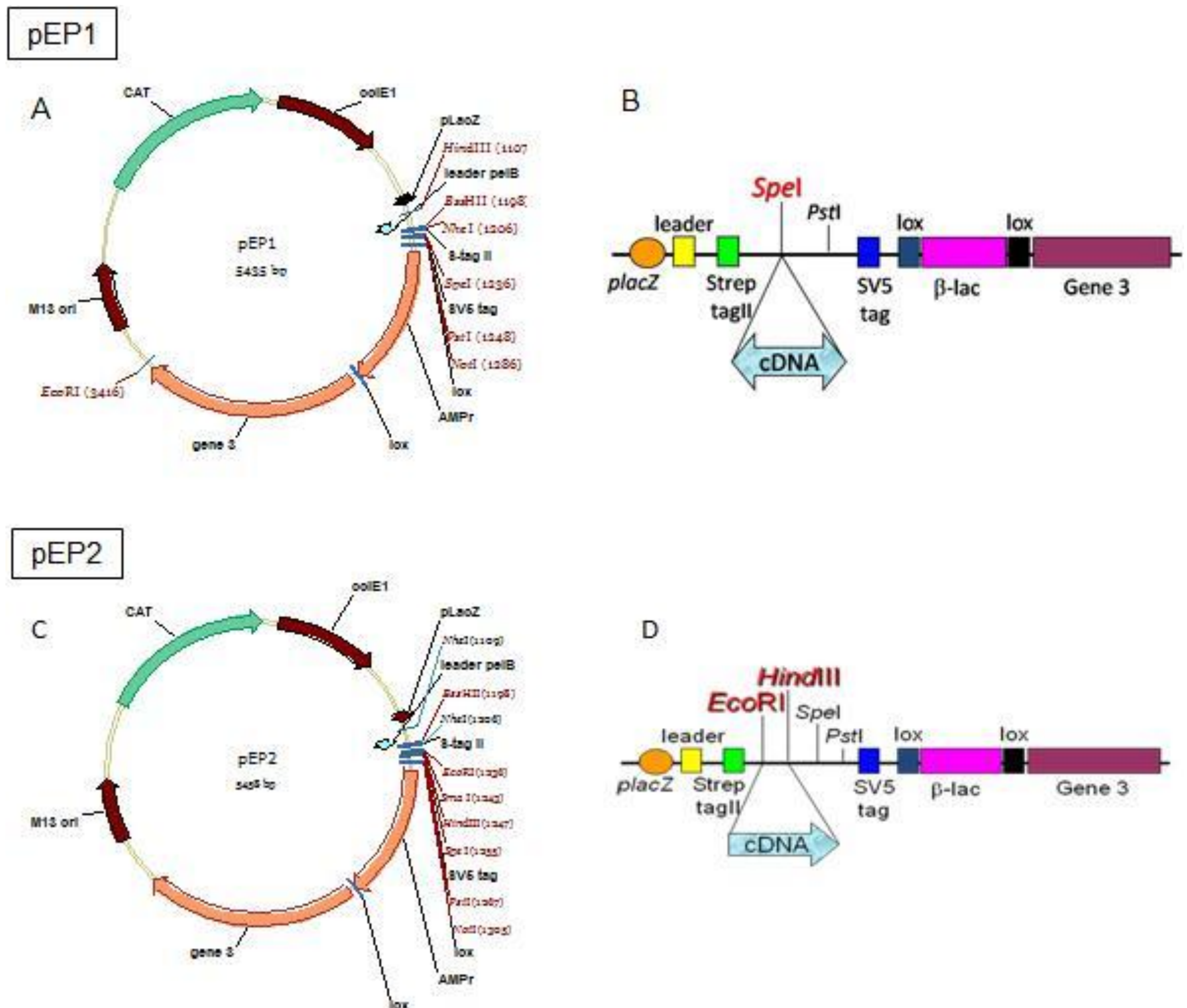



**Fig. 4.1** Schematic representation of pEP1 vector map (A), pEP1 polylinker (B), pEP2 vector map (C) and pEP2 polylinker (D). All the typical components of a phage display vector, able to select ORF are indicate both in pEP1 and pEP2: the constitutive Chloramphenicol resistance, the pLacZ promoter upstream the polylinker, a "leader" sequence downstream the promoter for the expression products periplasmic localization, the β-lactamase gene for ORFs selection flanked by two lox sequences for the Cre-mediate recombination, the gene III encoding for the phage coat pIII protein and a SV5 tag for the identification of the expression products using commercial antibodies anti-SV5.

Another important limitation for the selection of "in frame" clones using pPAO2 is the possible difficult folding of the β-lactamase enzyme when it is fused to exogenous polypeptides and to phage's pIII. To overcome this limitation, it is necessary the use of low Ampicillin concentration (12 μg/mL) with consequent potential contaminations by clones containing "out of frame" fragments.

In order to avoid the difficult folding of the β-lactamase enzyme the vector **pEP3** was generate deleting the gene III (see Materials and Methods for a detailed description of the used procedure). In this way ORFs selection and phages production were uncoupled. The final purpose was to obtain more representative epitope cDNA libraries of the entire expression pattern of the tissue from which the cDNA was synthetized.

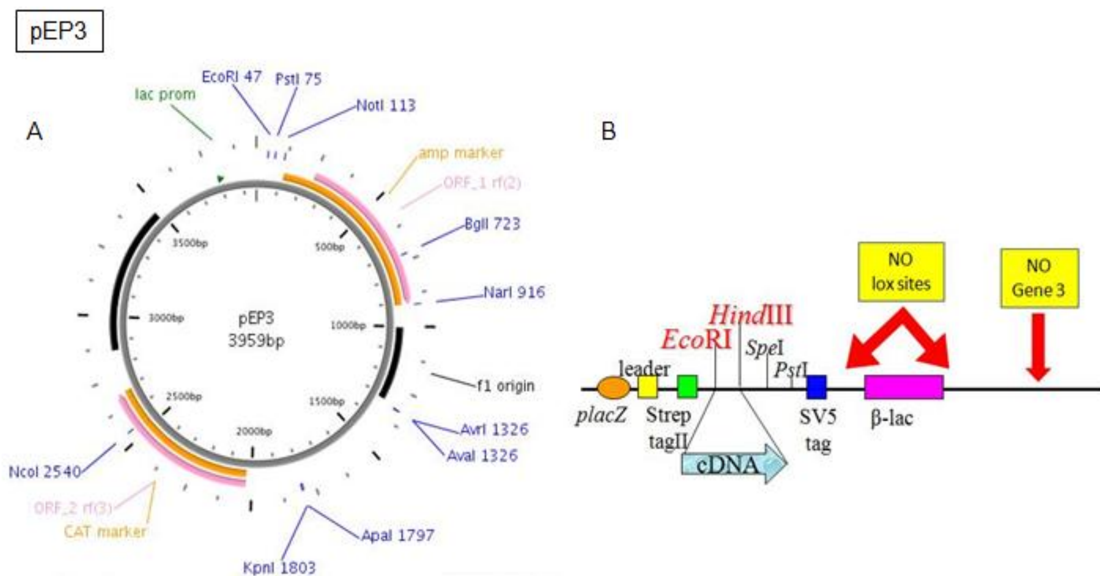



**Fig. 4.2** Schematic representation of pEP3 vector map (A) and pEP3 polylinker (B). pEP3, respect pEP2, has no gene III and lox sites for Cre-recombination.

To test the new vector an oriented cloning using both pEP2 and pEP3 vector was performed. A smear of *Eco*RI- *Hind*III cDNA fragments were cloned in both vectors and the results are shown in the table 4.1. Using pEP3, cloning efficiency and ORFs selection were increased.

| VECTOR | N°clones in CHL-AMP | N°clones in CHL | CHL-AMP/CHL RATIO |
|---|---|---|---|
| pEP2 | $6.3x10^2$ | $5.6x10^3$ | 1/9 |
| pEP3 | $5x10^3$ | $1.4x10^4$ | 1/3 |

**Tab. 4.1** The table shows the numbers of colonies obtained after transformation of electrocompetent *E. coli* DH5αONE SHOT cells with the ligation of oriented cDNA fragments in pEP2 and pEP3. Also the ratio “in frame”/ “out of frame” clones was indicated.

The cloning efficiency was increased about ten-fold using pEP3. One hypothesis could be that the optimisation of pEP3 vector, which means removing of gene III, gave origin to a smaller vector with a consequent better capability of been transformed. Further, the basal expression of fragments in fusion with gene III could be more toxic in the case of pEP2. The experimental ratio of “in frame” / “out of frame” clones with pEP2 was, as expected 1/9, whereas the ratio with pEP3 was 1/3. For the transformation the same bacterial strain and the same Ampicillin concentration (12 µg/mL) in the growth media were used. The use of low concentration of Ampicillin in the growth medium can led to contaminations in the collected “in frame” library by clones containing “out of frame” inserts.

We performed some tests using different Ampicillin concentration. For both vector, using an Ampicillin concentration of 12 µg/mL, clones number as indicated in the table 4.1 were obtained. Increasing the concentration of Ampicillin to 25 µg/mL, the number of clones was ten-fold reduced. Further, increasing the concentration to 50 µg/mL or more, only few clones (less than 10 clones using 200 ng of both vector) survived. On the basis of this results, according to literature data (Zacchi et al. 2003), 12 µg/mL Ampicillin was used for all subsequent experiment.

Without gene III, pEP3 doesn’t encode phage coat protein pIII so it is not able to produce recombinant pIII. In order to develop epitope phage display libraries for the expression of cDNA encoding ORF sequences, an appropriate phagemid vector must be used. Thanks to two specific restriction sites, flanking the pEP3 polylinker, all ORF selected inserts could be easily oriented sub-cloned into the phagemid vector pDAN5 (Sblattero and Bradbury 2000), maintaining the correct frame with gIII.

### Construction of ORF-selected library

The primary purpose of this thesis was to create a phage-display epitope library for the identification of biomarkers in autoimmune diseases of the Central Nervous System (CNS). This involved the improvement of our system

for the cloning of the entire repertoire of the molecules of the tissue target of the autoimmune response.

To represent all possible epitopes, we chose to generate a library of cDNA fragments of compatible size with the dimension of protein domain (Zacchi et al. 2003). Based on literature's data, the average length of the conformational epitopes is estimate around 300 base pairs, while for a linear epitope the best range is between 50-200 bp (Cui et al. 2003) so we chose to clone fragments having dimension to encode possibly both linear and conformational epitopes (over 70 amino acids) but maintaining a range in length not so high among the different fragments because in this way all the fragments may have the same probability of being cloned. Moreover, too short fragment (60 bp or less) could be positively select as "in frame" even when they don't correspond to real ORF; in fact, statistically, fragments that don't correspond to a real ORF present "stop" codons every 20 amino acids.

One of the aim of my PhD project was the production of a phage-display ORF-enriched and normalized cDNA library from human brain to discover new potential antigens involved in Central Nervous System autoimmune diseases. This initial purpose involved some preliminary steps to set (I) the new DNA fragmentation method and (II) the evolution of a specific protocol for the DNA cloning in the new vector version for the ORF selection (pEP3). We decided to set this steps not using cDNA form human brain because it was expensive and the quantity was not sufficient for several tests.

Initially, to check the new vector and the DNA fragmentation methods set up during my PhD project, an ORFs enriched library from total genomic DNA from *E. coli* was produced.

## 4.2 DNA fragmentation system

The first aim for the construction of an epitope library is the identification of a system to obtain DNA fragments with an appropriate size, useful to express linear and/or conformational epitopes. A perfect fragmentation ensures that the library contains representative copies of all genes of interest without imbalances in the representation of the fragments. For the construction of

genomic libraries, two methods for DNA fragmentation were mainly used: (I) mechanic fragmentation and (II) enzyme mediated fragmentation. Mechanical methods, as sonication, have the advantage that DNA fragmentation is totally random, but, on the other hand, it requires large amounts of DNA and the average size of DNA fragments is variable, fragmentation is weak and can be achieved with a minimum size fragments around 300 bp. In the case of cDNA, a mechanical method could degrade it completely because the fragmentation is not easily controlled. Another problem is that the mechanical fragmented DNA ends have a low cloning efficiency. The enzymatic DNA fragmentation can be carried out with restriction enzymes and nucleases. In this case, the fragmentation is not totally random; in fact, the length of the generated fragments depends on the number of recognized bases and every restriction enzyme has a sequence specificity. Another problem occurs if a target gene contains a recognition sites for the restriction enzyme used for the fragmentation. As a result, this gene is not represented in the library. To obviate this problem, a widely used method is the partial digestion, so, not all the recognition sequences in DNA are cut and the digestion time could be adjusted. Better results can be obtained using aspecific nucleases, even if in some cases these enzymes shown some specific recognition (Allan et al. 2012). But, the controlled use of these enzymes is difficult.

These problem can be resolved using a controlled fragmentation protocol (Azzoni et al. 2007) that used *Methanothermus fervidus* archaeal histones (HMf) to protect DNA from the *DNase*I digestion. The histones form a structure called NLS (Nucleosome Like Structure) very similar to the eukaryotic nucleosome. The purified HMf are able to form dimers, in solution, that wrap the DNA in a right-handed super helix and protect DNA sequences from nucleases cutting. They protect DNA sequences of 30 bp and, if increase the ratio HMf/DNA, its multiples. Varying HMf/DNA ratio, time of incubation and enzyme concentration, the fragments size can be modulate.

During my PhD, I performed this system that can rise to random DNA fragments with the possibility to modulate the fragments length.

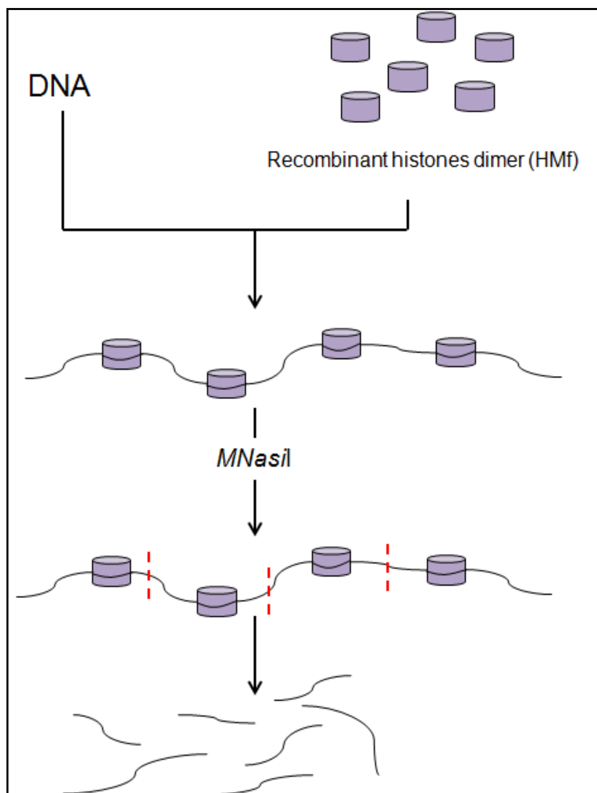



**Fig. 4.3** Schematic representation of DNA fragmentation with *MNase*I and protection of HMf. HMf wraps DNA and protects it from *MNase*I. After the removal of histones, fragments with predictable size were obtained.

For our aim we used a dimeric recombinant form of the archaeal histones, which features were compared with the monomeric form (previously tested in my hosting laboratory). For the comparison, both the monomeric and the dimeric recombinant form were produced and purified (see Materials and Methods). The purification was performed according with the *Strep-tag purification system* (*IBA*). Some changes were introduced to this commercial protocol. The solution used for the protein elution contains EDTA, a divalent ions chelating agent. Because cations act as *MNase*I cofactors, the enzyme used for the fragmentation, and are required for the correct enzyme working, the presence of EDTA in the purified histones collected could be a problem in the experiment reproducibility. So we decided to avoid a dialysis step hat can reduce the protein yield, and used solution without EDTA. We verified that this do not affect the efficiency of the purification system.

In figure 4.4 are shown the results of the purification of both monomeric and dimeric recombinant form.

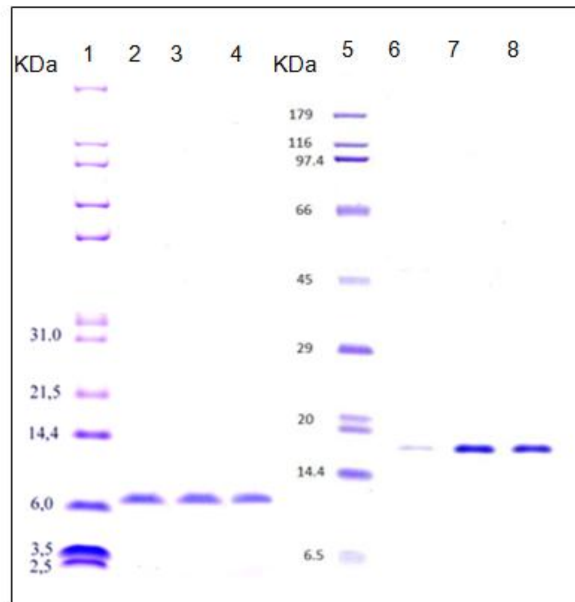
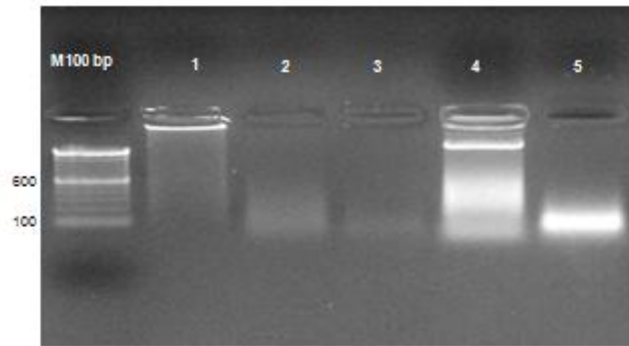



**Fig. 4.4** Gel electrophoresis SDS/PAGE 10% of the purified HMf. Purified HMf monomeric recombinant form (lane 2-4) and purified dimeric recombinant form (lane 6-8) can be observed. In lane 1 and 5 molecular weight. Histones monomer presents a molecular weight around 6.5 KDa, histones dimer has a molecular weight of about 16 KDa.

## 4.3 Comparison of DNA protection using histones monomer and dimer

For the comparison of DNA protection with histones monomer and dimer in presence of *MNase*I enzyme, a fragmentation test was performed using a cDNA SMART amplified from total CHO cells RNA. cDNA was incubated with the purified archeobacterial histones in specific ratio to allow the binding of cDNA to the histones, and then the enzyme *MNase*I was added. The digestion reaction was incubated for specific time to identify the perfect fragmentation time to obtain fragment with useful length. To identified the adequate time conditions, some tests were performed (data not shown) and the results indicated that the best time range for digestion reaction is between 2 and 5 minutes. We also want to investigate the possibility to control the digestion reaction changing DNA/HMf ratio, to obtain fragments of different size. The generated cDNA fragments for each tests were analyzed on agarose gel electrophoresis. The ratio used and the obtained results are showed in figure 4.5.



