ALMA MATER STUDIORUM – UNIVERSITY OF BOLOGNA

AGRICULTURAL FACULTY
Department of Food Science

PhD on **Food Science** AGR/15 (XIX Cycle)
Coordinator: **Prof. Claudio Cavani**

APPLICATION OF NUCLEAR MAGNETIC RESONANCE SPECTROSCOPY FOR THE EVALUATION OF FOLDING VARIABILITY IN CALCIUM BINDING PROTEINS AND ITS IMPLICATIONS IN FOOD ALLERGIES

Final Examination

PhD student:
Dr. Federica Casadei

Relator:
Dr. Mauro Andrea Cremonini

Corelators:
Prof. Francesco Capozzi
Dr. Elena Babini

Bologna 2007

“ to my family ”

INDEX

PREFACE

Food allergies are widespread in industrialized nations. The reasons of this phenomenon have been attributed to different factors: the consumption of new products, the evolution of food production techniques and the increase of the sensitivity to pollens' allergens that has brought an increase of food cross-reactions. Conduct studies document a 1-2% incidence of food allergy in the adult and 2-8% in the child (Pizzin G. et al., 2003). However, this reference value seems to be underestimated for an objective diagnostic difficulty. The scientific community has paid attention to this subject, that only recently has been recognized as a public healthiness problem, and currently the food allergies are object of normative activity at a communitarian level and extra communitarian too.

On March 2005 a trans-sectorial European research project, called EuroPrevall ("The Prevalence, Cost and Basis of Food Allergy across Europe") has been financed. The project' aim is to deliver the information and tools necessary for policy makers, regulators and the food industry to effectively manage food allergies across Europe and hence deliver an improved quality of life to food allergic consumers. Improved quality of life, including reduced risk, will be delivered by integrating information and developing tools for use by European food allergy scientists, health professionals, food and biotech industries, and consumers those who are food allergic and those who are not. One of the project objective concern the structural determination of allergens to provide a library (collection) of highly characterized, well defined food allergens which will allow the development of component resolved diagnosis, to developed novel methods with improved diagnostic power for food allergy. So, food sciences have to face the technology troubles concerning the protein structure characterization which can't be simply identified with the amminoacidic sequence. In fact, it's known that the BSE ezyologic origin is imputable not to a protein mutation but to a structural alteration which diffuses to the naturally presented molecules in healthy bovines (Chakraborty et al., 2005; Concepcion et al., 2005). Other studies have highlighted that certain molecular associations between milk proteins (α-lactalbumin) and unsatured fatty acids (C18:1) cause alterations in the protein structure that induce cellular death (Svensson et al., 2000). For this reason it is necessary to pay attention in the formation of "food experts" to improve their "familiarity" with techniques and approaches concerning the proteins structure characterization.

This last necessity is strictly connected with the developed PhD activity, whose aim is the study of the existing relations between the sequence of potential allergenic proteins and their

structure, easily connectable with the biologic activity. The application field has been focused on the calcium binding proteins, and in particular on those belonging to the EF-hand superfamily, one of the principal classes of calcium-binding macromolecules causing food allergies (i.e, parvalbumin and psoriasin).

The first identified calcium-allergen is the fish parvalbumin, that contains three EF-hand motives, one of which silent. Since then many calcium allergens have been described with different number of EF-hand. In many cases, it has been demonstrated that the loop-calcium bond induces conformational changes that causes the recognition of the IgE-allergen. These results indicate that the interaction of these proteins with the physiological calcium plays an important role in the biologic and allergenic activity (Hebenstreit D. et al., 2005).

The magnetic nuclear resonance (NMR) is used to value also small structural changes determined by different environmental conditions (temperature, pH and ionic concentration), also in relation to the peculiarities concerning the different specific sequence of each organism and then of the derived food. In fact, the real allergenic action is determined by the presence of specific protein surface regions, called epitopes, which could result deeply altered by the interaction of these macromolecules with ions or other molecules present in the environment. So, still once, the only information relative to the amminoacidic sequence results insufficient for this of proteins class that can undergo dramatic open-closed conformational changes depending on the interaction with the physiological targets.

The NMR study requires millimolar protein concentrations, correspondent to milligrams of protein, not easily obtainable by the direct extraction from natural matrixes. The alternative way is the protein over-expression by opportunely recombined bacteria and optimizing the extraction and purification conditions to obtain molecules with the same structural proprieties of the native protein. This strategy also allows to obtain protein samples enriched with magnetically active isotopes necessary to resolve the soluble structure.

The gene selection has been based on some protein requirements: 1) the protein, with unknown structure, must contain the EF-hand motive; 2) the protein must bind calcium; 3) the protein must be flexible so as to alter its structure with calcium-binding.

Additionally, to increase the number of soluble samples, many different cloning and expression conditions were tested, according to Gateway technology. The Gateway technology is a universal cloning method based on the site-specific recombination proprieties of bacteriophage lambda, that provides a rapid and highly efficient way to move DNA sequences into multiple vector systems for functional analysis and protein expression (Gateway Technology, 2003). The genes coding for selected proteins were recombined into

five different expression vectors (pETG-20A, pETG-30A, pETG-60A, pDEST17, pDEST15) carrying specific protein fusion tags. For proteins expression three different *E.coli* host strains were used (BL21 *Gold, pLys, Codon plus*). The proteins mainly expressed in a soluble form were produced on a large-scale and purified for spectroscopic analysis.