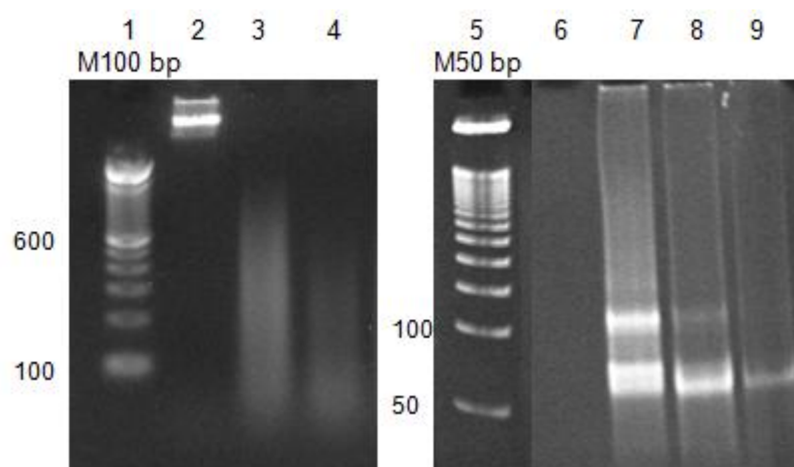

| M 100 bp | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| HMf form | / | Monomer | Monomer | Dimer | Dimer |
| DNA/HMf | / | 1:4 | 1:1 | 1:2 | 1:1 |

**Fig. 4.5** Electrophoresis on 2.5% agarose gel. *MNase*I protection test in presence of monomeric (lane 2, 3) and dimeric (lane 4, 5) HMf form. In lane 1 cDNA from CHO cells. In the table below, the characteristics of each different digestion tests are indicate. For each test, the ratio between DNA and HMf is the only parameter that we decide to change. The amount on *MNase*I enzyme was 0.005 U for each experiments, and the digestion time was 5 minutes.

As shown in figure 4.5, for both monomeric and dimeric form, if HMf decrease, a decrement of DNA protection was obtained; on the contrary, if HMf increase an increment of fragment size (protection) was achieved. In vivo, the HMfs make a dimer that is able to bind 30 bp of DNA (Sandman et al. 1990); however in vitro has been demonstrated that the tetrameric form is also present and bind unit of 60 bp. Using a 1 µg of HMf (ratio DNA/HMf 1:1) the obtained fragments size was 60 bp, for the monomeric and also dimeric form. Comparing the results obtained with the different forms of HMf, we demonstrated that the dimeric form protects the cDNA better than the monomeric one; in fact, as showed in figure 4.5 (lanes 3 and 5), when the same DNA/HMf ratio was used, using the dimeric form more cDNA was recovered. As previously shown, the digestion reaction reached an equilibrium and DNA could not be beyond the 60 bp. Another protection test was performed increasing the HMf concentration (data not shown), without modify the DNA concentration. In this case, a large smear of DNA fragments was obtained, confirming the multimeric unit formed by HMfs association and the high concentration of fragments around 60, 120, 180 bp (60 bp and multiple of it), as expected.

## 4.4 Preparation of genomic DNA and cloning

The genomic DNA was obtained from an overnight growth of *E. coli* One Shot[R] (see Materials and Methods). The resulting genomic DNA was subjected to controlled fragmentation in order to obtain DNA fragments of homogeneous size. The fragmentation was carried out using the over described fragmentation system developed during my PhD project, in presence of DNA/HMf dimers in specific ratio, definite amount of *MNase*I, and specific digestion time. The condition are shown in the figure 4.7. A smear ranging a population of multiple of 60 bp DNA fragments was generated and checked on agarose gel electrophoresis.




| 1 (M100 bp) | 2 | 3 | 4 | 5 (M50 bp) | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|
| DNA/HMf | 1:0 | 1:1 | 1:1 | | 1:0 | 1:2 | 1:2 | 1:2 |
| U *MNase*I | / | 0.005 | 0.005 | | 0.005 | 0.005 | 0.005 | 0.005 |
| Digestion time | / | 2 min | 5 min | | 2 min | 2 min | 4 min | 6 min |

**Fig. 4.6** Electrophoresis on 2.5% agarose gel. Genomic DNA fragmentation with different DNA/HMf ratio, 1:1 (lane 3 and 4) and 1:2 (lane 6-9) and different digestion time. In lane 2 genomic *E. coli* DNA not treated with *MNase*I enzyme and in line 6 genomic *E. coli* DNA treated with *MNase*I in absence of HMf. In line 1 and 5 molecular weight marker respectively 100 bp and 50 bp. In the table below, the characteristics of each different digestion tests are indicate. In lane 7 (yellow marked) were indicate the best condition for genomic DNA fragmentation. Based on these features, about 5 µg of genomic DNA was fragmented and used for the library construction.

As expected, incubating genomic DNA with 0.005 U of *MNase*I without HMf, after only 2 minutes, it was completely degraded. Increasing the HMf amount (from 1:1 to 1:2 ratio) the DNA was protected from nuclease, and the length

periodicity of the obtained fragments (60 and 120 bp) can be observed on agarose gel (see figure 4.6 lane 7 and 8).

Genomic DNA fragments were end-repaired and blunt-ended before the ligation with the Eco60LPE linkers, containing the *Eco*RI restriction site, necessary for the cloning in pEP3 vector. When this linkers form a dimer, an *Asc*I restriction site was recreated as shown in figure 4.7.

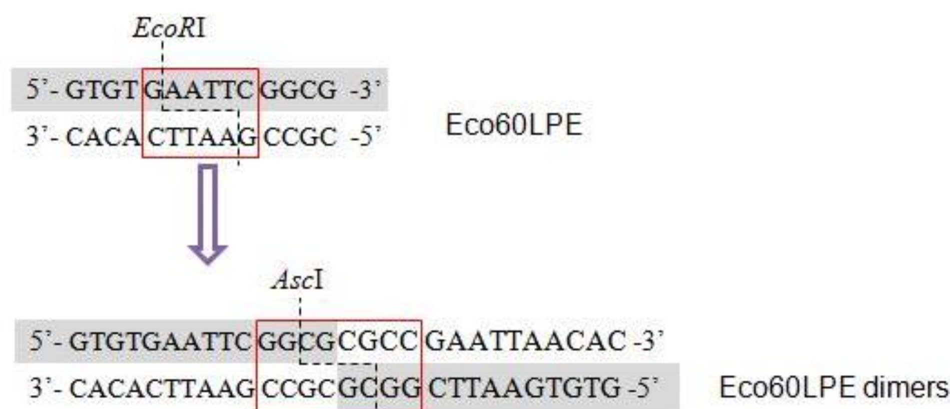



**Fig. 4.7** Representation of Eco60LPE linker with *Eco*RI restriction site, and the dimers formation with the creation of *Asc*I restriction site.

After the linkers ligation, to remove dimers of not ligated linkers and to create specific sticky ends for cloning, a simultaneous digestion with *Eco*RI and *Asc*I was performed. A restrict range of DNA fragment , from 60 to 300 bp, was purified from agarose gel and then ligated into pEP3 vector cut with *Eco*RI. The ligation products were transformed into electrocompetent DH5α One Shot cells. The transformed cells were selected for “in frame” fragments on solid media with Ampicillin 12 µg/mL and Chloramphenicol 25 µg/mL, and as control of transformation, on Chloramphenicol 25 µg/mL. In the tab. 4.3 the library dimension on both Chloramphenicol-Ampicillin and Chloramphenicol only, and the Chloramphenicol-Ampicillin/ Chloramphenicol ratio were indicated. This information gives an indication of how many clones “in frame” are present respect to those cloned.

| Library | N° independent clones | | Chl-Amp/ Chl ratio |
|---|---|---|---|
| | Chl-Amp | Chl | |
| DNA Genomic from *E. coli* | $8x10^4$ | $3.5x10^5$ | 1/4 |

**Tab. 4.2** The table shows the numbers of colonies obtained after transformation of electrocompetent *E. coli* DH5αONE SHOT cells with the ligation of genomic fragments in pEP3. Also the ratio “in frame”/ “out of frame” clones was indicated.

### 4.4.1 Genomic DNA library characterization and sequencing

The genomic DNA library was checked for the presence of fragments with the correct length. 96 randomly picked clones were checked by PCR to evaluate the presence and size of inserts. The used primers anneal external to the pEP3 cloning site (see Materials and Methods). An example of this analysis is shown in figure 4.8. To assess the clones diversity, the presence of ORFs and to evaluate the total representation of the original genomic DNA, the 96 clones were sequenced .

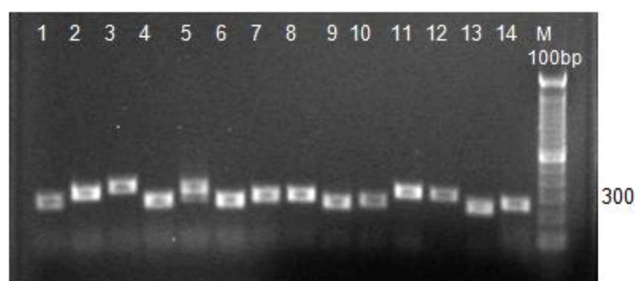
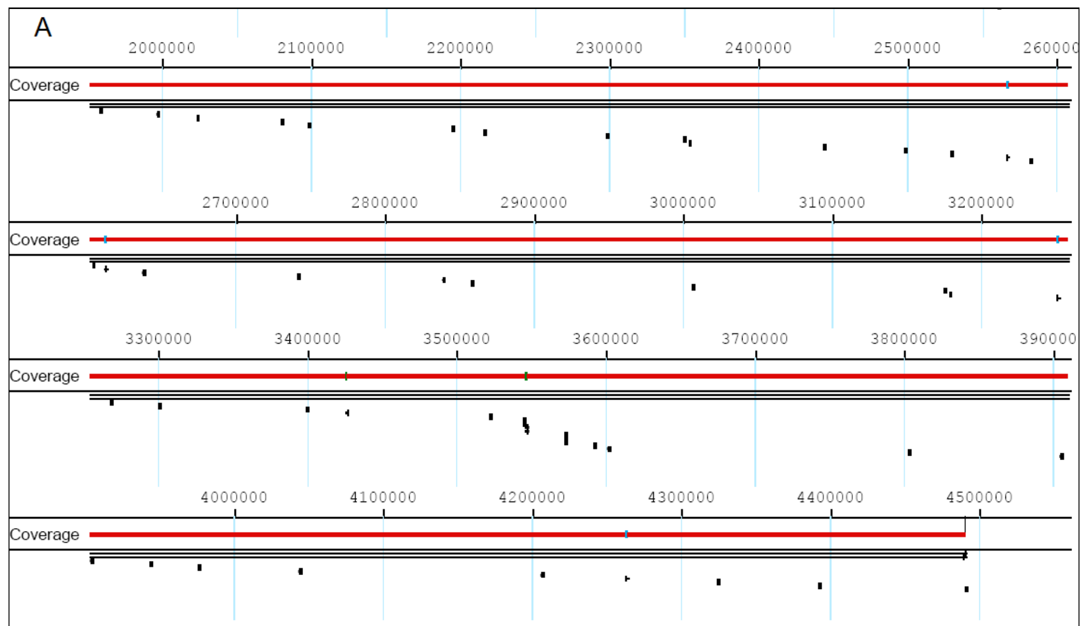



**Fig. 4.8** Electrophoresis on 1.5% agarose gel of PCR from 14 randomly picked clones of ORF selected genomic library of *E. coli* (lane 1-14). All the inserts presented a length between 200 bp and 300 bp, as expected. Lane 15 molecular weight 100 bp ladder.

Sequences were compared to *E. coli* K12 (NC_010473) genome. From sequences analysis emerged that among the 93 correctly sequenced inserts, all were mapped genes and, of this, 81 corresponded to CDS (that correspond to 87% of mapped genes). In figure 4.9 A the DNA sequences distribution was shown.



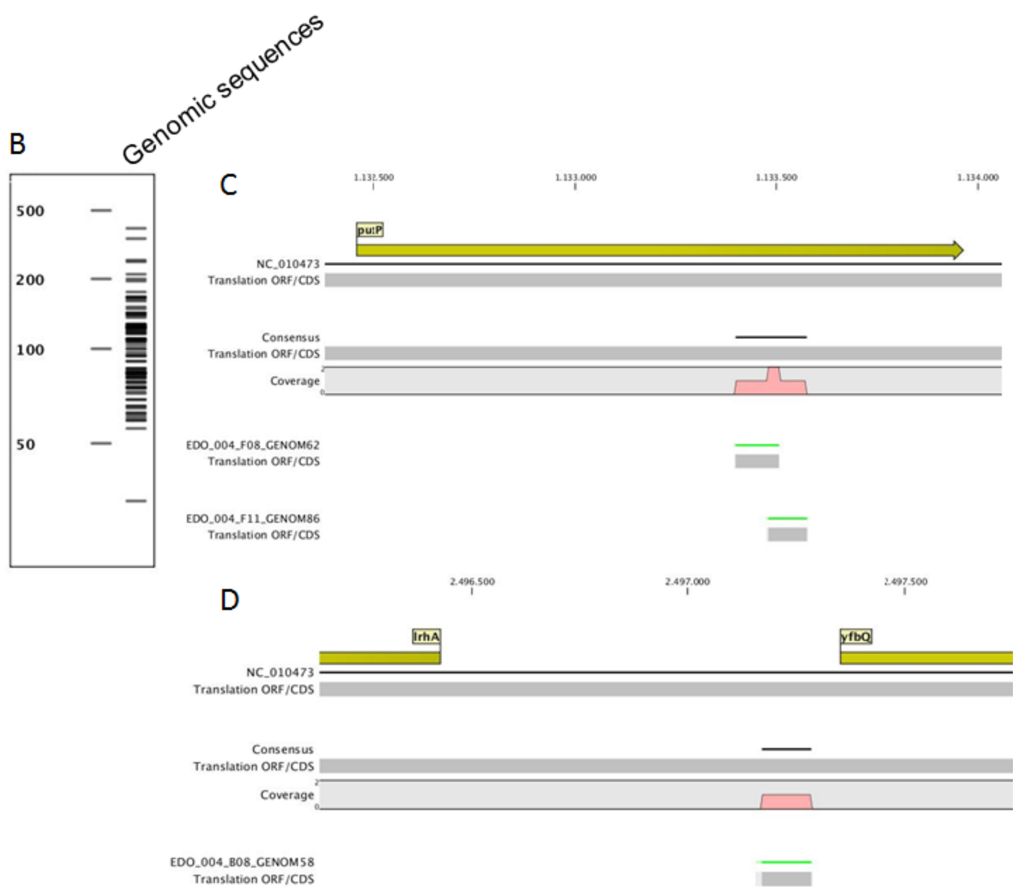




**Fig. 4.9** In A DNA sequences distribution. In B the schematic representation of the fragments length, in C and D examples of sequences matched in CDS (yellow) and not CDS regions, respectively.

## 4.5 Human Brain phage display library

A cDNA phage display library form Human Brain could be a useful tool to perform analysis of sera and CSFs of patients with autoimmune diseases of Central Nervous System (CNS), in fact such library potentially contains all the autoantigens of CNS. Through phage display technologies this library could be selected to identify the epitopes involved in the autoimmune response of a specific autoimmune disease of the CNS, and the eventually isolated autoantigens could be used to discriminate between different autoimmune diseases of CNS. One well recognized obstacle in the construction of expression cDNA libraries is the differential presence of the transcripts that reflects their level of expression in the target tissue. For this reason we decided to introduce a normalization step in the library construction protocol in order to equalize the relative abundance of each mRNA.

### First strategy

Initially we designed a protocol for the construction of ORF- enriched, oriented and normalized phage-display library from human brain. It was developed to achieve the following objectives: (I) cDNA fragmentation to obtain the total epitome representation of the target tissue, (II) Oriented cloning of the cDNA and (III) cDNA normalization.

cDNA fragmentation was introduced to display, on the phage surface, polypeptides of define length corresponding  to possible epitopes. In this protocol we decided to maintain the cDNA orientation because previous experience (Bembich 2004) showed that the ORFs enrichment is more efficient. Further, oriented cloning reduces the presence of clones with stops codons and “frame shift” events, that can be present in phage display (Baranov et al. 2001). The vector pEP3 contains sites for directional cloning and, like pEP2, is capable of ORF selection. The steps of the original protocol for the library construction are summarized in the figure 4.10.

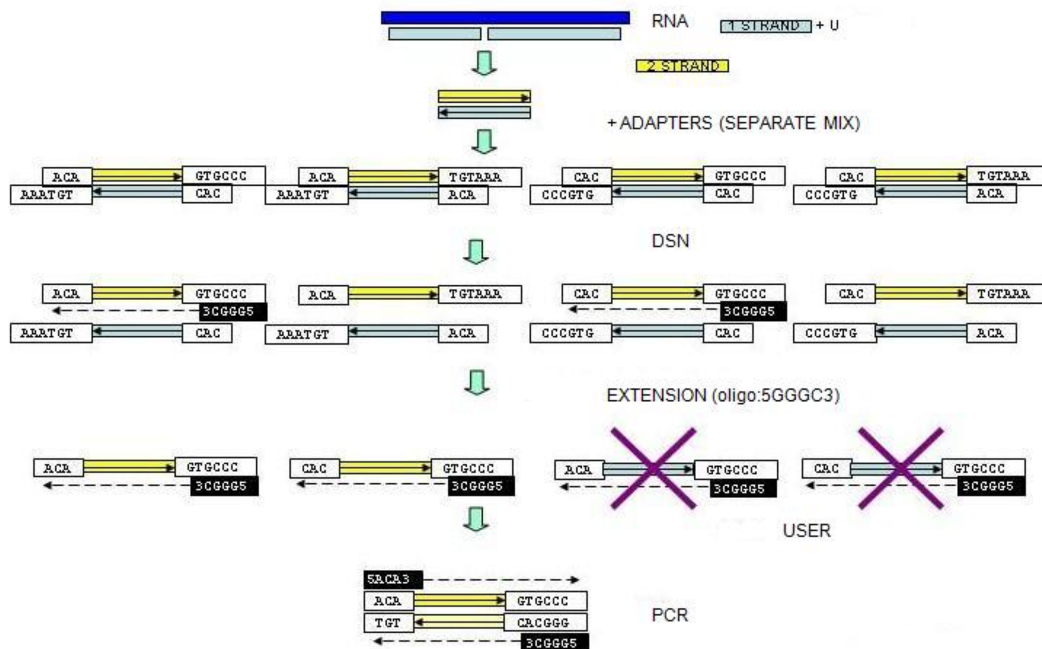



**Fig. 4.10** Schematic protocol for the construction of ORF- enriched, oriented and normalized phage-display library from human brain. The adapters and oligo sequences, shown in the figure, are merely indicative.

The first strand was obtained by reverse transcription in presence of a dNTPs mixture containing dUTP instead of dTTP, in order to mark this filament selectively. A cDNA fragmentation step (not shown on the figure) using HMf protection was also provided. The fragmentation step was performed immediately after the second strand synthesis, in order to have a sufficient cDNA amount to complete the entire protocol, and also because the adapters should be added after the fragmentation. Fragmentation, in fact, makes impossible to maintain the right cDNA orientation. After fragmentation, peculiar adapters (A1 and A2) were added. They contain the recognition site for *Eco*RI and *Hind*III enzyme that allow oriented cloning in pEP2 and pEP3 vectors.

Adapter 1:

ctaaacaaacgaccgatacgaattcagc LINORMFOR53

gagagatttgtttgctggctatgcttaagtcg P LINORMFOR35P

Adapter 2:

P ggcaagcttacgtcgcgaccctcccatatat LINORMREV53P

ccgttcgaatgcagcgctgggagggta LINORMREV35

The adapters are able to legate cDNA fragments in four different way: A1-A2, A1-A1, A2-A2, A2-A1, but the useful combination is only one: i.e. the adapters 2 to 3' of the cDNA.

After adapters ligation, the cDNA fragments were normalized using the Duplex Specific Nuclease enzyme (DSN). Subsequent phases of the protocol were performed to discriminate the oriented vs non oriented fragments by extension of only fragments showing the correct orientation and the right combination of adapters.

In the subsequent extension phase, only the cDNA fragments showing the adapter 2 ligated to 3' of the filament were extended. This discrimination was possible using specific primers, able to anneal only to this sequence. At the end of the extension, four type of double strand remain, two of which were dUTPs labeled. The final distinction of the correct oriented fragments was provided by the *USER* enzyme. This enzyme can recognize the presence of uracil in DNA and cuts in that sites. In this way the dUTP-labeled cDNA strand, that are not oriented, were degraded. The last step is the amplification with specific primers, able to amplify only the cDNA fragments bearing the correct adapters combination for the oriented cloning.

Some attempts has been made to generate the ORF- enriched, oriented and normalized phage-display cDNA library from mRNA of Human Brain with sufficient numbers of clones to represent all the potential autoantigens of the CNS, but the results were no satisfactory. It was emerged that the main obstacle in the library construction was the starting mRNA amount; in fact high mRNA concentration was required to follow the entire protocol. Starting with 4 µg of mRNA polyA$^{+}$ was not sufficient to obtain, in the final step, a cDNA amount for creating library with high numbers of clones. For this reasons it was decided to start simplifying the protocol: an amplification step after the mRNA retro-transcription was introduced. Should be noted that this amplification step can give a sufficient amount of cDNA to follow the entire protocol but, at the same time, made impossible to maintain the proper cDNA orientation. Among all purposes of the protocol (orientation, ORF enrichment

and fragmentation) we thought that the loss of orientation can be the most tolerable.

## Second strategy

The modified protocol was developed to construct an ORF- enriched and normalized phage display library from Human Brain. The steps of the modified protocol are summarize in the following figure.

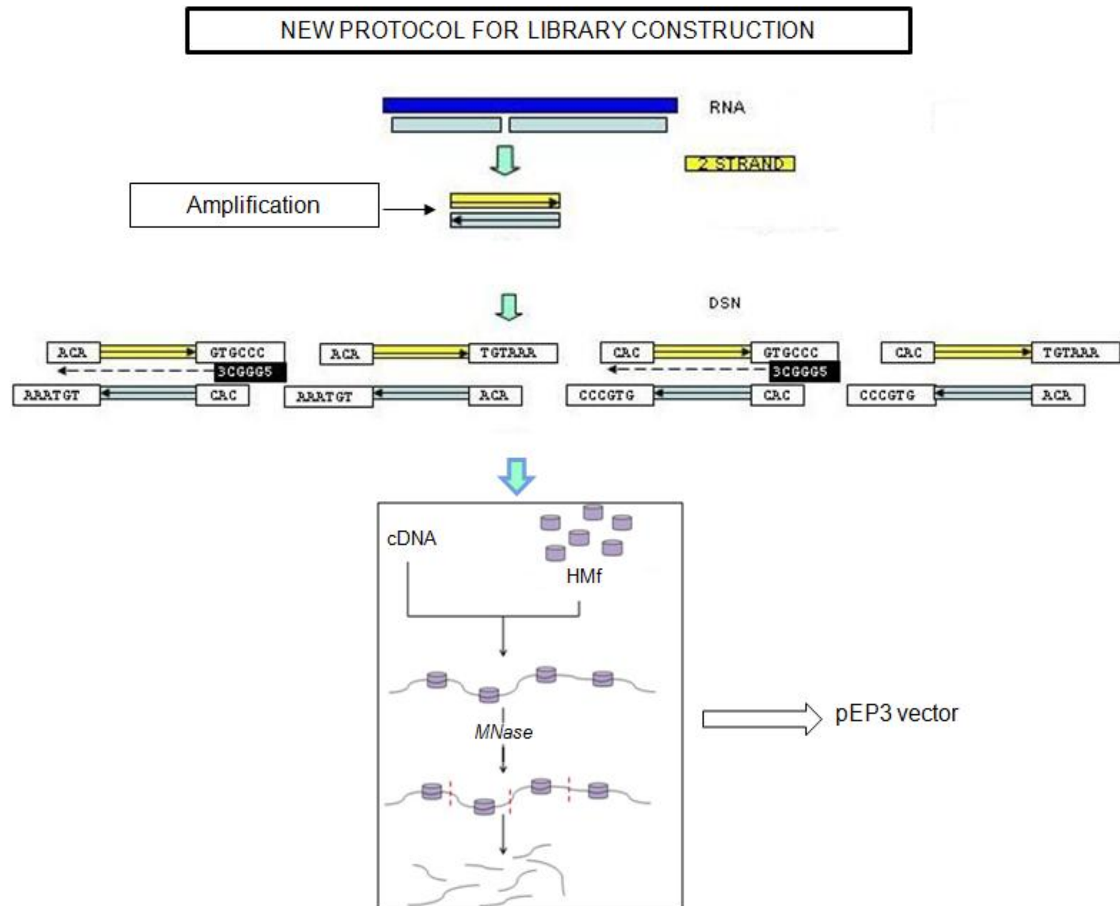



**Fig. 4.11** Modified protocol for the construction of ORF- enriched and normalized phage-display library from human brain. Shortly, mRNA was retro-transcribed and amplified following a mix between SMARTer$^{TM}$ PCR cDNA synthesis (*Clontech*) and MINT cDNA synthesis (*Evrogen*) protocols. The amplified cDNA was normalized with DSN and then fragmented with *MNase*I, in presence of HMf. The obtained cDNA fragments with a proper length were added with specific adapters and cloned in pEP3 vector.

## 4.5.1 cDNA preparation and normalization

The first difficult for the production of a cDNA library from human brain was the recovery of a good quality mRNA preparations from autoptic sample. When polyA$^+$ RNA from HB has been available from *Clontech*, it was

purchased and employed for the cDNA synthesis (Bembich 2004). The mRNA source was defined, by manufacturer, as a pool of whole forebrains from eight healthy caucasian males obtained immediately after sudden death. A second question was related to retro-transcription of the mRNA. In order to obtain, as planned, a large cDNA quantity, for cDNA synthesis and amplification a protocol, that is a mix between SMARTer$^{TM}$ PCR cDNA synthesis (*Clontech*) and MINT cDNA synthesis (*Evrogen*) protocols, was followed (as described in Materials and Methods). 1 µg of mRNA polyA$^{+}$ was a sufficient amount of cDNA to follow the entire protocol. Briefly, as suggested the two different protocols, the first strand cDNA synthesis was performed using the proper mRNA primers indicate in MINT protocol, and for the second strand synthesis was used the first strand as template, in presence of Advantage 2 polymerase mix. Once obtained the amplified cDNA from human brain, a normalization step has been performed.

## - Test on DSN activity

Duplex Specific Nuclease (DSN) is the enzyme used for the cDNA normalization (see Introduction). It has a high preference for the digestion of double strand DNA and DNA-RNA hybrids respect to single strand DNA. DSN normalization process involves (I) the double strand DNA denaturation, (II) the subsequent renaturation under specific conditions, and then (III) the enzymatic degradation of the DNA fraction composed by abundant transcript. This protocol allows to equalize the abundance of different expressed transcripts, and thus increase the presence of rare transcripts.

The amplified cDNA form human brain has been used. As suggest by the DSN protocol (*Evrogen*), 1200 ng of double strand cDNA human brain, in presence of specific buffer, were divided into four different sterile PCR tubes. After double strand denaturation for two minutes at 98 °C, the DNA, was incubated for 5 hours and 30 minutes at 68 °C. During this period, the single strand cDNAs reassociated following the kinetics of hybridization of double strand nucleic acid: the most abundant cDNA pair up quicker than less abundant. Several experiment, previously preformed, indicated that the

hybridization time of 5 hours and 30 minutes was the best condition to obtain normalized sequences from the used cDNA. After this hybridization, to three reaction tubes the DSN enzyme (undiluted, 1:2 and 1:3) and to one tube (the control) only the buffer was added. All reaction were incubate at 68 °C for 25 minutes, and then the DSN was inactivate.

In order to assess the DSN enzyme activity, a normalization test using a gene coding for an abundant transcript, the Glucose 6-phosphate dehydrogenase (GAPDH), have been carried out. The treated (undiluted, 1:2 and 1:3) and untreated (the control) samples were amplified using specific primers for GAPDH. The results of this test were shown in the figure 4.12.

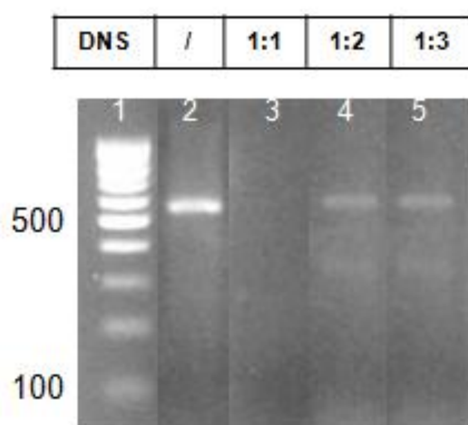



**Fig. 4.12** Electrophoresis on 1.5% agarose gel of PCR amplified GAPDH using as template the DSN treated and untreated samples. GAPDH was amplified starting form human brain cDNA treated with undiluted DSN enzyme (lane 3), 1:2 DSN dilution (lane 4) 1:3 DSN dilution (lane 5) and untreated (lane2). Lane 1 molecular weight 100 bp ladder.

As expected, the DSN untreated sample (lane 2) presents higher quantity of GAPDH products than the DSN treated samples, as can be noted from the band intensity of control sample. In addition, a different amount of amplification product for normalized sample can be correlated to the enzyme dilutions. Using the DSN enzyme undiluted, no amplification occurred (lane 3), while using 1:2 or 1:3 dilution a similar amplification was detectable (lane 4 and 5 respectively). This data indicate that a dilution 1:3 was sufficient to achieve a intensity reduction of the band, compared to the not normalized control, and that the normalization process was successful.

## - Normalized library construction

The efficiency of the DSN enzyme normalization system was further investigated generating a cDNA library from human brain in the commercial vector pBlueScript. To test only the protocol step that involves the use of DSN for the library normalization, the cDNA was not fragmented in presence of HMf.

As indicated by the DSN protocol (*Evrogen*), the normalized cDNA sample were re-amplified. The appropriate cycles number for the DSN treated sample amplification derived from the amplification analysis of the control (that was submitted to denaturation and renaturation processes but was not normalized). To determine the optimal amplification cycles number, the DSN untreated sample was PCR amplified and the amplified cDNA, at different amplification cycles (starting from cycle 7), were analyzed. The result was shown in the figure 4.13.

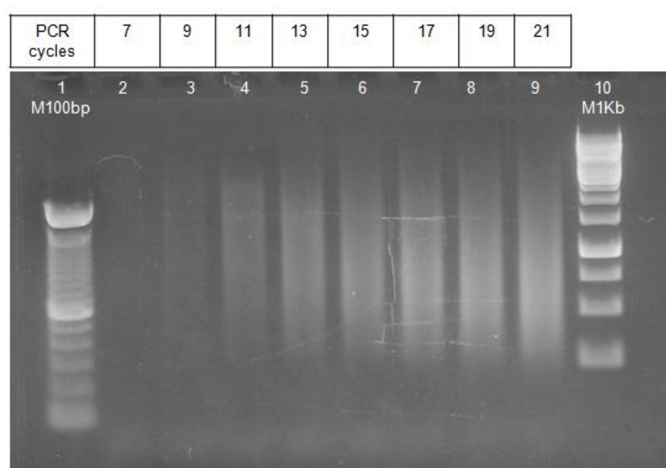



**Fig. 4.13** Electrophoresis on 1.5% agarose gel of PCR amplified DSN untreated sample. In lane 2-9 were shown control sample amplified at different amplification cycles, indicate in the table upstream. Lane 1 molecular weight 100 bp ladder, lane 10 molecular weight 1 Kb ladder.

The optimal amplification cycle number is the one after that the amplification is stable; in fact when the yield of PCR products stops increasing with every additional cycle, the reaction has reached its plateau. As can be seen in fig. 4.13, the optimal cycle number was found to be 13. After calculating the optimal cycles number (called number X by the manufacturer) for the control, it was calculate the amplification number for the DSN treated samples.

According to DSN protocol, it was calculate the number N, where N=X-7 (number of starting amplification cycles); in this case N is 13-7=6. The number of additional cycles was obtained by summing the number N to the number of cycles equal to 9, that in this case correspond to 15. The total number of cycles for the amplification of DSN treated sample was so calculated: number of additional cycles summed to number of starting cycles, in the specific case is 22.

The DSN (diluted 1:3) treated sample was amplified with a 22 amplification cycles. For the construction of the cDNA normalized library from human brain, the amplified sample was digested with *Rsa*I to partially fragmented cDNA to increase the cloning efficiency and to obtain blunted and for the specific linkers ligation. After the ligation with Eco60LPE linkers, the same used for the genomic DNA library construction, all the cDNAs were digested with specific restriction enzymes for the generation of sticky ends for the cloning (*Eco*RI) and for the deletion of linkers dimers (*Asc*I), and purified from agarose gel to eliminate the unligated linkers from the sample. The cDNAs ranging from 200 and 800 bp were ligated into *Eco*RI pBlueScript arms using a vector: cDNA ratio of about 1:10.

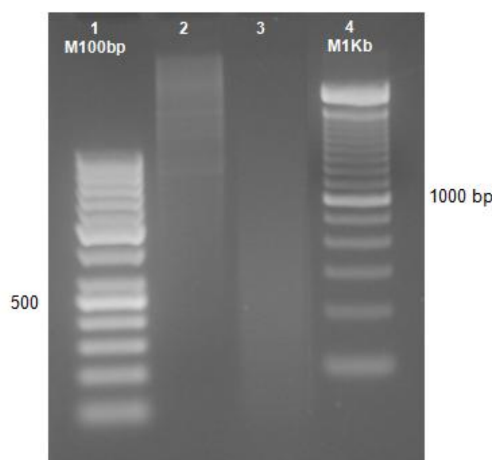



**Fig. 4.14** Electrophoresis on 1.5% agarose gel of PCR amplified DSN (1:3) treated sample (lane 2) and the same sample digested with *Rsa*I enzyme. Lane 1 molecular weight 100 bp ladder, lane 4 molecular weight 1 Kb ladder.

The ligation mixture was purified and used to transform chemical-competent DH5α One Shot *E. coli* cells. Being the purpose of the test only of indicative nature, i.e. to verify the correct normalization, chemical-competent cells were

chosen. After transformation, using 200 ng of pBlueScript vector, a normalized library with a total number of $1.3x10^3$ clones was obtained. First, it was checked for the presence of fragments. The size of the cloned inserts, assessed by PCR using the couple of primers external to the cloning site (see Materials and Methods), were performed on 96 randomly picked clones grown on Ampicillin agar plates. All the 96 analyzed clones gave a PCR products (see figure 4.15) and the insertion size were, as expected, between 200 bp and 800 bp, with a prevalence between 300 bp and 600 bp. To evaluate the normalization process, the 96 analyzed clones were sequenced.

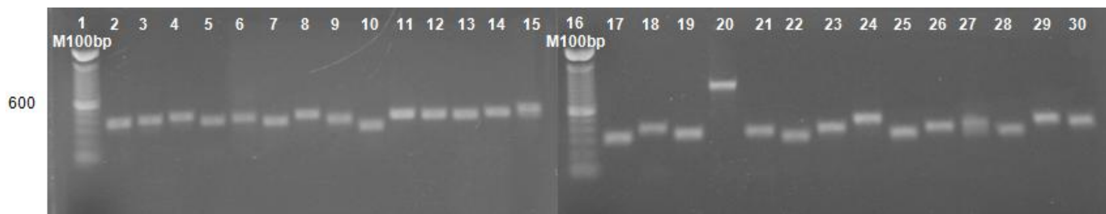



**Fig. 4.15** Electrophoresis on 1.5% agarose gel of PCR products relative to 28 randomly picked cones of the normalized human brain library. Lane 1 and 16 molecular weight 100 bp ladder.

The sequences were analyzed at www.ngs-trex.org site and the output results indicated that the normalization process performed with DSN enzyme can be considered successful, in fact, among the 72 mapped genes, 65 (that correspond to 90% of mapped genes) were unique (figure 4.16). Thanks to this positive results, it was decided to use this normalized cDNA for the next step of the library construction: the fragmentation.

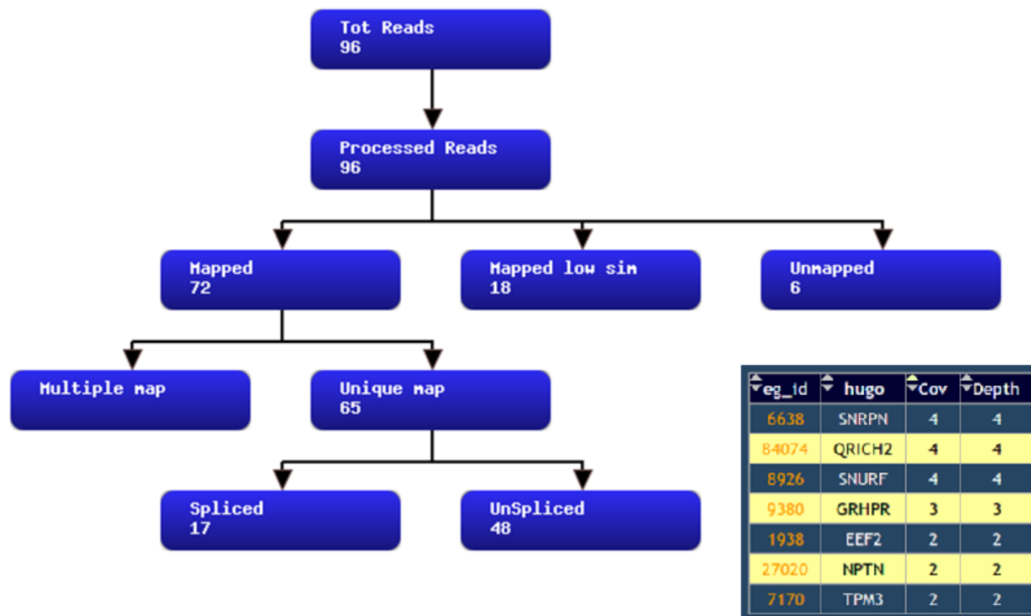



**Fig. 4.16** Output of the program used for the normalized sequences analysis.

### 4.5.2 Normalized cDNA fragmentation and HB library construction

The normalized cDNA was submitted to *MNase*I fragmentation in presence of histones dimers. The reaction was performed as for the genomic DNA fragmentation, changing specific parameters. Normalized cDNA was incubated with archeobacterial histones dimer in 1:2 ratio in presence of specific buffer (see Materials and Methods); after the *MNase*I enzyme addition, the digestion was carried out for 5 minutes. This parameters were chosen after some attempts to obtain the best conditions for the fragmentation.

The purified fragmented cDNA was checked on TBE 10% (figure 4.17) gel and end-repaired before the ligation with the Eco60LPE linkers, containing the *Eco*RI restriction site, as said before, necessary for the cloning in pEP3 vector. Thank to the specific Eco60LPE characteristics, to remove dimers of not legated linkers, and to create specific sticky end for cloning, a digestion with *Eco*RI and *Asc*I was performed. Two range of DNA fragments were purified from agarose gel (see figure 4.17), from 60 bp to 300 bp, and to 300 bp to 600 bp because it was decided to obtain two different libraries, one with

“low size fragments” encoding probably mostly for linear epitopes, the other with “high size fragments” encoding also for conformational epitopes.