1. INTRODUCTION

1.1 FOOD ALLERGY

Food allergy is an IgE-mediated abnormal response to a normally tolerated food protein. The reasons for an individual to become intolerant towards a specific food protein are unclear. The amount of protein required, the threshold, to elicit an allergic response in a sensitized person varies considerably from patient to patient and protein to protein (Sathe SK et al., 2005).

IgE-response is generated in presence of a target macromolecule. IgE-mediated food allergies consist of two phases, the first involving sensitisation. The second phase occurs on re-exposure to an allergen. In this phase an allergic reaction takes place, causing the release of histamine and other inflammatory mediators, and producing the physiological changes that are manifested as allergic reactions. The spectrum of food allergy symptoms may include flushing, urticaria, angioedema, laryngoedema, diarrhea, nausea/vomiting, bronchospasm, or hypotension (Roux KH, et al, 2005) and in rare instances some individuals manifest anaphylactic reactions.

IgE binding per se is not indicative that an allergen is able to elicit an allergic response. In fact, peanut allergic patients have serum IgE which cross-reacts with soy proteins but generally they don't suffer allergic reaction when they eat foods containing soy.

The sensitisation to the classical food allergens (milk, egg, peanut and fish) occurs via the gastrointestinal tract, while other types of food allergy (fresh fruits and vegetables) occur as a consequence of prior sensitisation to inhaled allergens such as pollen.

The cross-reactive allergy syndromes occurs for when individuals sensitised to an allergen, for exemple birch pollen, develop IgE to the major birch allergen Bet v 1, which also recognises homologous proteins present in fresh fruits and vegetables. In the same way, individuals with latex allergy develop IgE to the chitin-binding domain of the hevein polypeptides which make up the rubber latex network. An almost identical domain is found in the class I chitinases of a number of fruits including avocado, chestnut and banana, which the anti-latex IgE recognises, resulting in allergic reactions to these foods (Mills et al., 2005).

Only some types of foods are responsible for causing the majority of food allergies, including a number of foods of plant origin such as peanuts, tree nuts, wheat and soy, together with allergens of animal origin including cow's milk, egg, fish and shellfish (Bush et al.,1996).

It has been estimated that food allergies are most prevalent during infancy, affecting up to 6% of young children (Breiteneder H et al., 2005).

Although sensitivity to most food allergens such as milk, wheat, and eggs tends to remit in late childhood, persistence of certain food allergies such as peanut, tree nut, and seafood most commonly continues throughout one's lifetime. The most common food allergens in children in the United States are egg, milk, peanut, soy, and wheat. In a prospective study of adverse reactions to foods in infants, 80% of confirmed symptoms developed in the first year of life. Sensitivity of some food allergens (especially cow's milk, wheat, and egg) tends to remit in late childhood. For example, most infants who are sensitive to cow's milk lose their sensitivity by 2 years of age. However, persistence of childhood food allergies is common with certain foods, especially to peanuts, tree nuts, and seafood. Children diagnosed as having food allergy after 3 years of age are less likely to lose this sensitivity. Furthermore, children who develop one IgE-mediated food allergy have an increased risk of developing allergies to other foods and inhalant allergens (Bernstein IL et al., 2006).

1.1.1 Allergenic foods of animal origin

The major allergen in **fish** is parvalbumin, which acts as a calcium buffer protein in fast muscle, binding the Ca^{2+} ions by the EF-hand calcium binding motif. Parvalbumin is widely conserved across fish species (cod, salmon, mackerel, herring and plaice). This aspect is responsible for the cross-reactive nature of allergens. The cod allergen, Gad c 1, possesses five IgE-binding regions evenly distributed along the length of the protein, one of which encompassed one of the Ca^{2+} binding sites. Like other calcium binding proteins Gad c 1 is heat-stable, with the holo-form being both more IgE-reactive and more heat-stable than the apo form (Bugajska-Schretter A et al.,2000; Poulsen et al.,2001).

Cow's **milk** allergy is a common disease of infancy and childhood. Goat's milk cross-reacts with cow's milk. Ninety percent of cow's milk allergic patients will react to goat and/or sheep's milk (Bernstein IL et al., 2006).

Allergens are found in both the casein and whey fractions (h-lactoglobulin, a-lactalbumin).

h-Lactoglobulin is highly resistant to proteolysis. The IgE epitopes have been identified in four main regions located on the more mobile surface loops of the protein. The caseins also appear to contain thermostable epitopes, with IgE binding located in around seven different regions of the protein (Wal, 2002).

There is significant cross-reaction between different species of fish, including salt and fresh water (eg, salmon, trout, perch, carp, and eel) (Bernstein IL et al., 2006).

Ovomucoid (Gal d 1) and ovalbumin (Gal d 2), constituting the 10% and 50% of white proteins respectively, are the major allergens originate from **egg** white. Both proteins are

heavily glycosylated with 25% of the mass of ovomucoid comprising carbohydrate. In ovomucoid IgE epitopes are resistant to enzymic digestion and denaturation. They are also clustered in seven regions of the protein, in the N- and C-terminal regions but do not include glycosylation sites. Ovotransferrin (Gal d 3) and lysozyme (Gal d 4) are two minor allergens, identified in egg white. Cooking has been found to reduce allergenic activity of egg (Poulsen et al., 2001). The major cross-reacting antigens between hen's egg protein and bird dander are Livetins. Various bird egg whites (turkey, duck, goose, and seagull) contain proteins that cross-react with allergens in hen's egg white. Several proteins that cross-react with allergens in hen's egg white are also detected in egg yolk, hen sera, and meat (Bernstein IL et al., 2006).

The major allergen in **seafood and shellfish** is tropomyosin, a heat-stable muscle protein, with highly homologous proteins being found in the commonly edible crustaceans. These homologies cause the cross-reactive allergies observed between various types of seafood including shrimps, lobsters, crab, squid and abalone, and inhalant insect allergens, such as those from cockroaches. Two main linear IgE-binding sites have been identified in the shrimp allergen, Pen i 1, in the N- and C-terminal regions. The first two residues of the C-terminal epitope appear to be crucial for IgE binding and are not found in vertebrate tropomyosin (Mills et al., 2005). There is no IgE crossreactivity between crustacean and animal muscle tropomyosins since the lack of homology in the IgE epitopes. In addition to being found in cooked meat, the allergen also leaches into cooking water (Lehrer et al., 2003).

1.1.2 Allergenic foods of plant origin

Soy and peanut are the major allergenic **legumes**. Major allergens include the 7S (Ara h 1 for peanut, h-conglycinin for soy) and 11S seed storage globulins (Ara h 3 in peanut, glycinin in soy) (Mills et al., 2005). Peanut and soybean are members of the legume family and share several common antigenic fractions. Patients allergic to one of these foods have serum IgE antibodies that immunologically cross-react with other legumes(Bernstein IL et al., 2006).

Wheat and the related **cereals**, barley and rye, contain a range of allergens including the prolamins (alcohol soluble storage proteins) which are responsible for food dependent exercise induced anaphylaxis and atopic dermatitis. Inhibitors of proteases and a-amylases of cereals have also been described as both inhalant (e.g. Baker's asthma) and food allergens.

The 2S albumins and the 7S and 11/12S globulins are the major allergens of **seeds** and **nuts**. In addition to storage protein allergens, chestnut contains an endochitinase allergen Cas s 1 involved in the latex-fruit allergy syndrome (Mills et al., 2005).

Although IgE-mediated reactions to **fruits** and **vegetables** are commonly reported, clinically relevant cross-reactivity resulting in severe reactions is uncommon. The latex-fruit syndrome is the result of cross-reactivity between natural rubber latex proteins and fruit proteins. The most commonly reported cross-reactive foods include banana, avocado, kiwi, and chestnut, but many other fruits and some nuts have been identified in cross-reactivity studies. Reactions are typically midl and related to initial sensitization to pollens that share homologous proteins with the implicated fruits and vegetables; this is known as the pollen-food syndrome or oral allergy syndrome (Bernstein IL et al., 2006).

1.1.3 Food allergies and the relevance for industrial proteins

The 8 types of foods originally identified by Bush (Bush et al.,1996) as responsible for the majority of allergies represent important sources of industrial proteins. They could be involved in the production of foods but also in other products such as cosmetics. For this reason in the industrialized countries, the necessisty to regulate the diffusion of products containing allergenic proteins is born, particularly in the wake of recent legislation regarding labelling of allergens in foods.

Following on from the FAO–WHO Codex Alimentarius Commission amendment to the Codex General Standard for the Labelling of Prepackaged Foods, the European Union, the USA and other countries have been developing the necessary regulations to ensure that allergenic foods, or ingredients, are labelled on prepackaged foods (Mills et al., 2004)..

There are also associated concerns about so-called cross-contact allergens which may arise from use of common or adjacent processing lines to prepare foods that do and do not contain any allergenic ingredients. Given that a kiss from a person eating peanuts given to a severely peanut allergic child can be enough to trigger a severe reaction, this represents a considerable hazard for allergic consumers which is frequently “hidden”. Whilst manufacturers may protect themselves by placing “may contain” labels to deal with cross-contact allergens in foods, such a strategy runs the risk of devaluing labels and undermining consumers trust in food manufacturers. The lack of knowledge on how much of an allergen causes a problem, and the development of associated analytical methodology with appropriate performance criteria and reference materials are also issues that will have to be met in the coming years. Awareness of the problems facing allergic consumers, good manufacturing practices to manage allergenic ingredients coupled with an effective dialogue with allergy sufferers, including good quality information on products, will undoubtedly help in managing the apparently rising tide of food allergies we are faced with.

The number of additives used by the food industry is extensive. Only a small number of additives have been implicated in IgE-mediated or other (immunologic or nonimmunologic) adverse reactions. Adverse reactions to food additives, therefore, are rare (Bernstein IL et al., 2006).

Major allergenic foods listed in annex IIIa of the EU directive on labelling of foods
Cereals containing gluten (i.e. wheat, rye, barley, oats, spelt or their hybridized strains), and products thereof
Crustaceans and products thereof
Eggs and products thereof
Fish and products thereof
Peanuts and products thereof
Soybeans and products thereof
Milk and products thereof (including lactose)
Nuts i.e. Almond, Hazelnut, Walnut, Cashew, Pecan nut, Brazil nut, Pistachio nut, Macadamia nut and Queensland nut and products thereof
Celery and products thereof
Mustard and products thereof
Sesame seeds and products thereof
Sulphur dioxide and sulphites at concentrations of more than 10 mg/kg or 10 mg/l expressed as SO_2

Tab. 1.1 List of the allergenic foods (or derived ingredients) which will have to be labelled as allergens in accordance with the EU directive (Mills et al., 2005).