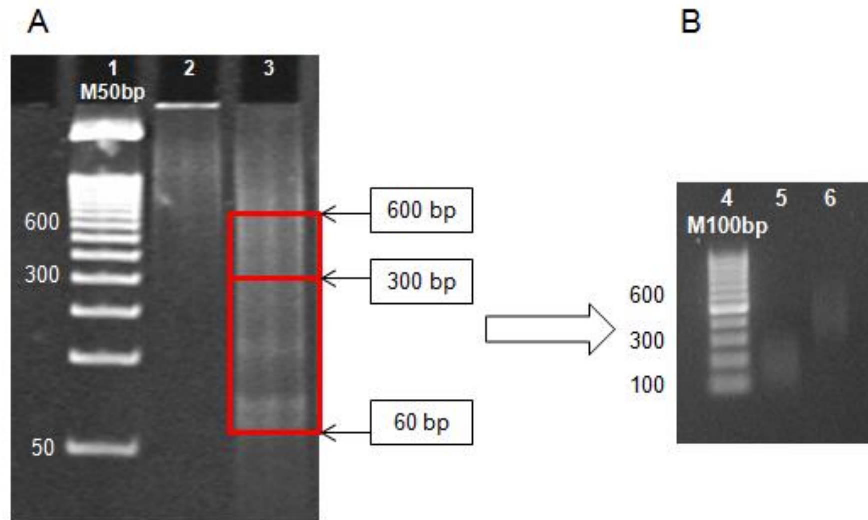



**Fig. 4.17** In A electrophoresis on 10% TBE gel. In lane 2 the normalized cDNA from human brain, in lane 3 the normalized cDNA fragmented with *MNase*I in presence of histones dimers. Red boxes indicate the two different cDNA fragments population extracted from gel (recovered range between 60 bp-300 bp and 300 bp-600 bp). In B electrophoresis on 2% agarose gel. In lane 5 the gel purified cDNA fragments from 60 bp to 300 bp, in lane 6 the gel purified cDNA fragments from 300 bp to 600 bp. Lane 1 molecular weight 50 bp ladder and lane 4 molecular weight 100 bp ladder.

The purified fragments were, separately, ligated into pEP3 vector cut with *Eco*RI, in a reaction with a molar ratio vector/DNA of 1:10. The purified ligation products were transformed into electrocompetent DH5α One Shot cells. This step was performed to obtain ORF selected fragments. The transformed cells were selected for “in frame” fragments on solid media with Ampicillin 12 μg/mL and Chloramphenicol 25 μg/mL, and as control of transformation, on Chloramphenicol 25 μg/mL only. In the table 4.3 the obtained library dimension, in the best performed test, on both Chloramphenicol-Ampicillin and Chloramphenicol only, and the Chloramphenicol-Ampicillin/ Chloramphenicol ratio were indicate.

| Library | N° independent clones | | Chl-Amp/ Chl ratio |
|---|---|---|---|
| | Chl-Amp | Chl | |
| HB 60-300 bp | $5.6x10^5$ | $6.4x10^6$ | 1/11 |
| HB 300-600 bp | $7.8x10^4$ | $5.7x10^5$ | 1/7 |

**Tab. 4.3** The table shows the numbers of colonies obtained after transformation of electrocompetent *E. coli* DH5αONE SHOT cells with the ligation of cDNA fragments in pEP3. Also the ratio “in frame”/ “out of frame” clones was indicated.

The experimental ratio of “in frame” / “out of frame” clones was around 1/11 for the HB 60-300 bp library and 1/7 for the HB 300-600 bp, instead of the expected 1/18.

The two human brain cDNA libraries were checked for the presence of fragments with the correct length. Some clones, randomly picked from agar plate with Ampicillin 12 µg/mL and Chloramphenicol 25 µg/mL were checked by PCR to evaluate the presence and sizes of inserts, by PCR with specific primers (pEPback and pEPseq). As we can observed in figure 4.18, all the clones, randomly picked, presented an insert with correct size.

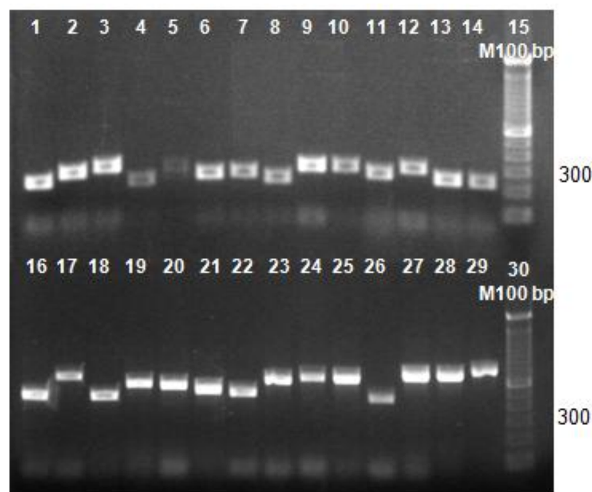



**Fig. 4.18** Electrophoresis on 1.5% agarose gel of PCR from 14 randomly picked clones of ORF selected HB 60-300 bp library (lane 1-14), and from 14 randomly picked clones of ORF selected HB 300-600 bp library (lane 16-29). All the inserts presented a proper length (between 60 bp and 300 bp for the first library, and between 300 bp and 600 bp for the second library). Lane 15 and 30 molecular weight 100 bp ladder.

### 4.5.3 Human Brain libraries subcloning in pDAN5 vector

Once obtained the ORF selected fragments, they were collected and used for the construction of a phage display library that could be used for the selection with specific target. The HB library 300-600 bp was digested with two specific restriction enzyme (*Bss*HII-*Not*I), to recovery all the ORF fragments and clone them directly in pDAN5 phagmidic vector (see Materials and Methods). This vector has the following characteristics: (I) the constitutive Ampicillin resistance, (II) the pLacZ promoter upstream the polylinker, (III) a "leader" sequence just downstream the promoter for the periplasmic localization of the expression products, (IV) the gene III encoding for the coat pIII of the phage particle, (V) a SV5 tag necessary for the identification of the expression products using commercial anti SV5 antibodies and (VI) amber codon between SV5 tag and gene III for the production of protein in soluble form.

The digestion products were separated by electrophoresis agarose gel and the ORF fragments were recovered. The HB 300-600 bp fragments were ligated in pDAN5 vector cut *Bss*HII-*Not*I in a reaction with a molar ratio vector/DNA of 1:10. The purified ligation products were transformed into electrocompetent DH5αF' cells. Finally, a phage display library of ORF enriched and normalized cDNA fragments from Human Brain with a total number of $1.9 \times 10^9$ clones was obtained. The library was first checked for the presence of fragments of proper length, and then it was harvested and stored in aliquots of 50 µL at -80 °C. The size of clones inserts, assessed by PCR using the couple of primers external to the cloning site (pELBS and pGENE3), were performed on randomly picked clones.

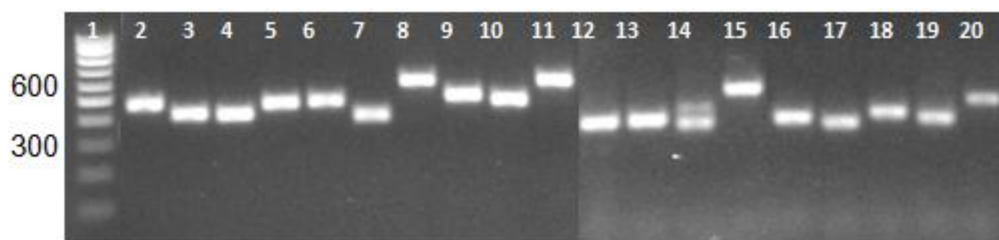



**Fig. 4.19** Electrophoresis on 1.5% agarose gel of PCR from 19 randomly picked clones of ORF selected HB 300-600 bp phage display library (lane 2-20). All the inserts presented a proper length (between 300 bp and 600 bp). Lane 1 molecular weight 100 bp ladder.

17 clones, randomly picked, were sequenced and in the table 4.4 an illustrarive analysis is shown.

| Clone | Feature ID | Identity | Chromosome | Exon |
|---|---|---|---|---|
| E03 | NMNAT2 | Nicotinamide nucleotide adenylyltransferase 2 | 1 | 11 |
| G07 | RGS7 | Regulator of G-protein signaling 7 | 1 | |
| E04 | ZNF238 | Zinc finger protein 238 | 1 | 2 |
| G10 | SVIL | Supervillin | 10 | |
| C05 | ENO2 | Enolase 2 (gamma, neuronal) | 12 | 11 |
| E09 | NDRG2 | NMYC downstrean-regulated gene 2 | 14 | 15 |
| E07 | RPL4 | Ribosomal protein L4 | 15 | 10 |
| H11 | RBFOX1 | RNA binding protein, fox1 homolog | 16 | 15 |
| B11 | LUC7L3 | RNA binding protein, fox1 homolog | 17 | |
| G12 | BABAM1 | BRISC and BRCA1-A complex member | 19 | 9 |
| D11 | SLC8A2 | Solute carrier family 8 | 19 | 10 |
| C03 | CRKL | Cyclin dependent kinase 6 | 22 | |
| G09 | RREB1 | Ras responsive element binding protein 1 | 6 | 13 |
| B01 | ST7 | ST7 antisense RNA 2 | 7 | |
| F05 | FDFT1 | Farnesyl-diphosphate farnesyl transferase 1 | 8 | 7 |
| B06 | NFIB | Nuclear factor I/B | 9 | 9 |
| E11 | ENSG00000238047 | MT-NDI pseudogene | X | |

**Tab. 4.4** Results of the HB sequences analysis.

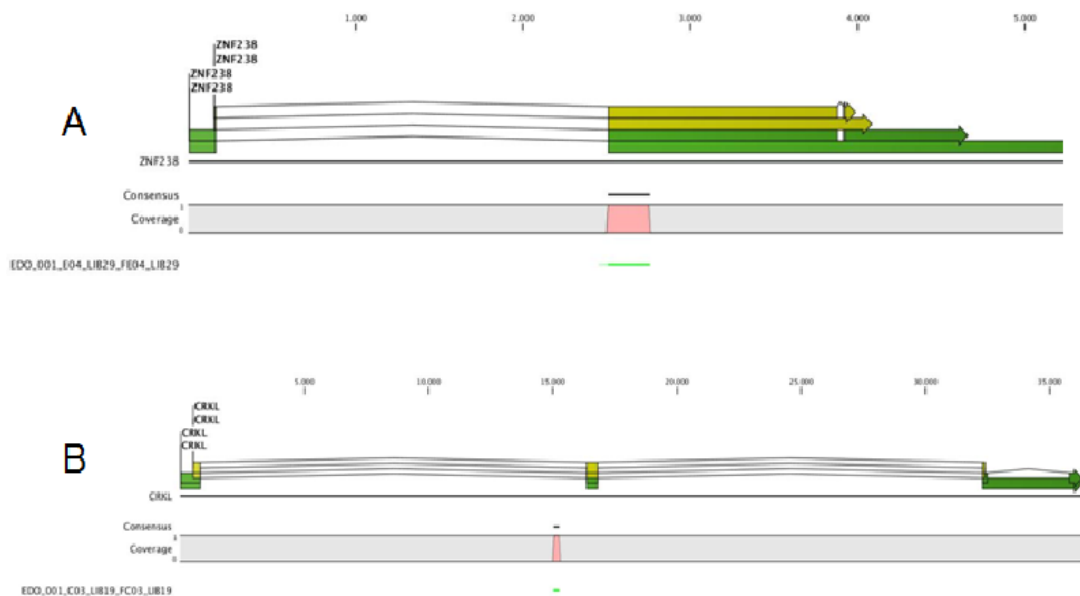



**Fig. 4.20** In A and B, examples of Human Brain sequences matched in CDS (yellow) and not CDS regions, respectively.

The human brain library with 300-600 bp ORF fragments was then used to discover novel potential antigens involved in autoimmune diseases of Central

Nervous System. To this aim, samples from Multiple Sclerosis patients were used. Although the cause of Multiple Sclerosis is still uncertain, many findings point toward an ongoing autoimmune pathogenesis, and in many cases the auto-antibodies production was demonstrated (Hartung and Kieseier 2010).

## Patients' data

Sera and Cerebrospinal fluid (CSF) samples used in this study were classified in the table 4.4. A total number of 136 Multiple Sclerosis sera ("MS Trieste-group 1,2,3") and 28 CSFs ("MS Trieste-group 1"), 36 Other Neurological Diseases (OND) sera with related CSFs ("OND Trieste") and 20 Healthy donors (HD) sera were kindly provided by Department of Clinical Medicine and Neurology**,** University of Trieste**.** A group of 171 MS sera ("MS Cagliari-group 1,2,3") and 129 CSFs ("MS Cagliari- group 1,3") and 50 OND sera with the related CSFs ("OND Cagliari") were kindly provided by the MS Centre of the University of Cagliari. A total number of 22 OND sera ("OND Padova") were kindly provided by the Clinical Neurological of the University of Padova. The characteristics of all MS and OND patients are summarized in the following tables.

| MS samples | CSF | N | Female/Male | Mean Age (range) in years | Diagnosis | Mean disease duration (range) in years | EDSS range |
|---|---|---|---|---|---|---|---|
| MS Trieste-group 1 | Yes | 28 | 18/ 12 | 42.7 (21-66) | 20 RR<br>7 CIS<br>1 PP | RR: 9.3 (3-28)<br>CIS: 5 (4-7)<br>PP:27 | 0-3.5 (CIS and RR) 8 (PP) |
| MS Trieste-group 2 | No | 90 | 71/ 19 | 40.4 (19-63) | All RR | 12 (1-36) | 0-6.5 |
| MS Trieste-group 3 | No | 18 | 7/ 11 | 48.7 (35-67) | All SP | 18.9 (6-35) | 1.5-7 |
| MS Cagliari-group 1 | Yes | 30 | 19/ 11 | 36.6 (22-54) | / | 5.8 (1-20) | / |
| MS Cagliari-group 2 | No | 15 | 9/ 6 | 45 (27-62) | 13 RR<br>1 CIS<br>1 SP | / | / |
| MS Cagliari-group 3 | Yes (99) | 126 | 82/ 44 | / | / | / | / |

| OND samples | CSF | N | Female/Male | Mean Age (range) in years | Diagnosis |
|---|---|---|---|---|---|
| OND Trieste | Yes | 36 | 12/ 24 | 62.8 (22-89) | NIND and OIND |
| OND Padova | No | 22 | 17/ 5 | / | 10 LES<br>12 Chronic polyradiculoneuritis |
| OND Cagliari-group 3 | Yes | 50 | 27/23 | / | / |

| Healthy samples | CSF | N | Female/Male | Mean Age (range) in years |
|---|---|---|---|---|
| Healthy Trieste-group 3 | No | 20 | 11/ 9 | 37.6 (27-63) |
| Healthy Cagliari-group 2 | No | 13 | 3/ 10 | / |

**Tab. 4.5** Classification of MS, OND and HD samples.

## 4.6 Selections of Human Brain library

The phage display library of ORF enriched and normalized cDNA fragments from human brain was selected using purified Immunoglobulin IgG form CSF of Multiple Sclerosis patients, with the purpose of isolating more reactive clones to auto-antibodies.

All the selections were performed following these specific step:

- An immunotube, for each panning, was coated overnight with 500 µL of purified IgGs from MS CSF pool at a concentration of 10 µg/mL in PBS 1X;
- Approximately $10^{10}$-$10^{12}$ phages resulting from cDNA library clones were added and incubated for the interaction with the IgG pool;
- Extensive washing steps to remove aspecific phages were performed, changing the washes numbers and the tween concentration for each panning (see Materials and Methods);
- Specific phages, bound to antibodies after the washing step, were eluted with DH5α *E. coli* cells and amplified through further panning (two in total) or plated in case of the last round of selection.
- A phage ELISA was performed using the clones resulting from the last selection round to discover the more reactive clones which were

submitted to secondary phage ELISA, sequencing analysis and databank screening to establish their identity.

### 4.6.1 Pooling of MS CSF samples

The pool of MS CSF samples were collected on the basis of a preliminary ELISA test in which the 96 CSF form “MS Cagliari-group3” were analysed for the presence of IgGs using an anti human IgG antibody. The result was shown in the figure 4.21.

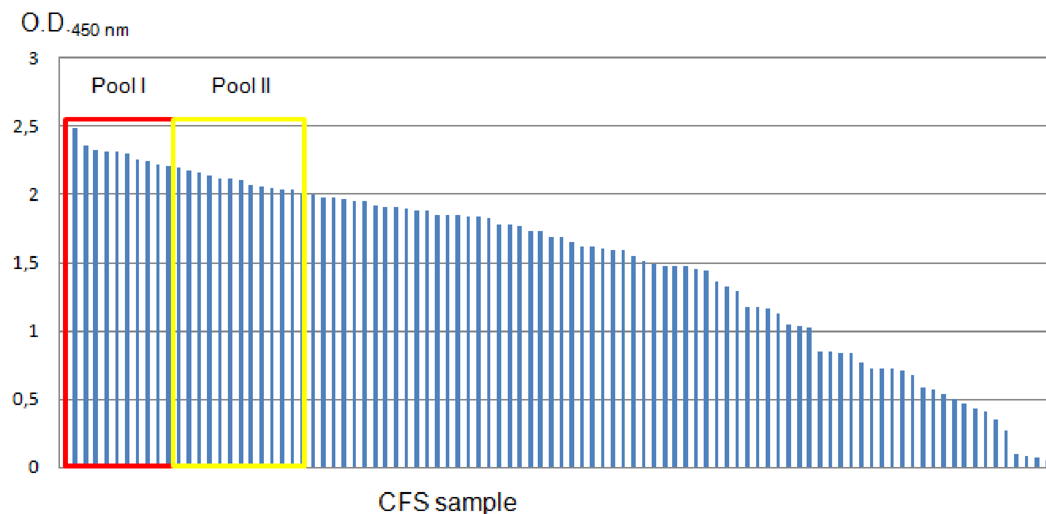



**Fig. 4.21** Schematic histogram representing the optical density (O.D.$_{450\ nm}$) values of ELISA preliminary screening conducted on 96 CSF samples. Red box indicates the 11 most reactive sample O.D.$_{450\ nm}$ >2.2. Yellow box indicates the 14 CSF samples with 2> O.D.$_{450\ nm}$ >2.2.

The 25 more reactive CFSs, showing a O.D.$_{450\ nm}$ >2 were divided into two populations based on optical density. Pool I included MS CSF samples with O.D.$_{450\ nm}$ >2.2 in ELISA assay (8 woman and 3 man); Pool II included MS CSF samples with O.D.$_{450\ nm}$ between 2 and 2.2 in ELISA assay (8 woman and 6 men). These CSF were collected and pooled 20 µL from each sample and affinity purified by protein G chromatography using *Millipore* kit. The eluted purified IgGs were checked on Coomassie gel for the presence of heavy (HC) and light (LC) IgGs chains, as we can observed in figure 4.22.

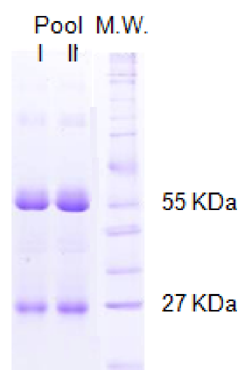



**Fig. 4.22** SDS/ PAGE 10% gel electrophoresis of the purified IgGs of the two pools. The heavy and light chains were localized around 55 and 27 KDa, respectively.

## 4.7 HB selection against purified IgGs from CSF

The HB library was subjected to three distinct selection:

1) Against purified IgGs from pool I. 11 CSF Multiple Sclerosis samples with O.D.$_{450nm}$>2.2 in ELISA assay (8 woman and 3 men);

2) Against purified IgGs from pool II. 14 CSF Multiple Sclerosis samples with 2>O.D.$_{450nm}$>2.2 in ELISA assay (8 woman and 6 men);

3) Against purified IgGs from pool I and II.

The titer of input phages, output phages and the library enrichment were summarized in the table 4.6.