1.2 STRUCTURAL BIOLOGY OF ALLERGENS

Immunogenicity and *cross-reactivity* are the two aspects which have to be distinguished in predicting the allergenic potential of a protein, particularly in novel foods. Immunogenicity represents the potential of a protein to induce IgE antibodies, while cross-reactivity is the reactivity of (usually preexisting) IgE antibodies with the target protein. Moreover, the relation between IgE-binding potential and clinical symptoms is of interest and is influenced by physical properties (eg, stability and size) and immunologic properties (affinity and epitope valence). Discussions on immunogenicity and cross-reactivity of allergens rely on the establishment of structural similarities and differences among allergens and between allergens and non allergens (Aalberse RC, 2000).

1.2.1 Determinants of allergenicity

The *solubility*, *stability*, *size*, and the *compactness of the overall fold* are relevant aspects of protein structure for its allergenicity. These aspects reflect dependency of allergenicity on transport over mucosal barriers and susceptibility to proteases. Size and solubility of the intact protein would be relevant factors for airborne allergens more than for food allergens (for which limited proteolysis might enhance mucosal transport and hence allergenicity) or parenteral antigens, such as insect venoms, insect salivary allergens, invasive organisms (helminths and fungi), vaccines, or therapeutic proteins.

The allergenicity can be affected by post-translational modification in different ways. A post-translational modification may influence solubility, stability, size, and susceptibility toward proteases and it may induce new epitopes. Moreover, uptake and processing by antigen-presenting cells are also known to be markedly influenced. Although glycosylation affects many of these processes, it is not a critical factor for allergenicity in general. Many allergens are not glycosylated, whereas some important allergens (eg, Gal d 1 ovomucoid) are heavily glycosylated. It will become clear from this overview that few, if any, structural features are currently known to be common for allergens in general, even though most allergens can be grouped into a small number of structural classes. Some of these classes have been suggested to be intrinsically more allergenic, for example, 2S albumin from seeds and lipocalin. It is, however, likely that features other than structure are more relevant for allergenicity. In addition, the search for common structural features relevant for allergenicity will become more relevant. It is not realistic to assume that the requirements for sensitization to food allergens in early childhood are very similar to those for late-onset sensitization to airborne occupational allergens (Aalberse RC, 2000).

1.2.2 Determinants of cross-reactivity

Cross-reactivity is largely determined by structural aspects: two proteins are crossreactive only (almost) if they share structural features, in contrast to allergenicity.

There are some exceptions to this rule. Antibody affinity is an important consideration: low affinity antibodies (particularly of the IgM class) have been found to be reactive with antigens with very little structural similarity. However, IgE-producing plasma cells evolved from a strongly T cell–dependent immune reaction in which only high-affinity B cells survive. Moreover, IgE-mediated cellular reactions are triggered by trace amounts of allergens. However, that information on the threshold of the affinity required for triggering allergic reactions by using cross-reactive allergens is still lacking. Another apparent exception is the cross-reactivity between glycoprotein allergens with unrelated protein structures (Aalberse et al., 1997). In this situation cross-reactivity is due to shared glycans and thus similarity in structure. Apart from these exceptions, all IgE cross-reactions described so far have been found to reflect shared features on the level of both primary and tertiary structure of the cross-reactive proteins. Whereas all cross-reactive proteins have a similar fold, the reverse is not true: proteins with a similar fold are not necessarily cross-reactive. This is partially due to immunologic tolerance induced by autologous proteins with a similar fold. More importantly, protein folding is liberal with respect to single amino acid substitutions for many positions in the sequence. Such substitutions may markedly affect the outer protein surface and thus reduce antibody reactivity.

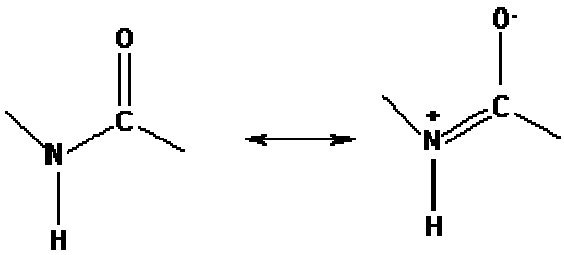
Similar protein folds are found with as little as 25% amino acid identity, whereas cross-reactivity is rare below 50% identity. In most situations cross-reactivity requires more than 70% identity.

The likelihood of cross-reactivity depends not only on the similarity between the primary sequences of the two antigens. If a human homologue exists and is available to the immune system, this is likely to induce immune tolerance to the common epitopes. In some instances tolerance to the human protein appears to be absent, resulting in autoreactive IgE antibodies. The first description of such an autoreactive IgE antibody presumably induced by contact with a cross-reactive environmental allergen was the autoreactivity described for human profilin in patients with pollinosis (Bettelheim P et al., 1991). Another striking example is the reactivity of IgE antibodies induced by the fungal allergen Asp f 6 (manganese-dependent superoxide dismutase [MnSOD]). The recombinant human protein was found to give strongly positive skin reactions in subjects with reactivity to the *Aspergillus* protein.

To *predict potential cross-reactivity* of a new allergen (eg, an engineered food protein), it is informative to compare protein folds. In the absence of similarity in folding with allergens, protein cross-reactivity is virtually excluded (apart from the above-mentioned exceptions).
If similarity in folding is observed, cross-reactivity needs to be investigated. The knowledge of the folding family allows a search restricted to subjects with sensitization to allergens with that particular protein fold. However, antibodies are highly idiotypical. Therefore absence of crossreactivity of one antibody (or even many antibodies) does not exclude cross-reactivity of another antibody, even with the same epitope specificity. As with all risk assessments, statistical evaluation on the basis of a sufficiently large number of cases is important.

1.2.3 Protein structure

Protein structure can be described at different levels: primary structure (ie, the amino acid sequence), protein fold, domain structure, and surface structure. The surface structure is the most relevant for antibody binding, particularly the epitope, which is that part of the surface that on an atomic level interacts with the antibody. The surface structure was given by the protein fold that is the results of the interaction of the secondary structure elements.
The primary structure of a segment of a polypeptide chain or of a protein is the amino-acid sequence of the polypeptide chain(s), without regard to spatial arrangement (apart from configuration at the alpha-carbon atom). The commonly occurring amino acids are of 20 different kinds which contain the same dipolar ion group $H_3N^+.CH.COO^-$. They all have in common a central carbon atom to which are attached a hydrogen atom, an amino group (NH_2) and a carboxyl group (COOH). The central carbon atom is called the C_{alpha}-atom and is a chiral centre. All amino acids found in proteins encoded by the genome have the L-configuration at this chiral centre.
There are 20 side chains found in proteins encoded by the genetic machinery of the cell. The side chains confer important properties on a protein such as the ability to bind ligands and catalyse biochemical reactions. They also direct the folding of the nascent polypeptide and stabilise its final conformation. Amino acids in proteins (or polypeptides) are joined together by peptide bonds. The sequence of R-groups along the chain is called the **primary structure.**
The C-N bond length of the peptide is 10% shorter than that found in usual C-N amine bonds. This is because the peptide bond has some double bond character (40%) due to resonance which occurs with amides. The two canonical structures are showed below.
As a consequence of this resonance all peptide bonds in protein structures are found to be almost planar, i.e. atoms C_{alpha}, C, O, N and $C_{alpha+1}$ are approximately co-planar.



This rigidity of the peptide bond reduces the degrees of freedom of the polypeptide during folding. The peptide bond nearly always has the *trans* configuration since it is more favourable than *cis*, which is sometimes found to occur with proline (Pro) residues (Schirmer et al., 1990). Steric hindrance between the functional groups attached to the C_{alpha} atoms will be greater in the *cis* configuration. However, for proline residues, the cyclic nature of the side chain means that both *cis* and *trans* configurations have more equivalent energies. Thus proline is found in the *cis* configuration more frequently than the other amino acids. The sequence and properties of side chains determine all that is unique about a particular protein, including its biological function and its specific three-dimensional structure.

The rotation around bounds N–C_α e C_α–C_o is permitted. The rotation angles φ (phi) and ψ (psi) can assume values from -180° to +180°, but many of these can't be formed for the steric interactions between atoms of the polypeptidic and lateral chains. For this reason the conformation's value zero of φ and ψ is forbidden, but it is used like a reference point only.

Every possible secondary structure is completely described by φ e ψ angles (Lehninger *et al.*, 1994). The permitted values of φ and ψ are indicated on a two-dimensional map of the φ - ψ plane.

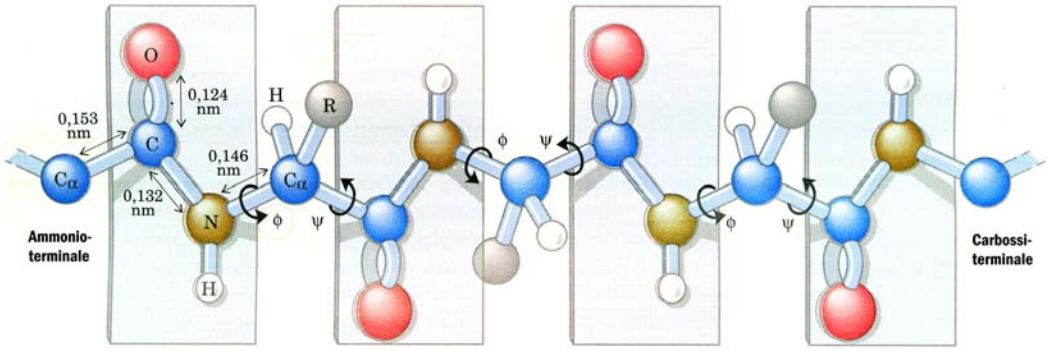


Fig.1.1 Representation of the possible rotation movements in the polypeptidic chain. The C-N bonds of the planar peptidic groups aren't free to rotate. The possible obstacled rotation of the other skeleton's single bonds depends on the charge and dimension of R groups (Lehninger *et al.*, 1994).

The **secondary structure** of a segment of polypeptide chain is the local spatial arrangement of its main-chain atoms without regard to the conformation of its side chains or to its relationship with other segments. There are three common secondary structures in proteins: alpha helices, beta sheets and turns. Regular secondary structure conformations in segments of a polypeptide chain occur when all the φ bond angles in that polypeptide segment are equal to each other, and all the ψ bond angles are equal.