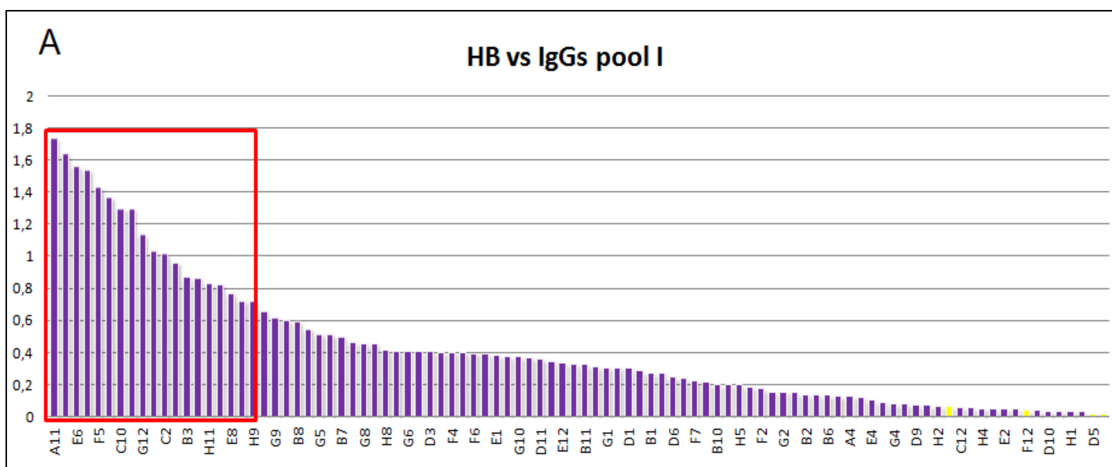
| | I Round | | II Round | | |
|---|---|---|---|---|---|
| | Input phages | Output phages | Input phages | Output phages | Library enrichment |
| Pool I | $6x10^{10}$ | $5.1x10^{5}$ | $1.1x10^{11}$ | $1x10^{7}$ | 10 |
| Pool II | $1x10^{9}$ | $8.4x10^{5}$ | $9x10^{10}$ | $7.3x10^{6}$ | 9 |
| Pool I+II | $7x10^{10}$ | $6.6x10^{5}$ | $9x10^{10}$ | $1.3x10^{7}$ | 1.2 |

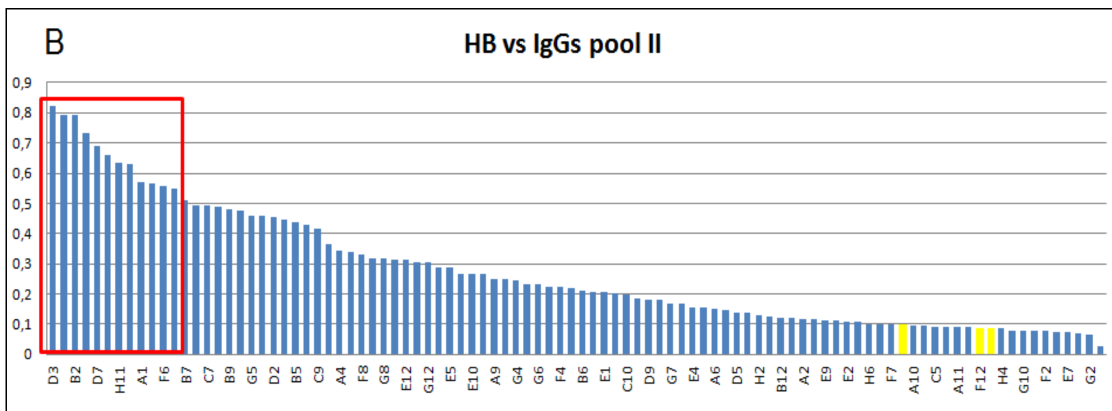
**Tab. 4.6** Results of the HB library selections. In the table the number of input phages, output phages and the library enrichment were indicated.

A final enrichment of about 10-fold, 9-fold and 1.2-fold for the presence of specific clone for the three selection respectively was obtained.

92 clones from the last rounds of each selection were randomly picked and individually subjected to phage-ELISA. For this assays, a monoclonal

antibody anti human IgGs- Fc specific was used as coating in order to better display the pooled immunoglobulins for the following interaction with the phage library clones. Considering the obtained O.D. values, the clones with an O.D. values > 0.7 for the first and third selections and an O.D. values > 0.54 for the second selection were considered for further analysis. In total 19 clones were considered for the first selection, 12 for the second selection and 14 for the third selection. A phage ELISA experiment was replicated on these 45 clones and the data was confirmed. The insert presence for all clones was verified by PCR amplification (data not shown).



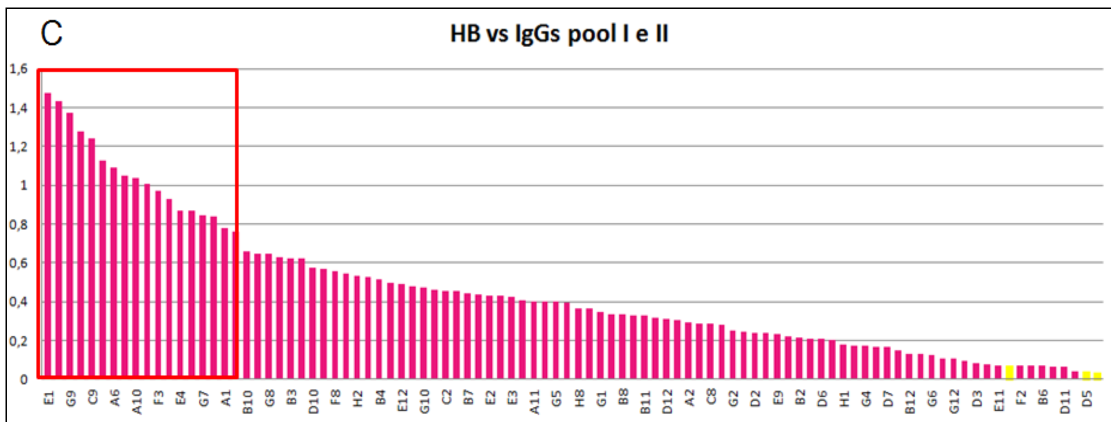



**Fig. 4.23** Phage-ELISA values read at $A_{450\ nm}$ relative to 92 (for each phage ELISA) individual HB clones selected against CSF pool I (A), pool II (B) and pool I e II (C). In the red boxes were indicate the positive clones. In yellow are indicated the negative controls, performed using helper phage.

The three clones that showed the higher optical density in both phage ELISA were chosen for a secondary phage ELISA test on single serum sample. All these clones (E6, F5 and C10) derived from the first selection (pool I).

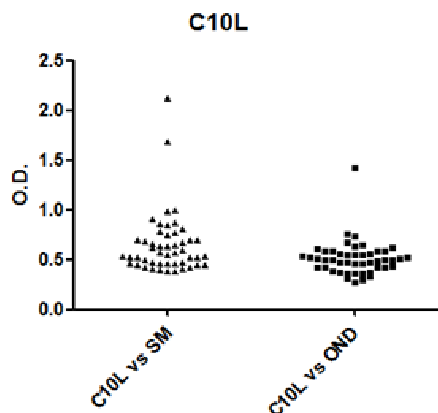
## 4.8 Secondary ELISA on MS selected antigens

In secondary ELISA test, the more reactive clones were tested, as phages, to assess their potentiality in the discrimination between Multiple Sclerosis patients and patients with Other Neurological Diseases. Briefly, phage-ELISA were performed using the three more interesting clones. They were tested using the optimal conditions previously set: $10^{10}$ as phage titer, PBS1X - Tween 0.1% as wash stringency and 1:1000 sera as coating. All washing steps were performed using the hydro speed washer (*TECAN*) to standardize the condition and increase the reproducibility. The interaction was detected with a peroxidase conjugated-secondary antibody against the phage envelope (anti M13- HRP). Each O.D. value obtained for the interaction between single MS serum and HB clone was compared with the O.D. value obtained for the interaction between single OND serum and HB clone. Interestingly, the discrimination between the two populations of patients was statistically significant ($p > 0.05$) considering all the three antigens. The data were analyzed using the *GraphPad Prism version 5.0* program and one illustrative result is shown in the figure 4.20. For each antigen using 49 MS sera and 47

OND sera, it was possible to identify a discrimination, statistically significant, between the two populations. Setting a cut-off given by the mean + 2SD (Standard Deviation) of the OND population, it is possible to calculate the presumed “sensitivity” and “specificity”. Among the MS patients, 4/49 (8% = sensitivity) had elevated reactivity for the “E6” phage clone, while 46/47 (98% = specificity) of the OND individuals resulted under the cut-off; the value considering “F5” were: 4/49 (8% = sensitivity) and 46/47 (98% = specificity); while for “C10” 5/49 MS patients had elevated reactivity (10% = sensitivity) and 46/47 OND individuals resulted under the cut-off (98% = specificity).

Moreover, additional diagnostic parameters were evaluated: the positive predictive value (PPV) and the negative predictive value (NPV), both useful in predicting the likelihood of someone with a positive result having the disease. Therefore, PPV represents the proportion of patients with positive test result who have the disease, whereas NPV represents the proportion of patients with negative test result who do not have the disease. Another parameter considered is the “Likelihood ratio” (LR); it indicates the likelihood that a given test result would be expected in a patient with the target disorder compared to the likelihood that same result would be expected in a patient without the target disorder. High LRs (>10) are usually considered to be of diagnostic value, while low LRs (<0.1) can rule out the disease. LRs around 1 indicate no useful information in the diagnosis prediction.

An example of the obtained data is summarized in the figure 4.24.



| Statistical parameters calculated for C10L | | | |
|---|---|---|---|
| Parameter | Definition | Calculation | Values |
| Sensitivity | True positives/ True positives+ False negatives | 5/49=0.1 | 10% |
| Specificity | True negatives/ True negatives+ False positives | 46/47=0.98 | 98% |
| PPV | True positives/ True+ False positives | 5/(5+1)=0.83 | 83% |
| NPV | True negatives/ True+ False negatives | 46/(46+44)=0.5 | 50% |
| LR+ | Sensitivity/ (1- Specificity) | 0.1/(1-0.98)=5 | 5 |
| LR- | (1-Sensitivity)/ Specificity | (1-0.1)/0.98=0.91 | 0.91 |
| Cut-off | mean + 2 SD of the OND population | 0.5126+2(0.1740) | 0.86 |

**Fig. 4.24** Summary of the analyzed phage ELISA data with *GraphPad Prism version 5.0* program using the most reactive clones (C10L). In the table are shown the statistical values obtained.

A statistically significant difference was established between MS patients and OND for their reactivity against all the three encoding phage clones, but considering LRs these clones didn’t have a significant diagnostic value.

The sequence analysis showed that some of post-selections clones are identical reducing their number to a total of 13, but some sequences find high homology clone with apparently not coding or not annotated genome regions. So a Blastx analysis has been performed and the results are summarized in the table 4.6.

| Clone | Code-N Blast | Identity – Blastx | Lenght nt | Homology | aa/ tot | CDS (a.a.) | Frame | Frameshift | Cellular component | Phage ELISA |
|---|---|---|---|---|---|---|---|---|---|---|
| F5C | BAG64005.1 | unnamed protein product [Homo sapiens] | 228 | 100% | 74/76 | 320-393 | IF | | Cytoplasm | ++ |
| G11L | | | 240 | 100% | 77/80 | | | | | ++ |
| F11L | NP_055591.2 | ARF GTPase-activating protein GIT2 isoform 3 | 329 | 100% | 70/109 | 430-499 | OOF | 1 | Nucleus | ++ |
| D3G | | | | | | | | | | ++ |
| H11L | NP_002013.1 | dimethylaniline monooxygenase [N-oxide-forming] 4 | 125 | 100% | 39/41 | 472-510 | OOF | 1 | Cytoplasm/Mitochondrion | ++ |
| B12L | NP_001193826.1 | retinal dehydrogenase 2 isoform 4 | 293 | 65% | 62/98 | 305-399 | IF | | Nucleus | ++ |
| H9L | AAM00184.2 | MCEF protein | 122 | 63% | 19/40 | 3-33 | OOF | 1 | Nucleus | ++ |
| G9G | Q96L62.1 | Full=HERV-H_1q41 provirus ancestral Env polyprotein | 199 | 54% | 20/66 | 228-264 | IF | | / | ++ |
| A12L | AAH29366.1 | ZFP106 protein [Homo sapiens] | 122 | 53% | 18/40 | 152-168 | OOF | 1 | / | +++ |
| C2L | AAI31600.1 | ras-responsive element-binding protein 1 | 122 | 53% | 18/41 | 170-202 | OOF | 1 | Nuclear speck/Cytoplasm/Nucleus | +++ |
| H12G | | | | | | | | | | + |
| E6C | | | | | | | | | | +++ |
| A10C | | | | | | | | | | +++ |
| C9C | | | | | | | | | | +++ |
| B5C | | | | | | | | | | ++ |
| D7G | | | 218 | 53% | 18/72 | | | | | + |
| G7C | | | | | | | | | | ++ |
| G9C | NP115599.1 | Pecanex-like protein 3 | 323 | 51% | 20/107 | 253-291 | OOF | 1 | Multi-pass membrane protein | +++ |
| D8L | NP_001185968.1 | serine incorporator 2 isoform 5 | 250 | 50% | 28/83 | 76-126 | IF | | Integral to membrane | +++ |
| A1G | NP_002324.2 | low-density lipoprotein receptor-related protein 3 precursor | 208 | 50% | 16/69 | 73-126 | IF | | Single-pass type I membrane protein | + |
| B3L | NP_006030.2 | C-type mannose receptor 2 precursor | 145 | 49% | 24/49 | 598-643 | IF | | Single-pass type I membrane protein | +++ |
| F5L | | | 185 | 49% | 24/62 | | | | | +++ |
| B1C | | | | | | | | | | +++ |
| E6L | | | | | | | | | | +++ |
| A11L | | | | | | | | | | +++ |
| A6C | | | | | | | | | | +++ |
| E1C | | | | | | | | | | +++ |
| C10L | | | 296 | 39% | 31/99 | 598-672 | | | | +++ |
| A8L | CAI16861.1 | AT-Hook transcription factor | 346 | 41% | 22/115 | 397-461 | IF | | Nucleus | ++ |

**Tab. 4.7** Results of the post selection sequences analysis. In the last column: + phage-ELISA (O.D.<0.7); ++ phage-ELISA (0.7<O.D.>1); +++ phage-ELISA (O.D.>1).

With the knowledge of the human and other genomes and, more recently, with the advent of next-generation sequencing technologies, the construction of cDNA libraries and developing of the concerning strategies for a mere sequencing purpose or gene discovery have become less important. On the other hand, the new high-throughput screening and the so called omics technologies have increased the demand of biotechnological tools, like combinatorial libraries, in many research fields, as drug discovery and biomarker detection (Glokler et al. 2010).

Among this tool, the display technology have been largely improved and successfully employed in affinity peptides or proteins identification and searching (Bradbury et al. 2011; Lofblom 2011; Ullman et al. 2011; Beghetto and Gargano 2011
). To characterize antigen-antibody affinity in epitope targeting and mapping, several methods including recombinant techniques have been developed to increase the diversity of phage display libraries (Pande et al. 2010).

The principal aim of my PhD was the setting of a protocol for the creation of phage libraries to display cDNA fragments encoding real ORF sequences, that could correspond to potential epitopes. A similar phage display library contains all the potential ORF repertoire of a cell or tissue, with single ORF represented by all the fragments in which has been randomly split. This tool can be specially used in the study of autoimmune diseases to perform different kind of analysis, such as the identification of epitopes involved in pathological reaction, the comparison between healthy and pathological conditions, or between different pathological conditions (Puccetti and Lunardi 2010).

For this purpose the phage display technology provides several advantages over conventional immunoscreening of plasmid or lambda-phage cDNA libraries. First, screening the cDNA library in a fluid phase, the interaction of antibody with conformational epitopes is possible because the denaturation of

the proteins displayed on the surface of the phage is avoided. In addition, repeated rounds of selection allow the specific enrichment of phage particles that interact with the antibodies; in this way, the possibility of detection antigens that are present at low abundance can be increased. Moreover, this chance is also increased by screening as many as $10^{10}$–$10^{11}$ individual clones in a single experiment.

Two kind of phage display libraries have been used in epitope targeting: the random peptide libraries and the gene fragments libraries. In the second ones protein domains can be displayed on phage particles and directly identified.

Respect to conventional approach, in which complete cDNAs are cloned, libraries of gene fragments increase the chance to express stable and functionally folded domains (Prodromou et al. 2007). This combinatorial approach require two fundamental step: a controlled fragmentation of starting DNA and a selection of ORF fragments.

In fact, cDNA fragment libraries suffer from the problem that the vast majority of clones are non functional, since only one clone in eighteen, if starting with DNA encoding an open reading frame, will be correctly in frame. This problem can be overcome using strategies that confer to “in frame” clones a selectable characteristic (Li and Caberoy 2009). Further, the coupling of screening ORF enriched libraries and massive sequencing of selected clones allow the identification and structural determination of functional genic ORFs (D'Angelo et al. 2011).

It has been demonstrated that the β-lactamase can act as a selectable folding reporter allowing to filter out ORFs (D'Angelo et al. 2011). So, to construct epitope phage display libraries, we used this system developing a vector obtained from the backbone of pPAO2 described by Zacchi and colleagues (2003). In pPAO2 the ß-lactamase is located between the DNA insert and the gene III of M13, allows the initial selection of “in frame” cDNA fragments and must be removed prior to obtain the final library. To overcame some pPAO2 limitations such as LIC cloning system involves a T7 DNA polymerase treatment, two derivative vectors, pEP1 and pEP2, were previously created

(Bembich 2004; Bembich 2008). They allowed to construct different phage display mini-libraries from single genes (Bembich 2004; Cortini 2005) useful for epitope mapping, and also phage display library from human brain (Bembich 2008) and phage display library from MS CIS patients B-cells (Cortini 2008) both useful for antigen discovery. The limitation in the use of this vectors is the low cloning efficiency, that reduces the possibility to obtain highly representative phage display library. To improve the cloning efficiency, we constructed, starting from pEP2 backbone, a new vector, named pEP3, in which the gene III was deleted. In this way ORFs selection and phage production were uncoupled. The idea was that the β-lactamase alone, fused to the DNA insert, could be more properly folded and improve the ORF selection. Further, the following biopanning could have be performed in an appropriate vector, as pDAN5 (Sblattero and Bradbury 2000).

Using pEP3, the cloning efficiency was increased about ten-fold in comparison with pEP2; this data could be explained with the smallest pEP3 size that gives a better capability of being transformed. Further, it can be hypothesized that a basal expression of fragments in fusion with gene III, that occurs in pEP2 system, could be more toxic for the clones.

When a population of ORF fragments was cloned, only 1/18 are expected to be in frame because only 1/3 start correctly, only 1/3 finish in a correct way and only 1/2 is cloned with the correct orientation. Using pEP2 in a comparison with pEP3, we obtained exactly a 1/9 ratio because in pEP2 the cDNA fragment were oriented cloning. The ratio of “in frame”/ “out of frame” clones obtained with different libraries constructed using pEP3, ranges from 1/3 to 1/11. To filter ORF fragments with pEP3 we used a low Ampicillin concentration (12 µg/mL). It is possible that a higher ampicillin concentration could be required in order to obtain a better selection. It is in some extent confirmed by the analysis of the clone sequences after biopanning, in which the percentage of out of frame clones was higher than that obtained with pEP2 (Bembich, 2008). It should be not exclude that same frame shift events allowed the “out of frame” clones to survive. It is known that transcriptional readtrough can occur in bacteria producing functional phages particles even

in presence of frame-shifted sequences (Carcamo et al. 1998). Finally, in some cases, the use of suppressive bacteria strain DH5α *E. coli* can allow the translation of the amber "stop" codon (TAG) as glutamine, leading to a positive selection of "out of frame" clones with TAG codons.