The **alpha-helix** and **beta-structure** conformations for polypeptide chains are generally the most thermodynamically stable of the regular secondary structures. However, particular amino acid sequences of a primary structure in a protein may support regular conformations of the polypeptide chain other than alpha-helical or beta-structure. Thus, whereas alpha-helical or beta-structure are found most commonly, the actual conformation is dependent on the particular physical properties generated by the sequence present in the polypeptide chain and the solution conditions in which the protein is dissolved. In addition, in most proteins there are significant regions of unordered structure in which the φ and ψ angles are not equal. A large proportion of (85%) of helices are distorted in some way i.e. radius of curvature greater than 90Å and deviation of axis from straight line is equal to or greater than 0.25Å. These may be due to a number of reasons:

- CO groups form hydrogen bonds with NH groups 3 residues along the chain forming a 3_{10} helix. A substantial amount of all 3_{10} helices occur at the ends of alpha-helices. They are called 3_{10} because there are 3 residues per turn and 10 atoms enclosed in a ring formed by each hydrogen bond. Dipoles are not aligned as in a normal right-handed alpha-helix.
- Packing of buried helices against other secondary structural elements in the core of a protein can lead to distortions since the side chains are on the surface of helices.
- Proline residues induce distortions of around 20deg in the direction of a helix. This is because proline cannot form a regular alpha-helix due to steric hindrance arising from its cyclic sidechain which blocks the main chain NH group. Proline causes 2 hydrogen bonds in the helix to be broken. Helices containing proline are usually long because shorter helices would be destabilized.
- Exposed helices are often bent away from the solvent. This is because the exposed C=O groups tend to point towards solvent to maximize their hydrogen bonding capacity i.e tend to form hydrogen-bonds to solvent as well as N-H groups. This gives rise to a bend in the helix axis.

- The pi helix is an extremely rare secondary structural element in proteins. Like the 3_{10} helix, one turn of pi helix is sometimes found at the ends of regular alpha helices. The infrequency of this particular form of secondary structure stems from the following properties: (i) the phi and psi angles of the pure pi helix (-57.1, -69.7) lie at the very edge of an allowed minimum energy region of the Ramachandran map, (ii) the pi helix requires that the angle tau (N-C_{alpha}-C') be larger (114.9) than the standard tetrahedral angle of 109.5degrees, (iii) the large radius of the pi helix means the polypeptide backbone is no longer in van der Waals contact across the helical axis forming an axial hole too small for solvent water to fill, and (iv) side chains are more staggered than the ideal 3_{10} helix but not as well as the alpha helix (Schirmer et al., 1990).

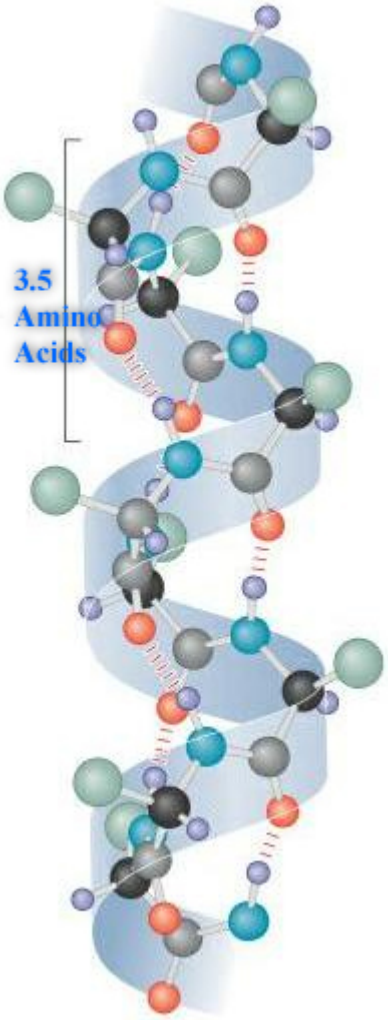


Fig.1.2 Representation of alpha-helix one of the structural element of the regular secondary structures in globular proteins.

Besides the alpha-helix, beta-sheets are another major structural element in globular proteins containing 20-28% of all residues. The basic unit of a beta-sheet is a beta strand (which can be thought of as a helix with n=2 residues/turn) with approximate backbone dihedral angles phi = -120 and psi = +120 producing a translation of 3.2 to 3.4 Å / residue for residues in antiparallel and parallel strands, respectively. The beta strand is then like the alpha-helix, a repeating secondary structure. However, since there are no intra-segment hydrogen bonds and van der Waals interactions between atoms of neighbouring residues are not significant due to the extended nature of the chain, this extended conformation is only stable as part of a beta-sheet where contributions from hydrogen bonds and van der Waals interactions between aligned strands exert a stabilizing influence. The beta-sheet is sometimes called the beta pleated sheet since sequential neighbouring C_{alpha} atoms are alternately above and below the plane of the sheet giving a pleated appearance. beta-sheets are found in two forms designated as "Antiparallel" or "Parallel" based on the relative directions of two interacting beta strand as showed in figure 1.3.

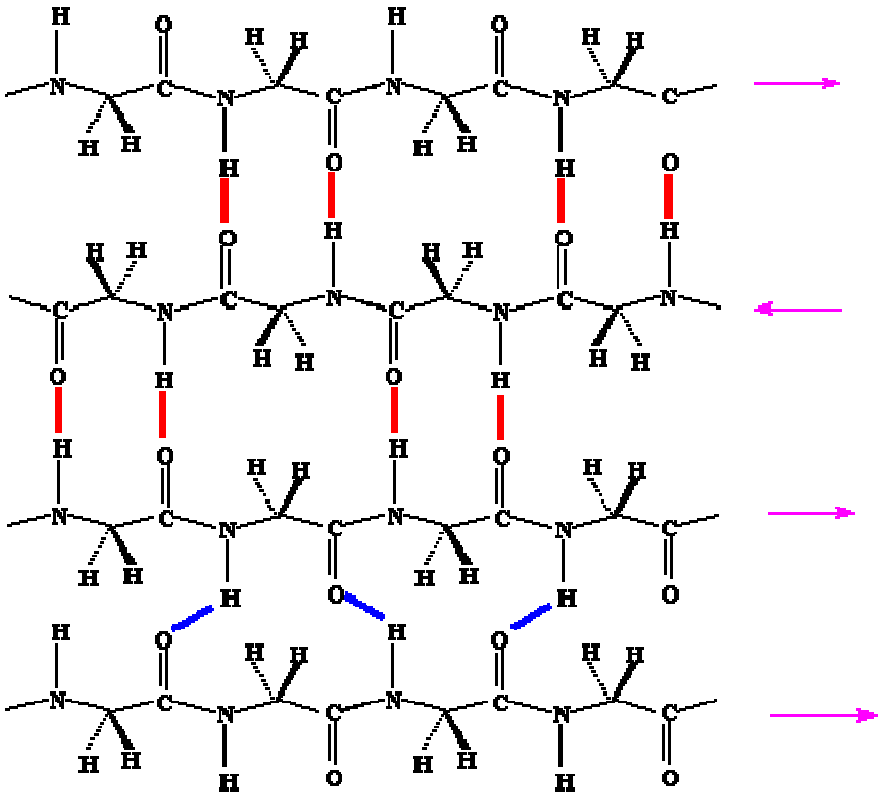


Fig.1.3 Rapresentation of hydrogen bond patterns in Antiparallel (red lines) or Parallel (blue lines) beta-sheets.

Hydrogen bond patterns in beta-sheets. A four-stranded beta-sheet is drawn schematically which contains three antiparallel and one parallel strand. Hydrogen bonds are indicated with red lines (antiparallel strands) and blue lines (parallel strands) connecting the hydrogen and receptor oxygen. Like alpha-helices, beta-sheets have the potential for amphiphilicity with one face polar and the other apolar. However, unlike alpha-helices which are composed of residues from a continuous polypeptide segment, beta-sheets are formed from strands that are very often from distant portions of the polypeptide sequence. Hydrogen bonds in beta-sheets are on average 0.1 Å shorter than those found in alpha-helices (Baker and Hubbard, 1984). The classical beta-sheets originally proposed are planar but most sheets observed in globular proteins are twisted (0 to 30 degrees per residue).

Antiparallel beta-sheets are more often twisted than parallel sheets. Another irregularity found in antiparallel beta-sheets is the hydrogen-bonding of two residues from one strand with one residue from another called a beta bulge (as shown above).

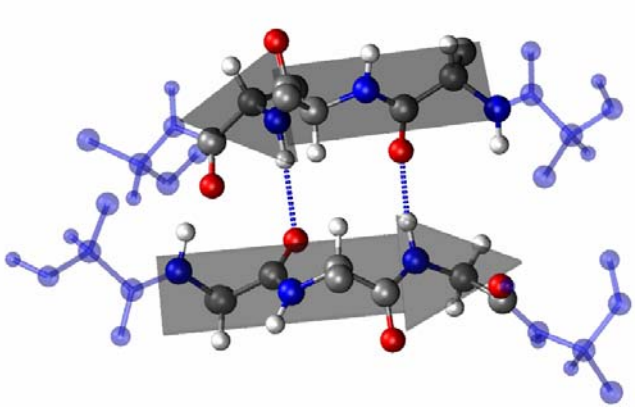


Fig.1.4 Atomistic illustration of two beta strands in an antiparallel beta hairpin motif. A very simple structural motif involving β sheets is the β hairpin, in which two antiparallel strands are linked by a short loop of two to five residues, of which one is frequently a glycine or a proline, both of which can assume the unusual dihedral-angle conformations required for a tight turn. However, individual strands can also be linked in more elaborate ways with long loops that may contain alpha helices or even entire protein domains.

Turns are the third of the three "classical" secondary structures that serve to reverse the direction of the polypeptide chain. They are located primarily on the protein surface and accordingly contain polar and charged residues. Turns may be viewed as a weak link in the polypeptide chain, allowing the other secondary structures (helix and sheet) to determine the conformational outcome. There were structural components comprising a few alpha-helices or beta-strands which were frequently repeated within structures, called "**supersecondary structures**" (being intermediate to secondary and tertiary structure) and suggested that these structures might be due to convergence. A variety of recurring structures were subsequently recognised such as the "Greek key" (Richardson, 1977). Some of these **structural motifs** can

be associated with a particular function whilst others have no specific biological function alone but are part of larger structural and functional assemblies. proteins containing mostly alpha-helix. The simplest motif with a specific function consists of two alpha-helices joined by a loop region. Two such motifs are (i) a motif specific for DNA binding and (ii) a motif specific for calcium binding and is present in parvalbumin, calmodulin, troponin-C, and other proteins that bind calcium and thereby regulate cellular activities.