Summing up, pEP3 meets the purpose for which it was built, the need to increase the library size, improving the cloning efficiency. Regarding ORFs selection based on Ampicillin media survival, the efficiency is comparable to previously used vectors. We decided to use pEP3 vector for the human brain library construction because we thought that an high cloning efficiency could be determinant to obtain high representative libraries.

As stated above, the other important step in the construction of a phage display cDNA libraries, to be used in biopanning of samples of patients with CNS autoimmune diseases, is the controlled fragmentation of the starting cDNA.

One of the major issue to consider is the cDNA fragments size to clone. Epitope phage display libraries can be employed to identify both linear and conformational epitopes (Mackay and Rowley 2004). This aspect can be fundamental in some pathologies; for example, in Multiple Sclerosis conformational epitopes are recognized by pathogenic autoantibodies (Zhou et al. 2006). In this case the use of a cDNA phage display library is more appropriate than a random peptide library. Until now peptide-based research has been important in attempts to identify autoantigens in MS (Alcaro and Papini 2006) and random peptide libraries have been almost always used. Only recently the use of a phage display library derived from MS brain plaques for a serological Ag selection was reported (Somers et al. 2008), but only one potential antigen was identified.

It seemed appropriate to predetermine an insert size that allows also the selection of conformational epitopes. For this purpose the size of the cDNA fragments must be precise.

Several fragmentation methods to obtain random fragments were developed (Prodromou et al. 2007); these can be divided into two groups: the physical

and the enzymatic methods. For both systems some limitations occurred: the fragments dimension can not go below the 300 bp, the DNA quality is compromised (particularly for physical methods), and the enzymatic reactions are difficult to control. So we decided to adopt a system previously set up in my hosting laboratory (Azzoni and colleagues 2007) and based on the digestion with *MNase*I and protection with archeal histones from Methanothermus fervidus (HMf). In the original protocol histones monomer was used to protect DNA from the complete digestion. During my PhD we think to produce a recombinant form of a covalent dimer of histone, in order to make the DNA protection more reproducible, because the formation of dimer in vitro is transient (Grayling et al. 1997).

Our hypothesis was confirmed by the comparison between the use of the monomeric and the dimeric form: the digestion reactions are more controlled, and this affect also the yield of DNA after the post-reaction recovery. Further, varying the DNA:histones ratio it is possible to modulate the length of DNA fragments to be obtained. As expected, the dimer protects modules of DNA of 60 bp and multiples; below this length we cannot detect DNA fragments while the size of DNA fragments show peaks around 60, 120, 180 bp. The fragmentation can be also modulated changing the digestion time. In this way it is possible to generate libraries of predetermined length.

The procedure of DNA fragmentation and ORF selection was initially tested on the total genomic DNA of *E. coli*. The mapping of 93 randomly chosen sequences showed that 87% correspond to ORFs demonstrating the capability of the system of select efficiently “in frame” clones.

The ORF genome representation obtained with a dimension of $8x10^4$ ORF clones can be estimated considering that the ORF sequences in *E. coli* genome amount for approximately to $4x10^6$ bp (4290 ORFs of 951 nt of medium length) (Blattner et al. 1997) and that the fragments have a medium length of 120 bp: with a 2,35X coverage there is the 90% of probability to include all sequences.

Another improvement in the protocol for the construction of a phage display cDNA libraries, to be used in biopanning of samples of patients with CNS autoimmune diseases, is the introduction of a library normalization step.

It is well known that one important obstacle in the construction of expression cDNA libraries is the differential presence of the transcripts that reflects their level of expression in the target tissue. A normalization step must be introduced in order to obtain, ideally, equal representation of each transcript.

To produce an ORF- enriched, oriented and normalized phage display library from human brain cDNA fragments a sophisticate strategy was conceived. The attempts to follow the entire protocol demonstrated that high mRNA polyA$^+$ amount was required; in fact, 4 µg of mRNA polyA$^+$ were considered not sufficient to obtain representative libraries.

To overcome this obstacle, we decided to amplified the starting cDNA, so postponing the aim of library orientation, thinking that, among all the purposes of the protocol, this step is the less essential.

The mRNA poly$^+$ was retro-transcribed and amplified, and then submitted to a normalization process using DSN enzyme. To check the normalization protocol, a normalized library in pBluescript was constructed and 96 clones, randomly chosen, were sequenced. The bioinformatic analysis indicated that the normalization process performed with DSN enzyme can be considered successful, in fact, among the 72 mapped genes, 65 (that correspond to 90% of mapped genes) were unique.

The normalized cDNA was subjected to fragmentation with *MNase*I thanks to the protection with histones dimer. The reaction conditions were performed to obtained cDNA fragments with useful size both to encode linear and conformational epitopes. In order to distinguish between different epitopes we cloned separately two fragments population, one from 60 bp to 300 bp and the other from 300 bp to 600 bp, obtaining two ORF- enriched and normalized cDNA libraries from human brain. Probably conformational epitopes can be more abundant in the “high size fragments” library. This library, with fragments between 300 bp and 600 bp, was sub-cloned into the phagemid pDAN5.

Considering that the ORF enriched library dimension, before the subcloning, was comparable to that of a previously constructed, but not normalized, library (Bembich, 2008), it can be concluded that the procedure produces library of sufficient size for the next biopanning experiments. Considering that in a human cell are expressed from 10.000 to 15.000 transcripts (Jongeneel et al. 2003) and that the medium length of the coding part is 1186 bp (the MGC Project Team 2004), an ORF enriched cDNA library of $7.8x10^4$ clones is sufficient to provide a 2.5-fold clone coverage of the transcripts present corresponding to a 92% of probability to include a particular sequence.

A first biopanning with this library was performed using purified immunoglobulins from a pool of cerebrospinal fluid samples of Multiple Sclerosis patients. In fact, the aim of this thesis was the developing of a protocol to generate epitope libraries suitable for fishing novel antigens of autoimmune diseases like Multiple Sclerosis.

Although the debate in Multiple Sclerosis study is still open, many findings point toward an ongoing autoimmune pathogenesis. The study of autoimmune response in Multiple Sclerosis is complicated by epitope spreading, its heterogeneity and pathogenetic relevance of the autoantibodies. The targets of the autoimmune response in MS are believed to be cellular components of the CNS. The immune response presents both a cellular and a humoral component, but especially in the last decade, several data have demonstrated a strong implication of B cells in the development of the disease (Oh et al. 2008; Racke 2008). It is well known that in MS there is an unexplained increase of IgGs in the CSF from 15% to 30%, visualized as oligoclonal bands after electrophoresis (Gilden et al. 2001). Several studies on the antigen-binding regions of antibodies found in MS brain demyelinating plaques and cerebrospinal fluid revealed that the presence of oligoclonal bands can not be considered a nonspecific bystander response (Fujimori et al. 2011), because of data consistent with a specific antigen-targeted processes and the clonal expansion of B lymphocytes populations.

Despite a large number of studies on the humoral response in MS, that suggest the involvement of an infectious agent, the disease etiology is still

unknown (Correale et al. 2008; Brennan et al. 2011). The most accepted hypothesis about the mechanism of triggering MS is the molecular mimicry. This phenomena leads to the activation of the autoreactive immune cells because of cross reactivity between foreign agents, such as viral epitopes, and self antigens, for example during a virus infection. During the course of the disease, the antibody profile was complicated by the epitope spreading, a process whereby epitopes distinct from and non-cross-reactive with an inducing epitope become major targets of an ongoing immune response. This phenomenon has been defined in experimental and natural situations as a consequence of acute or persistent infection and secondary to chronic tissue destruction that occurs during progressive autoimmune disease (Vanderlugt et al. 1998). The hypothetical mechanism of epitope spreading, proposed to explain the progression of CNS autoimmunity observed in MS, is based on the damaging of self-tissue, by immune mediated mechanisms or apoptosis, which leads to the release of self-antigens then presented to self-reactive T cells with other self-specificities (Croxford et al. 2002). This mechanisms are able to create a “fertile field” for the development of inflammatory autoimmune disease. So, the specificity of antibodies in MS can vary between patients and in the time in the same patient. For this motif, increasing the list of candidate autoantigens of MS might implement the possibility to distinguish between different patients and/or different states of the diseases and to monitor the therapeutic effects. Since it is recognized that MS is a patient-specific disease, the definition of a panel of biomarkers for the autoantibody profiling is really informative.

The idea of having an immunological profile, or signature, of multiple antigens from each single patient is progressively emerging. This can be obtained using innovative approach as protein microarray. The final goal of this project is the construction of a diagnostic protein microarray using all the potential autoantigens discovered during previously selections (Bembich 2008, Cortini 2008) and those that will be performed using the new generated libraries.

We started with these biopanning of the constructed phage display library from Human Brain performing three different selections with three partially

overlapped pools of IgGs from CSF of MS patients. The first results are encouraging. Of 92 clones tested for each selection, about 50% showed a high recognition (O.D.> 0.5) in Phage ELISA assays. Among the clones that exhibited the greatest recognition, some have been tested in secondary Phage ELISA. They showed high specificity (98%) compared with a low sensitivity (8%). Except that the assay using the phage as antigen is not optimal for establishing diagnostic parameters, it must be considered that the data of the low sensitivity is not unexpected considering that the MS is a patient-specific disease. The further sequences analysis revealed that these clones are identical; totally they are 8 clones on the 29 sequenced. It could be suggestive, but can also mean that the phage clone displays a particularly sticky but aspecific peptide. In fact the Blastx analysis reveals that only the 49% of the clone shows sequence homology with the C-type mannose receptor 2 precursor.

The sequence analysis also shows that other clones are identical and that 5 on 13 seems to be out of frame. This could be due to a non-specific binding of phage particles, but it should be noted that all these clones have only one frame shift which could be read-trough during the translation in bacteria.

Among these clones, two in particular could be interesting having a high homology (100%) on the aminoacid sequence in Blastx analysis. They encode for ARF GTPase-activating protein GIT2 isoform 3 and dimethylaniline monooxygenase [N-oxide-forming] 4. Further, four proteins code for membrane proteins that are most likely antibody targets.

The G-protein-coupled receptor (GPCR)-kinase interacting proteins 1 and 2 (GIT1 and GIT2) are ubiquitous multidomain proteins involved in diverse cellular processes. They traffic between three distinct cellular compartments (cytoplasmic complexes, focal adhesions and the cell periphery) through interactions with proteins including ARF and PAK-interacting exchange factor (PIX). GITs and PIX cooperate to form large oligomeric complexes to which other proteins are transiently recruited (Hoefen and Berk 2006). GITs also participate in receptor internalization by regulating membrane trafficking between the plasma membrane and endosomes, targeting ARF GTPases

through their ARF GTPase-activating protein (ARF-GAP) activity. Furthermore, GITs act as scaffolds to control spatial activation of several signaling molecules. Finally, recent results suggest pathogenic roles for GIT proteins in Huntington's disease and HIV infection. The GITs appear to have specialized cytoskeletal functions in neurons. Neurite outgrowth requires Rac and ARF6 activity, and GIT1 mutations that abrogate ARF-GAP activity inhibit this process (Albertinazzi et al. 2003). GIT1 is enriched in pre and postsynaptic densities (Kim et al. 2003; Ko et al. 2003; Zhang et al. 2003), and inhibiting formation of GIT-PIX oligomers in cultured hippocampal neurons decreases the number of synapses formed along dendrites (Zhang et al. 2003). Recents findings indicated that GIT 2 is also implicated in the Regulation of thymocyte positive selection and motility. Thymocytes are highly motile cells that migrate under the influence of chemokines in distinct thymic compartments as they mature. The motility of thymocytes is tightly regulated; however, the molecular mechanisms that control thymocyte motility are not well understood. GIT2 was required for efficient positive selection in fact, Git2(-/-) double-positive thymocytes showed greater activation of the small GTPase Rac, actin polymerization and migration toward the chemokines CXCL12 (SDF-1) and CCL25 in vitro (Phee et al. 2010).

Obviously, all these observations are purely speculative, but, in general, confirm the idea that the autoimmune response in MS is so complex to require a diagnostic analysis that includes multiple markers simultaneously.

Adams, T. M., A. Wentzel and H. Kolmar (2005). "Intimin-mediated export of passenger proteins requires maintenance of a translocation-competent conformation." J Bacteriol **187**(2): 522-33.

Albertinazzi, C., L. Za, S. Paris and I. de Curtis (2003). "ADP-ribosylation factor 6 and a functional PIX/p95-APP1 complex are required for Rac1B-mediated neurite outgrowth." Mol Biol Cell **14**(4): 1295-307.

Alcaro, M. C. and A. M. Papini (2006). "Contribution of peptides to multiple sclerosis research." Biopolymers **84**(4): 349-67.

Allan, J., R. M. Fraser, T. Owen-Hughes and D. Keszenman-Pereyra (2012). "Micrococcal Nuclease Does Not Substantially Bias Nucleosome Mapping." J Mol Biol.

Alonso, A., M. A. Hernan and A. Ascherio (2008). "Allergy, family history of autoimmune diseases, and the risk of multiple sclerosis." Acta Neurol Scand **117**(1): 15-20.

Archelos, J. J., M. K. Storch and H. P. Hartung (2000). "The role of B cells and autoantibodies in multiple sclerosis." Ann Neurol **47**(6): 694-706.

Azzazy, H. M. and W. E. Highsmith, Jr. (2002). "Phage display technology: clinical applications and recent innovations." Clin Biochem **35**(6): 425-45.

Azzoni, E., D. Sblattero, M. Licciulli, R. Marzari and P. Edomi (2007). "Using archaeal histones for precise DNA fragmentation." Protein Eng Des Sel **20**(6): 267-71.

Baggio, R., P. Burgstaller, S. P. Hale, A. R. Putney, M. Lane, D. Lipovsek, M. C. Wright, R. W. Roberts, R. Liu, J. W. Szostak and R. W. Wagner (2002). "Identification of epitope-like consensus motifs using mRNA display." J Mol Recognit **15**(3): 126-34.

Baker, M. (2011). "Protein engineering: navigating between chance and reason." Nat Methods **8**(8): 623-6.

Baranov, P. V., O. L. Gurvich, O. Fayet, M. F. Prere, W. A. Miller, R. F. Gesteland, J. F. Atkins and M. C. Giddings (2001). "RECODE: a database of frameshifting, bypassing and codon redefinition utilized for gene expression." Nucleic Acids Res **29**(1): 264-7.

Beghetto, E., F. De Paolis, F. Montagnani, C. Cellesi and N. Gargano (2009). "Discovery of new Mycoplasma pneumoniae antigens by use of a whole-genome lambda display library." Microbes Infect **11**(1): 66-73.

Beghetto, E. and N. Gargano (2011
). "Lambda-display: a powerful tool for antigen discovery." Molecules **16**(4): 3089-105.

Beghetto, E., N. Gargano, S. Ricci, G. Garufi, S. Peppoloni, F. Montagnani, M. Oggioni, G. Pozzi and F. Felici (2006). "Discovery of novel Streptococcus pneumoniae antigens by screening a whole-genome lambda-display library." FEMS Microbiol Lett **262**(1): 14-21.

Beghetto, E., A. Spadoni, W. Buffolano, M. Del Pezzo, O. Minenkova, E. Pavoni, A. Pucci, R. Cortese, F. Felici and N. Gargano (2003). "Molecular dissection of the human B-cell response against Toxoplasma gondii infection by lambda display of cDNA libraries." Int J Parasitol **33**(2): 163-73.

Beghetto, E. G., N. (2011). "Lambda-display: a powerful tool for antigen discovery." Molecules **16**(4): 3089-105.

Bembich, S. (2004). Nuove strategie per l'identificazione di auto-antigeni nella Sclerosi Multipla. Università degli Studi di Trieste

Bembich, S. (2008). Identification of new diagnostic markers of Multiple Sclerosis using ORF-selected cDNA phage display libraries. Università degli Studi di Trieste.

Bennett, C. L., J. Christie, F. Ramsdell, M. E. Brunkow, P. J. Ferguson, L. Whitesell, T. E. Kelly, F. T. Saulsbury, P. F. Chance and H. D. Ochs (2001). "The immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndrome (IPEX) is caused by mutations of FOXP3." Nat Genet **27**(1): 20-1.

Besgen, P., P. Trommler, S. Vollmer and J. C. Prinz (2010). "Ezrin, maspin, peroxiredoxin 2, and heat shock protein 27: potential targets of a streptococcal-induced autoimmune response in psoriasis." J Immunol **184**(9): 5392-402.

Bessette, P. H., X. Hu, H. T. Soh and P. S. Daugherty (2007). "Microfluidic library screening for mapping antibody epitopes." Anal Chem **79**(5): 2174-8.

Bessette, P. H., J. J. Rice and P. S. Daugherty (2004). "Rapid isolation of high-affinity protein binding peptides using bacterial display." Protein Eng Des Sel **17**(10): 731-9.

Binz, H. K., M. T. Stumpp, P. Forrer, P. Amstutz and A. Pluckthun (2003). "Designing repeat proteins: well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins." J Mol Biol **332**(2): 489-503.

Blattner, F. R., G. Plunkett, 3rd, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau and Y. Shao (1997). "The complete genome sequence of Escherichia coli K-12." Science **277**(5331): 1453-62.

Bogdanova, E. A., D. A. Shagin and S. A. Lukyanov (2008). "Normalization of full-length enriched cDNA." Mol Biosyst **4**(3): 205-12.

Bradbury, A. R. (2010). "The use of phage display in neurobiology." Curr Protoc Neurosci **Chapter 5**: Unit 5 12.

Bradbury, A. R., S. Sidhu, S. Dubel and J. McCafferty (2011). "Beyond natural antibodies: the power of in vitro display technologies." Nat Biotechnol **29**(3): 245-54.

Brennan, K. M., F. Galban-Horcajo, S. Rinaldi, C. P. O'Leary, C. S. Goodyear, G. Kalna, A. Arthur, C. Elliot, S. Barnett, C. Linington, J. L. Bennett, G. P. Owens and H. J. Willison (2011). "Lipid arrays identify myelin-derived lipids and lipid complexes as prominent targets for oligoclonal band antibodies in multiple sclerosis." J Neuroimmunol **238**(1-2): 87-95.

Bright, J. J. (2007). "Curcumin and autoimmune disease." Adv Exp Med Biol **595**: 425-51.

Burnet, M. (1963). "Experimental Production of Auto-Antibodies or of Auto-Immune Disease." Br Med Bull **19**: 245-50.

Carcamo, J., M. W. Ravera, R. Brissette, O. Dedova, J. R. Beasley, A. Alam-Moghe, C. Wan, A. Blume and W. Mandecki (1998). "Unexpected frameshifts from gene to expressed protein in a phage-displayed peptide library." Proc Natl Acad Sci U S A **95**(19): 11146-51.

Carninci, P., Y. Shibata, N. Hayatsu, Y. Sugahara, K. Shibata, M. Itoh, H. Konno, Y. Okazaki, M. Muramatsu and Y. Hayashizaki (2000). "Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes." Genome Res **10**(10): 1617-30.

Cassan, C. and R. S. Liblau (2007). "Immune tolerance and control of CNS autoimmunity: from animal models to MS patients." J Neurochem **100**(4): 883-92.

Cesareni, G. (1992). "Peptide display on filamentous phage capsids. A new powerful tool to study protein-ligand interaction." FEBS Lett **307**(1): 66-70.

Cho, J. H. and P. K. Gregersen (2011). "Genomics and the multifactorial nature of human autoimmune disease." N Engl J Med **365**(17): 1612-23.

Choi, C. and E. N. Benveniste (2004). "Fas ligand/Fas system in the brain: regulator of immune and apoptotic responses." Brain Res Brain Res Rev **44**(1): 65-81.

Coche, T. and M. Dewez (1994). "Reducing bias in cDNA sequence representation by molecular selection." Nucleic Acids Res **22**(21): 4545-6.

Correale, J., M. Farez and W. Gilmore (2008). "Vaccines for multiple sclerosis: progress to date." CNS Drugs **22**(3): 175-98.

Cortese, I., S. Capone, S. Luchetti, R. Cortese and A. Nicosia (2001). "Cross-reactive phage-displayed mimotopes lead to the discovery of mimicry between HSV-1 and a brain-specific protein." J Neuroimmunol **113**(1): 119-28.

Cortini, A. (2005). Clonaggio della risposta immunitaria di pazienti con Sclerosi Multipla in librerie di Phage display e selezione di anticorpi ricombinanti contro antigeni del SNC.
. Università degli Studi di Trieste.

Cortini, A. (2008). Serum profiling and autoantibodies identification in Multiple Sclerosis using epitope and CSF IgG phage display libraries. Università degli studi di Trieste.

Cose, S., C. Brammer, K. M. Khanna, D. Masopust and L. Lefrancois (2006). "Evidence that a significant number of naive T cells enter non-lymphoid organs as part of a normal migratory pathway." Eur J Immunol **36**(6): 1423-33.

Crameri, R. and M. Suter (1993). "Display of biologically active proteins on the surface of filamentous phages: a cDNA cloning system for selection of functional gene products linked to the genetic information responsible for their production." Gene **137**(1): 69-75.

Croxford, J. L., J. K. Olson and S. D. Miller (2002). "Epitope spreading and molecular mimicry as triggers of autoimmunity in the Theiler's virus-induced demyelinating disease model of multiple sclerosis." Autoimmun Rev **1**(5): 251-60.

Cui, X., H. S. Nagesha and I. H. Holmes (2003). "Mapping of conformational epitopes on capsid protein VP2 of infectious bursal disease virus by fd-tet phage display." J Virol Methods **114**(1): 109-12.

D'Angelo, S., N. Velappan, F. Mignone, C. Santoro, D. Sblattero, C. Kiss and A. R. Bradbury (2011). "Filtering "genic" open reading frames from genomic DNA samples for advanced annotation." BMC Genomics **12 Suppl 1**: S5.

Dalakas, M. C. (2006). "B cells in the pathophysiology of autoimmune neurological disorders: a credible therapeutic target." Pharmacol Ther **112**(1): 57-70.

Danner, S. and J. G. Belasco (2001). "T7 phage display: a novel genetic selection system for cloning RNA-binding proteins from cDNA libraries." Proc Natl Acad Sci U S A **98**(23): 12954-9.

Daugherty, P. S. (2007). "Protein engineering with bacterial display." Curr Opin Struct Biol **17**(4): 474-80.

De Jager, P. L., X. Jia, J. Wang, P. I. de Bakker, L. Ottoboni, N. T. Aggarwal, L. Piccio, S. Raychaudhuri, D. Tran, C. Aubin, R. Briskin, S. Romano, S. E. Baranzini, J. L. McCauley, M. A. Pericak-Vance, J. L. Haines, R. A. Gibson, Y. Naeglin, B. Uitdehaag, P. M. Matthews, L. Kappos, C. Polman, W. L. McArdle, D. P. Strachan, D. Evans, A. H. Cross, M. J. Daly, A. Compston, S. J. Sawcer, H. L. Weiner, S. L. Hauser, D. A. Hafler and J. R. Oksenberg (2009). "Meta-analysis of genome scans and replication identify CD6, IRF8 and TNFRSF1A as new multiple sclerosis susceptibility loci." Nat Genet **41**(7): 776-82.

De Paolis, F., E. Beghetto, A. Spadoni, F. Montagnani, F. Felici, M. R. Oggioni and N. Gargano (2007). "Identification of a human immunodominant B-cell epitope within the immunoglobulin A1 protease of Streptococcus pneumoniae." BMC Microbiol **7**: 113.

Deroo, S. and C. P. Muller (2001). "Antigenic and immunogenic phage displayed mimotopes as substitute antigens: applications and limitations." Comb Chem High Throughput Screen **4**(1): 75-110.

Diatchenko, L., S. Lukyanov, Y. F. Lau and P. D. Siebert (1999). "Suppression subtractive hybridization: a versatile method for identifying differentially expressed genes." Methods Enzymol **303**: 349-80.

Dokland, T. and H. Murialdo (1993). "Structural transitions during maturation of bacteriophage lambda capsids." J Mol Biol **233**(4): 682-94.

Faix, P. H., M. A. Burg, M. Gonzales, E. P. Ravey, A. Baird and D. Larocca (2004). "Phage display of cDNA libraries: enrichment of cDNA expression using open reading frame selection." Biotechniques **36**(6): 1018-22, 1024, 1026-9.

Feldhaus, M. J., R. W. Siegel, L. K. Opresko, J. R. Coleman, J. M. Feldhaus, Y. A. Yeung, J. R. Cochran, P. Heinzelman, D. Colby, J. Swers, C. Graff, H. S. Wiley and K. D. Wittrup (2003). "Flow-cytometric isolation of human antibodies from a nonimmune Saccharomyces cerevisiae surface display library." Nat Biotechnol **21**(2): 163-70.

Forooghian, F., R. K. Cheung, W. C. Smith, P. O'Connor and H. M. Dosch (2007). "Enolase and arrestin are novel nonmyelin autoantigens in multiple sclerosis." J Clin Immunol **27**(4): 388-96.

Forrer, P., M. T. Stumpp, H. K. Binz and A. Pluckthun (2003). "A novel strategy to design binding molecules harnessing the modular nature of repeat proteins." FEBS Lett **539**(1-3): 2-6.

Forrester, J. V., H. Xu, T. Lambe and R. Cornall (2008). "Immune privilege or privileged immunity?" Mucosal Immunol **1**(5): 372-81.

Fujimori, J., I. Nakashima, K. Fujihara, T. Misu, S. Sato and Y. Itoyama (2011). "Epitope analysis of cerebrospinal fluid IgG in Japanese multiple sclerosis patients using phage display method." Mult Scler Int **2011**: 353417.

Garufi, G., O. Minenkova, C. Lo Passo, I. Pernice and F. Felici (2005). "Display libraries on bacteriophage lambda capsid." Biotechnol Annu Rev **11**: 153-90.

Geysen, H. M., R. H. Meloen and S. J. Barteling (1984). "Use of peptide synthesis to probe viral antigens for epitopes to a resolution of a single amino acid." Proc Natl Acad Sci U S A **81**(13): 3998-4002.

Gibbs, K. A., D. D. Isaac, J. Xu, R. W. Hendrix, T. J. Silhavy and J. A. Theriot (2004). "Complex spatial distribution and dynamics of an abundant Escherichia coli outer membrane protein, LamB." Mol Microbiol **53**(6): 1771-83.

Glokler, J., T. Schutze and Z. Konthur (2010). "Automation in the high-throughput selection of random combinatorial libraries--different approaches for select applications." Molecules **15**(4): 2478-90.

Grayling, R. A., K. A. Bailey and J. N. Reeve (1997). "DNA binding and nuclease protection by the HMf histones from the hyperthermophilic archaeon Methanothermus fervidus." Extremophiles **1**(2): 79-88.

Guerau-de-Arellano, M., M. Martinic, C. Benoist and D. Mathis (2009). "Neonatal tolerance revisited: a perinatal window for Aire control of autoimmunity." J Exp Med **206**(6): 1245-52.

Guggino, G., A. R. Giardina, F. Ciccia, G. Triolo, F. Dieli and G. Sireci (2011). "Are Toll-like receptors and decoy receptors involved in the immunopathogenesis of systemic lupus erythematosus and lupus-like syndromes?" Clin Dev Immunol **2012**: 135932.

Hanes, J., C. Schaffitzel, A. Knappik and A. Pluckthun (2000). "Picomolar affinity antibodies from a fully synthetic naive library selected and evolved by ribosome display." Nat Biotechnol **18**(12): 1287-92.

Hardwick, K. G., J. C. Boothroyd, A. D. Rudner and H. R. Pelham (1992). "Genes that allow yeast cells to grow in the absence of the HDEL receptor." EMBO J **11**(11): 4187-95.

Hartung, H. P. and B. C. Kieseier (2010). "Atacicept: targeting B cells in multiple sclerosis." Ther Adv Neurol Disord **3**(4): 205-16.

Hemmer, B., S. Nessler, D. Zhou, B. Kieseier and H. P. Hartung (2006). "Immunopathogenesis and immunotherapy of multiple sclerosis." Nat Clin Pract Neurol **2**(4): 201-11.
Hoefen, R. J. and B. C. Berk (2006). "The multifunctional GIT family of proteins." J Cell Sci **119**(Pt 8): 1469-75.
Hu, S., A. Vissink, M. Arellano, C. Roozendaal, H. Zhou, C. G. Kallenberg and D. T. Wong (2011). "Identification of autoantibody biomarkers for primary Sjogren's syndrome using protein microarrays." Proteomics **11**(8): 1499-507.
Hueber, W. and W. H. Robinson (2006). "Proteomic biomarkers for autoimmune disease." Proteomics **6**(14): 4100-5.
Hueber, W., P. J. Utz, L. Steinman and W. H. Robinson (2002). "Autoantibody profiling for the study and treatment of autoimmune disease." Arthritis Res **4**(5): 290-5.
Hust, M., M. Meysing, T. Schirrmann, M. Selke, J. Meens, G. F. Gerlach and S. Dubel (2006). "Enrichment of open reading frames presented on bacteriophage M13 using hyperphage." Biotechniques **41**(3): 335-42.
Ice, J. A., H. Li, I. Adrianto, P. C. Lin, J. A. Kelly, C. G. Montgomery, C. J. Lessard and K. L. Moser (2011). "Genetics of Sjogren's syndrome in the genome-wide association era." J Autoimmun.
James, L. C., P. Roversi and D. S. Tawfik (2003). "Antibody multispecificity mediated by conformational diversity." Science **299**(5611): 1362-7.
Jarius, S., P. Eichhorn, C. Jacobi, B. Wildemann, M. Wick and R. Voltz (2009). "The intrathecal, polyspecific antiviral immune response: Specific for MS or a general marker of CNS autoimmunity?" J Neurol Sci **280**(1-2): 98-100.
Jestin, J. L. (2008). "Functional cloning by phage display." Biochimie **90**(9): 1273-8.
Jongeneel, C. V., C. Iseli, B. J. Stevenson, G. J. Riggins, A. Lal, A. Mackay, R. A. Harris, M. J. O'Hare, A. M. Neville, A. J. Simpson and R. L. Strausberg (2003). "Comprehensive sampling of gene expression in human cell lines with massively parallel signature sequencing." Proc Natl Acad Sci U S A **100**(8): 4702-5.
Jung, H. C., J. M. Lebeault and J. G. Pan (1998). "Surface display of Zymomonas mobilis levansucrase by using the ice-nucleation protein of Pseudomonas syringae." Nat Biotechnol **16**(6): 576-80.
Kabat, E. A., D. A. Freedman and et al. (1950). "A study of the crystalline albumin, gamma globulin and total protein in the cerebrospinal fluid of 100 cases of multiple sclerosis and in other diseases." Am J Med Sci **219**(1): 55-64.
Kalnina, Z., K. Silina, I. Meistere, P. Zayakin, A. Rivosh, A. Abols, M. Leja, O. Minenkova, D. Schadendorf and A. Line (2008). "Evaluation of T7 and lambda phage display systems for survey of autoantibody profiles in cancer patients." J Immunol Methods **334**(1-2): 37-50.
Kaplan, H. J. and J. Y. Niederkorn (2007). "Regional immunity and immune privilege." Chem Immunol Allergy **92**: 11-26.
Kehoe, J. W. and B. K. Kay (2005). "Filamentous phage display in the new millennium." Chem Rev **105**(11): 4056-72.
Kemp, E. H., L. M. Herd, E. A. Waterman, A. G. Wilson, A. P. Weetman and P. P. Watson (2002). "Immunoscreening of phage-displayed cDNA-encoded polypeptides identifies B cell targets in autoimmune disease." Biochem Biophys Res Commun **298**(1): 169-77.
Khoury, S. J., W. W. Hancock and H. L. Weiner (1992). "Oral tolerance to myelin basic protein and natural recovery from experimental autoimmune encephalomyelitis are associated with downregulation of inflammatory cytokines and differential upregulation of transforming growth factor beta, interleukin 4, and prostaglandin E expression in the brain." J Exp Med **176**(5): 1355-64.

Kim, S., J. Ko, H. Shin, J. R. Lee, C. Lim, J. H. Han, W. D. Altrock, C. C. Garner, E. D. Gundelfinger, R. T. Premont, B. K. Kaang and E. Kim (2003). "The GIT family of proteins forms multimers and associates with the presynaptic cytomatrix protein Piccolo." J Biol Chem **278**(8): 6291-300.

Knappik, A., L. Ge, A. Honegger, P. Pack, M. Fischer, G. Wellnhofer, A. Hoess, J. Wolle, A. Pluckthun and B. Virnekas (2000). "Fully synthetic human combinatorial antibody libraries (HuCAL) based on modular consensus frameworks and CDRs randomized with trinucleotides." J Mol Biol **296**(1): 57-86.

Ko, J., S. Kim, J. G. Valtschanoff, H. Shin, J. R. Lee, M. Sheng, R. T. Premont, R. J. Weinberg and E. Kim (2003). "Interaction between liprin-alpha and GIT1 is required for AMPA receptor targeting." J Neurosci **23**(5): 1667-77.

Ko, M. S. (1990). "An 'equalized cDNA library' by the reassociation of short double-stranded cDNAs." Nucleic Acids Res **18**(19): 5705-11.

Kohl, A., H. K. Binz, P. Forrer, M. T. Stumpp, A. Pluckthun and M. G. Grutter (2003). "Designed to be stable: crystal structure of a consensus ankyrin repeat protein." Proc Natl Acad Sci U S A **100**(4): 1700-5.

Kondo, A. and M. Ueda (2004). "Yeast cell-surface display--applications of molecular display." Appl Microbiol Biotechnol **64**(1): 28-40.

Korndorfer, I. P. and A. Skerra (2002). "Improved affinity of engineered streptavidin for the Strep-tag II peptide is due to a fixed open conformation of the lid-like loop at the binding site." Protein Sci **11**(4): 883-93.

Krogsgaard, M., K. W. Wucherpfennig, B. Cannella, B. E. Hansen, A. Svejgaard, J. Pyrdol, H. Ditzel, C. Raine, J. Engberg and L. Fugger (2000). "Visualization of myelin basic protein (MBP) T cell epitopes in multiple sclerosis lesions using a monoclonal antibody specific for the human histocompatibility leukocyte antigen (HLA)-DR2-MBP 85-99 complex." J Exp Med **191**(8): 1395-412.

Kruger, D. H. and C. Schroeder (1981). "Bacteriophage T3 and bacteriophage T7 virus-host cell interactions." Microbiol Rev **45**(1): 9-51.

Kuhle, J., C. Pohl, M. Mehling, G. Edan, M. S. Freedman, H. P. Hartung, C. H. Polman, D. H. Miller, X. Montalban, F. Barkhof, L. Bauer, S. Dahms, R. Lindberg, L. Kappos and R. Sandbrink (2007). "Lack of association between antimyelin antibodies and progression to multiple sclerosis." N Engl J Med **356**(4): 371-8.

Kuroda, K., M. Ueda, S. Shibasaki and A. Tanaka (2002). "Cell surface-engineered yeast with ability to bind, and self-aggregate in response to, copper ion." Appl Microbiol Biotechnol **59**(2-3): 259-64.

Lalive, P. H., T. Menge, C. Delarasse, B. Della Gaspera, D. Pham-Dinh, P. Villoslada, H. C. von Budingen and C. P. Genain (2006). "Antibodies to native myelin oligodendrocyte glycoprotein are serologic markers of early inflammation in multiple sclerosis." Proc Natl Acad Sci U S A **103**(7): 2280-5.

Larman, H. B., Z. Zhao, U. Laserson, M. Z. Li, A. Ciccia, M. A. Gakidis, G. M. Church, S. Kesari, E. M. Leproust, N. L. Solimini and S. J. Elledge (2011). "Autoantigen discovery with a synthetic human peptidome." Nat Biotechnol **29**(6): 535-41.

Lee, S. Y., J. H. Choi and Z. Xu (2003). "Microbial cell-surface display." Trends Biotechnol **21**(1): 45-52.

Lesko, L. J. and A. J. Atkinson, Jr. (2001). "Use of biomarkers and surrogate endpoints in drug development and regulatory decision making: criteria, validation, strategies." Annu Rev Pharmacol Toxicol **41**: 347-66.

Leslie, D., P. Lipsky and A. L. Notkins (2001). "Autoantibodies as predictors of disease." J Clin Invest **108**(10): 1417-22.

Li, M. (2000). "Applications of display technology in protein analysis." Nat Biotechnol **18**(12): 1251-6.
Li, R., R. H. Hoess, J. S. Bennett and W. F. DeGrado (2003). "Use of phage display to probe the evolution of binding specificity and affinity in integrins." Protein Eng **16**(1): 65-72.
Li, W. and N. B. Caberoy (2009). "New perspective for phage display as an efficient and versatile technology of functional proteomics." Appl Microbiol Biotechnol **85**(4): 909-19.
Li, W. and N. B. Caberoy (2010). "New perspective for phage display as an efficient and versatile technology of functional proteomics." Appl Microbiol Biotechnol **85**(4): 909-19.
Lipke, P. N. and J. Kurjan (1992). "Sexual agglutination in budding yeasts: structure, function, and regulation of adhesion glycoproteins." Microbiol Rev **56**(1): 180-94.
Lipovsek, D. and A. Pluckthun (2004). "In-vitro protein evolution by ribosome display and mRNA display." J Immunol Methods **290**(1-2): 51-67.
Liston, A., S. Lesage, D. H. Gray, R. L. Boyd and C. C. Goodnow (2005). "Genetic lesions in T-cell tolerance and thresholds for autoimmunity." Immunol Rev **204**: 87-101.
Lofblom, J. (2011). "Bacterial display in combinatorial protein engineering." Biotechnol J **6**(9): 1115-29.
Lu, C. F., R. C. Montijn, J. L. Brown, F. Klis, J. Kurjan, H. Bussey and P. N. Lipke (1995). "Glycosyl phosphatidylinositol-dependent cross-linking of alpha-agglutinin and beta 1,6-glucan in the Saccharomyces cerevisiae cell wall." J Cell Biol **128**(3): 333-40.
Lu, Z., E. R. LaVallie and J. M. McCoy (2003). "Using bio-panning of FLITRX peptide libraries displayed on E. coli cell surface to study protein-protein interactions." Methods Mol Biol **205**: 267-80.
Lu, Z., K. S. Murray, V. Van Cleave, E. R. LaVallie, M. L. Stahl and J. M. McCoy (1995). "Expression of thioredoxin random peptide libraries on the Escherichia coli cell surface as functional fusions to flagellin: a system designed for exploring protein-protein interactions." Biotechnology (N Y) **13**(4): 366-72.
Luk'ianov, K. A., I. G. Gurskaia, M. V. Matts, G. L. Khaspekov, B. D'Iachenko L, A. A. Chenchik, S. G. Il'evich-Stuchkov and S. A. Luk'ianov (1996). "[A method for obtaining the normalized cDNA libraries based on the effect of suppression of polymerase chain reaction]." Bioorg Khim **22**(9): 686-90.
Luk'ianov, S. A., N. G. Gurskaia, K. A. Luk'ianov, V. S. Tarabykin and E. D. Sverdlov (1994). "[Highly-effective subtractive hybridization of cDNA]." Bioorg Khim **20**(6): 701-4.
Lunemann, J. D., T. Kamradt, R. Martin and C. Munz (2007). "Epstein-barr virus: environmental trigger of multiple sclerosis?" J Virol **81**(13): 6777-84.
Mackay, I. R. and M. J. Rowley (2004). "Autoimmune epitopes: autoepitopes." Autoimmun Rev **3**(7-8): 487-92.
Mannie, M. D., K. D. Prevost and C. A. Marinakis (1995). "Prostaglandin E2 promotes the induction of anergy during T helper cell recognition of myelin basic protein." Cell Immunol **160**(1): 132-8.
Manz, R. A., A. E. Hauser, F. Hiepe and A. Radbruch (2005). "Maintenance of serum antibody levels." Annu Rev Immunol **23**: 367-86.
Maruyama, I. N., H. I. Maruyama and S. Brenner (1994). "Lambda foo: a lambda phage vector for the expression of foreign proteins." Proc Natl Acad Sci U S A **91**(17): 8273-7.
Matsumoto, T., H. Fukuda, M. Ueda, A. Tanaka and A. Kondo (2002). "Construction of yeast strains with high cell surface lipase activity by using novel display systems based on the Flo1p flocculation functional domain." Appl Environ Microbiol **68**(9): 4517-22.

Matthews, L. J., R. Davis and G. P. Smith (2002). "Immunogenically fit subunit vaccine components via epitope discovery from natural peptide libraries." J Immunol **169**(2): 837-46.

Maurer, J., J. Jose and T. F. Meyer (1997). "Autodisplay: one-component system for efficient surface display and release of soluble recombinant proteins from Escherichia coli." J Bacteriol **179**(3): 794-804.

Menard, L., D. Saadoun, I. Isnardi, Y. S. Ng, G. Meyers, C. Massad, C. Price, C. Abraham, R. Motaghedi, J. H. Buckner, P. K. Gregersen and E. Meffre (2011). "The PTPN22 allele encoding an R620W variant interferes with the removal of developing autoreactive B cells in humans." J Clin Invest **121**(9): 3635-44.

Menge, T., H. C. von Budingen, P. H. Lalive and C. P. Genain (2007). "Relevant antibody subsets against MOG recognize conformational epitopes exclusively exposed in solid-phase ELISA." Eur J Immunol **37**(11): 3229-39.

Monson, N. L., P. D. Cravens, E. M. Frohman, K. Hawker and M. K. Racke (2005). "Effect of rituximab on the peripheral blood and cerebrospinal fluid B cells in patients with primary progressive multiple sclerosis." Arch Neurol **62**(2): 258-64.

Murai, T., M. Ueda, M. Yamamura, H. Atomi, Y. Shibasaki, N. Kamasawa, M. Osumi, T. Amachi and A. Tanaka (1997). "Construction of a starch-utilizing yeast by cell surface engineering." Appl Environ Microbiol **63**(4): 1362-6.

Nakamura, Y., S. Shibasaki, M. Ueda, A. Tanaka, H. Fukuda and A. Kondo (2001). "Development of novel whole-cell immunoadsorbents by yeast surface display of the IgG-binding domain." Appl Microbiol Biotechnol **57**(4): 500-5.

O'Dell, J. R. (2004). "Therapeutic strategies for rheumatoid arthritis." N Engl J Med **350**(25): 2591-602.

Odendahl, M., H. Mei, B. F. Hoyer, A. M. Jacobi, A. Hansen, G. Muehlinghaus, C. Berek, F. Hiepe, R. Manz, A. Radbruch and T. Dorner (2005). "Generation of migratory antigen-specific plasma blasts and mobilization of resident plasma cells in a secondary immune response." Blood **105**(4): 1614-21.

Oh, S., C. Cudrici, T. Ito and H. Rus (2008). "B-cells and humoral immunity in multiple sclerosis. Implications for therapy." Immunol Res **40**(3): 224-34.

Oh, S., A. L. Rankin and A. J. Caton (2010). "CD4+CD25+ regulatory T cells in autoimmune arthritis." Immunol Rev **233**(1): 97-111.

Pande, J., M. M. Szewczyk and A. K. Grover (2010). "Phage display: concept, innovations, applications and future." Biotechnol Adv **28**(6): 849-58.

Paschke, M. (2006). "Phage display systems and their applications." Appl Microbiol Biotechnol **70**(1): 2-11.

Patanjali, S. R., S. Parimoo and S. M. Weissman (1991). "Construction of a uniform-abundance (normalized) cDNA library." Proc Natl Acad Sci U S A **88**(5): 1943-7.

Phee, H., I. Dzhagalov, M. Mollenauer, Y. Wang, D. J. Irvine, E. Robey and A. Weiss (2010). "Regulation of thymocyte positive selection and motility by GIT2." Nat Immunol **11**(6): 503-11.

Plebani, M., M. Pittoni, M. Celadin, D. Bernardi and M. M. Mion (2009). "Recent advances in diagnostic technologies for autoimmune diseases." Autoimmun Rev **8**(3): 238-43.

Polman, C. H. and J. Killestein (2006). "Anti-myelin antibodies in multiple sclerosis: clinically useful?" J Neurol Neurosurg Psychiatry **77**(6): 712.

Prodromou, C., R. Savva and P. C. Driscoll (2007). "DNA fragmentation-based combinatorial approaches to soluble protein expression Part I. Generating DNA fragment libraries." Drug Discov Today **12**(21-22): 931-8.

Puccetti, A. and C. Lunardi (2010). "The role of peptide libraries in the identification of novel autoantigen targets in autoimmune diseases." Discov Med **9**(46): 224-8.

Qin, Q. P., L. L. Zhang, N. Y. Li, Y. Y. Cui and K. Xu (2010). "[Optimizing of cDNA preparation for next generation sequencing]." Yi Chuan **32**(9): 974-7.

Racke, M. K. (2008). "The role of B cells in multiple sclerosis: rationale for B-cell-targeted therapies." Curr Opin Neurol **21 Suppl 1**: S9-S18.

Ransohoff, R. M., P. Kivisakk and G. Kidd (2003). "Three or more routes for leukocyte migration into the central nervous system." Nat Rev Immunol **3**(7): 569-81.

Reindl, M., M. Khalil and T. Berger (2006). "Antibodies as biological markers for pathophysiological processes in MS." J Neuroimmunol **180**(1-2): 50-62.

Rhyner, C., R. Kodzius and R. Crameri (2002). "Direct selection of cDNAs from filamentous phage surface display libraries: potential and limitations." Curr Pharm Biotechnol **3**(1): 13-21.

Rice, J. J., A. Schohn, P. H. Bessette, K. T. Boulware and P. S. Daugherty (2006). "Bacterial display using circularly permuted outer membrane protein OmpX yields high affinity peptide ligands." Protein Sci **15**(4): 825-36.

Robinson, W. H. (2006). "Antigen arrays for antibody profiling." Curr Opin Chem Biol **10**(1): 67-72.

Rolan, P., A. J. Atkinson, Jr. and L. J. Lesko (2003). "Use of biomarkers from drug discovery through clinical practice: report of the Ninth European Federation of Pharmaceutical Sciences Conference on Optimizing Drug Development." Clin Pharmacol Ther **73**(4): 284-91.

Rott, S. and U. Mrowietz (2005). "Recent developments in the use of biologics in psoriasis and autoimmune disorders. The role of autoantibodies." BMJ **330**(7493): 716-20.

Sandman, K., J. A. Krzycki, B. Dobrinski, R. Lurz and J. N. Reeve (1990). "HMf, a DNA-binding protein isolated from the hyperthermophilic archaeon Methanothermus fervidus, is most closely related to histones." Proc Natl Acad Sci U S A **87**(15): 5788-91.

Santini, C., D. Brennan, C. Mennuni, R. H. Hoess, A. Nicosia, R. Cortese and A. Luzzago (1998). "Efficient display of an HCV cDNA expression library as C-terminal fusion to the capsid protein D of bacteriophage lambda." J Mol Biol **282**(1): 125-35.

Sasaki, Y. F., D. Ayusawa and M. Oishi (1994). "Construction of a normalized cDNA library by introduction of a semi-solid mRNA-cDNA hybridization system." Nucleic Acids Res **22**(6): 987-92.

Sblattero, D. and A. Bradbury (2000). "Exploiting recombination in single bacteria to make large phage antibody libraries." Nat Biotechnol **18**(1): 75-80.

Schekman, R. (1992). "Genetic and biochemical analysis of vesicular traffic in yeast." Curr Opin Cell Biol **4**(4): 587-92.

Schreuder, M. P., S. Brekelmans, H. van den Ende and F. M. Klis (1993). "Targeting of a heterologous protein to the cell wall of Saccharomyces cerevisiae." Yeast **9**(4): 399-409.

Scofield, R. H. (2004). "Autoantibodies as predictors of disease." Lancet **363**(9420): 1544-6.

Shagin, D. A., D. V. Rebrikov, V. B. Kozhemyako, I. M. Altshuler, A. S. Shcheglov, P. A. Zhulidov, E. A. Bogdanova, D. B. Staroverov, V. A. Rasskazov and S. Lukyanov (2002). "A novel method for SNP detection using a new duplex-specific nuclease from crab hepatopancreas." Genome Res **12**(12): 1935-42.

Sherer, Y., A. Gorstein, M. J. Fritzler and Y. Shoenfeld (2004). "Autoantibody explosion in systemic lupus erythematosus: more than 100 different antibodies found in SLE patients." Semin Arthritis Rheum **34**(2): 501-37.

Shibasaki, S., H. Maeda and M. Ueda (2009). "Molecular display technology using yeast--arming technology." Anal Sci **25**(1): 41-9.

Shibasaki, S., M. Ueda, T. Iizuka, M. Hirayama, Y. Ikeda, N. Kamasawa, M. Osumi and A. Tanaka (2001). "Quantitative evaluation of the enhanced green fluorescent protein

displayed on the cell surface of Saccharomyces cerevisiae by fluorometric and confocal laser scanning microscopic analyses." Appl Microbiol Biotechnol **55**(4): 471-5.

Shikama, N., G. Nusspaumer and G. A. Hollander (2009). "Clearing the AIRE: on the pathophysiological basis of the autoimmune polyendocrinopathy syndrome type-1." Endocrinol Metab Clin North Am **38**(2): 273-88, vii.

Shimojyo, R., H. Furukawa, H. Fukuda and A. Kondo (2003). "Preparation of yeast strains displaying IgG binding domain ZZ and enhanced green fluorescent protein for novel antigen detection systems." J Biosci Bioeng **96**(5): 493-5.

Silverstein, A. M. (2005). "Paul Ehrlich, archives and the history of immunology." Nat Immunol **6**(7): 639.

Smith, G. P. (1985). "Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface." Science **228**(4705): 1315-7.

Soares, M. B., M. F. Bonaldo, P. Jelene, L. Su, L. Lawton and A. Efstratiadis (1994). "Construction and characterization of a normalized cDNA library." Proc Natl Acad Sci U S A **91**(20): 9228-32.

Somers, V., C. Govarts, K. Somers, R. Hupperts, R. Medaer and P. Stinissen (2008). "Autoantibody profiling in multiple sclerosis reveals novel antigenic candidates." J Immunol **180**(6): 3957-63.

Svejgaard, A. (2008). "The immunogenetics of multiple sclerosis." Immunogenetics **60**(6): 275-86.

Tektonidou, M. G. and M. M. Ward (2011). "Validation of new biomarkers in systemic autoimmune diseases." Nat Rev Rheumatol.

Tomassini, V., L. De Giglio, M. Reindl, P. Russo, I. Pestalozza, P. Pantano, T. Berger and C. Pozzilli (2007). "Anti-myelin antibodies predict the clinical outcome after a first episode suggestive of MS." Mult Scler **13**(9): 1086-94.

Tong, A. H., B. Drees, G. Nardelli, G. D. Bader, B. Brannetti, L. Castagnoli, M. Evangelista, S. Ferracuti, B. Nelson, S. Paoluzi, M. Quondam, A. Zucconi, C. W. Hogue, S. Fields, C. Boone and G. Cesareni (2002). "A combined experimental and computational strategy to define protein interaction networks for peptide recognition modules." Science **295**(5553): 321-4.

Uccelli, A., F. Aloisi and V. Pistoia (2005). "Unveiling the enigma of the CNS as a B-cell fostering environment." Trends Immunol **26**(5): 254-9.

Ueda, M. and A. Tanaka (2000). "Cell surface engineering of yeast: construction of arming yeast with biocatalyst." J Biosci Bioeng **90**(2): 125-36.

Ullman, C. G., L. Frigotto and R. N. Cooley (2011). "In vitro methods for peptide display and their applications." Brief Funct Genomics **10**(3): 125-34.

van Bloois, E., R. T. Winter, H. Kolmar and M. W. Fraaije (2010). "Decorating microbes: surface display of proteins on Escherichia coli." Trends Biotechnol **29**(2): 79-86.

van der Vaart, J. M., L. H. Caro, J. W. Chapman, F. M. Klis and C. T. Verrips (1995). "Identification of three mannoproteins in the cell wall of Saccharomyces cerevisiae." J Bacteriol **177**(11): 3104-10.

Vanderlugt, C. L., W. S. Begolka, K. L. Neville, Y. Katz-Levy, L. M. Howard, T. N. Eagar, J. A. Bluestone and S. D. Miller (1998). "The functional significance of epitope spreading and its regulation by co-stimulatory molecules." Immunol Rev **164**: 63-72.

Veiga, E., V. de Lorenzo and L. A. Fernandez (2004). "Structural tolerance of bacterial autotransporters for folded passenger protein domains." Mol Microbiol **52**(4): 1069-80.

von Boehmer, H. and F. Melchers (2010). "Checkpoints in lymphocyte development and autoimmune disease." Nat Immunol **11**(1): 14-20.

Walsh, K. and M. Sata (1999). "Is extravasation a Fas-regulated process?" Mol Med Today **5**(2): 61-7.
Watari, J., Y. Takata, M. Ogawa, H. Sahara, S. Koshino, M. L. Onnela, U. Airaksinen, R. Jaatinen, M. Penttila and S. Keranen (1994). "Molecular cloning and analysis of the yeast flocculation gene FLO1." Yeast **10**(2): 211-25.
Weichhart, T., M. Horky, J. Sollner, S. Gangl, T. Henics, E. Nagy, A. Meinke, A. von Gabain, C. M. Fraser, S. R. Gill, M. Hafner and U. von Ahsen (2003). "Functional selection of vaccine candidate peptides from Staphylococcus aureus whole-genome expression libraries in vitro." Infect Immun **71**(8): 4633-41.
Wenkel, H., J. W. Streilein and M. J. Young (2000). "Systemic immune deviation in the brain that does not depend on the integrity of the blood-brain barrier." J Immunol **164**(10): 5125-31.
Wilbanks, G. A. and J. W. Streilein (2005). "Fluids from immune privileged sites endow macrophages with the capacity to induce antigen-specific immune deviation via a mechanism involving transforming growth factor-beta." Eur J Immunol **22**(4): 1031-6.
Willats, W. G. (2002). "Phage display: practicalities and prospects." Plant Mol Biol **50**(6): 837-54.
Wing, K. and S. Sakaguchi (2010). "Regulatory T cells exert checks and balances on self tolerance and autoimmunity." Nat Immunol **11**(1): 7-13.
Winter, G., A. D. Griffiths, R. E. Hawkins and H. R. Hoogenboom (1994). "Making antibodies by phage display technology." Annu Rev Immunol **12**: 433-55.
Xu, L., P. Aha, K. Gu, R. G. Kuimelis, M. Kurz, T. Lam, A. C. Lim, H. Liu, P. A. Lohse, L. Sun, S. Weng, R. W. Wagner and D. Lipovsek (2002). "Directed evolution of high-affinity antibody mimics using mRNA display." Chem Biol **9**(8): 933-42.
Yang, T. H., J. G. Pan, Y. S. Seo and J. S. Rhee (2004). "Use of Pseudomonas putida EstA as an anchoring motif for display of a periplasmic enzyme on the surface of Escherichia coli." Appl Environ Microbiol **70**(12): 6968-76.
Yi, H., Y. J. Cho, S. Won, J. E. Lee, H. Jin Yu, S. Kim, G. P. Schroth, S. Luo and J. Chun (2011). "Duplex-specific nuclease efficiently removes rRNA for prokaryotic RNA-seq." Nucleic Acids Res **39**(20): e140.
Yu, X., D. H. Gilden, A. M. Ritchie, M. P. Burgoon, K. M. Keays and G. P. Owens (2006). "Specificity of recombinant antibodies generated from multiple sclerosis cerebrospinal fluid probed with a random peptide library." J Neuroimmunol **172**(1-2): 121-31.
Zacchi, P., D. Sblattero, F. Florian, R. Marzari and A. R. Bradbury (2003). "Selecting open reading frames from DNA." Genome Res **13**(5): 980-90.
Zahnd, C., S. Spinelli, B. Luginbuhl, P. Amstutz, C. Cambillau and A. Pluckthun (2004). "Directed in vitro evolution and crystallographic analysis of a peptide-binding single chain antibody fragment (scFv) with low picomolar affinity." J Biol Chem **279**(18): 18870-7.
Zhang, H., D. J. Webb, H. Asmussen and A. F. Horwitz (2003). "Synapse formation is regulated by the signaling adaptor GIT1." J Cell Biol **161**(1): 131-42.
Zhang, Y., J. L. Davis and W. Li (2005). "Identification of tribbles homolog 2 as an autoantigen in autoimmune uveitis by phage display." Mol Immunol **42**(11): 1275-81.
Zhao, Y., H. Hoshiyama, J. W. Shay and W. E. Wright (2008). "Quantitative telomeric overhang determination using a double-strand specific nuclease." Nucleic Acids Res **36**(3): e14.
Zhou, D., R. Srivastava, S. Nessler, V. Grummel, N. Sommer, W. Bruck, H. P. Hartung, C. Stadelmann and B. Hemmer (2006). "Identification of a pathogenic antibody

response to native myelin oligodendrocyte glycoprotein in multiple sclerosis." Proc Natl Acad Sci U S A **103**(50): 19057-62.

Zhu, C., A. C. Anderson, A. Schubart, H. Xiong, J. Imitola, S. J. Khoury, X. X. Zheng, T. B. Strom and V. K. Kuchroo (2005). "The Tim-3 ligand galectin-9 negatively regulates T helper type 1 immunity." Nat Immunol **6**(12): 1245-52.

Zhulidov, P. A., E. A. Bogdanova, A. S. Shcheglov, I. A. Shagina, L. L. Vagner, G. L. Khazpekov, V. V. Kozhemiako, S. A. Luk'ianov and D. A. Shagin (2005). "[A method for the preparation of normalized cDNA libraries enriched with full-length sequences]." Bioorg Khim **31**(2): 186-94.

Zhulidov, P. A., E. A. Bogdanova, A. S. Shcheglov, L. L. Vagner, G. L. Khaspekov, V. B. Kozhemyako, M. V. Matz, E. Meleshkevitch, L. L. Moroz, S. A. Lukyanov and D. A. Shagin (2004). "Simple cDNA normalization using kamchatka crab duplex-specific nuclease." Nucleic Acids Res **32**(3): e37.

// Ringraziamenti

Desidero ringraziare il dott. Paolo Edomi per avermi dato la possibilità di svolgere il dottorato presso il suo laboratorio.

Voglio inoltre ringraziare tutti coloro che hanno condiviso con me questi tre anni di dottorato, compagni non solo di laboratorio ma anche di vita. In particolare Marta, Ilenia, Imma, Gabriella, Sara, Sandro, il dott. Giulianini, Sabrina e tutte le persone che per un periodo più o meno lungo hanno rallegrato questi anni.

Alla mia famiglia e a Mario va un grazie particolare, che va oltre a questo lavoro di Tesi. Per il profondo amore che ogni giorno ricevo da voi.