The **tertiary structure** of a protein molecule, or of a subunit of a protein molecule, is the arrangement of all its atoms in space, without regard to its relationship with neighbouring molecules or subunits. As an example of a protein's tertiary structure, the structure of lysozyme is shown below. The **quaternary structure** of a protein molecule is the arrangement of its subunits in space and the ensemble of its inter-subunit contacts and interactions, without regard to the internal geometry of the subunits. The subunits in a quaternary structure must be in no-covalent association. Haemoglobin contains four polypeptide chains ($alpha_2b_2$) held together no-covalently in a specific conformation as required for its function. To be biologically active, proteins must adopt specific folded three-dimensional, tertiary structures. Yet the genetic information for the protein specifies only the primary structure, that is the linear sequence of amino acids in the polypeptide backbone. Many purified proteins can spontaneously refold *in vitro* after being completely unfolded, so the three-dimensional structure must be determined by the primary structure (Anfinsen, 1973). Different conformations of a protein differ only in the angle of rotation about the bonds of the backbone and amino acid side-chains. It may therefore, appear surprising that a protein folds into a single unique conformation from all the possible rotational conformations available around single bonds in the primary structure of a protein. The question arises as to when a protein folds up to its native conformation, does this structure represent a local or a global energy minimum? When a protein folds it samples a number of conformations. Does the structure which results from the folding depend on its stability or on the energy barriers encountered by the polypeptide? The polypeptide whilst folding may become trapped in the local energy well and cannot fold to the global energy minimum (Wetlaufer DB, 1973).

The amino acid sequence deduced from the nucleotide sequence is usually not the complete and final primary structure because cotranslational and posttranslational modifications are common. Most allergens are extracellular proteins. Secretion of a protein by a cell requires the presence of an N-terminal leader peptide for transport of the protein through the membrane of the endoplasmic reticulum (ER). After in vitro unfolding, most proteins will fold at least partially back into their native structure. Therefore the primary sequence contains

all the clues for the 3-dimensional structure, and we should be able to predict the 3-dimensional structure of the protein. Homology modeling proves to be reliable for sequences that can be aligned well. With sequence homology as low as 25%, overall backbone folds can be predicted often with high accuracy, with the exception of some of the more complicated loops. Configurations of solvent-exposed side chains are far more difficult to predict, and therefore even if the 3-dimensional structure of two proteins with similar folds is known, prediction of cross-reactivity is not yet possible. Fold recognition is, however, an important help in the search for potential cross-reactivity. Fold recognition on the basis of the primary sequence is dependent on the classification of folds, which can be automated to a certain extent, but remains to some degree a subjective process (Aalberse RC, 2000).

1.2.4 Linear and conformational epitopes

The part of the surface of the allergen that interacts with the Fab fragment of the antibody (ie, the epitope) is of particular interest. The atomic details of the interaction between antibody and antigen are known for more than 30 antibody-antigen combinations. The interaction between two proteins is usually described in terms of buried surface (ie, the surface area that becomes inaccessible to water molecules as a consequence of the formation of the immune complex). This value is calculated from crystallographic data by using a spherical probe the size of a water molecule. The review by Padlan (Padlan EA, 1996) gives a range of 54.0 to 89.0 nm2 for the immune complexes with intact protein antigens (as opposed to peptides or haptens) that were available at that time. For a typical globular 20-kd allergen, the radius is approximately 2.0 nm, with a spherical surface area of 500.0 nm2. The water-accessible surface area varies depending on the compactness of the folded structure structure but will usually be approximately twice this value or 1000.0 nm2 in this example. Thus the interaction with a Fab fragment of an antibody will bury approximately 5% to 10% of the surface of such a 20-kd allergen. The surface area that is accessible to a water molecule is obviously larger than the surface area accessible to a Fab fragment (and even more so for a complete antibody). It is therefore likely that a single allergen of this size cannot accommodate more than 5 to 10 antibodies at the same time. Before crystallographic information was available, information on the structure of epitopes was derived, among others, from the reactivity of the antibody with peptides derived from the linear sequence of the allergen or with the unfolded allergen on an immunoblot. On the basis of this type of information, epitopes were classified as either linear or conformational. An antibody was assumed to react with a linear epitope if it was reactive with a small peptide (usually 8-15 amino acids) derived from the linear sequence

or if it was reactive (in an immunoblot) after the allergen had been unfolded by boiling in SDS in the presence of a reducing agent. However, from the crystallographic studies, a general picture emerges that does not support the dichotomy of linear versus conformational epitopes: all protein epitopes are conformational. In all examples studied this buried surface area involves amino acids from different parts of the linear sequence; that is, all these epitopes are conformational to some extent. Figure 1.5 shows the epitope of lysozyme for mAb D1.3 (Bentley GA, et al., 1991).

How can this observation be reconciled with the observation that many antibodies react with small peptides derived from the linear sequence of the antigen? The main factor is the huge difference in affinity. On a molar basis, the peptides are very inefficient compared with the intact antigen, inhibiting the interaction between antigen and antibody. The peptide represents a fraction of the epitope, and the strength of the interaction with the antibody is a fraction of the strength with the complete antigen. The strength of interaction with the peptide is usually decreased even more because of the higher flexibility of the free peptide compared with the peptide in the complete antigen, resulting in a larger loss of entropy on binding (Aalberse RC, 2000).

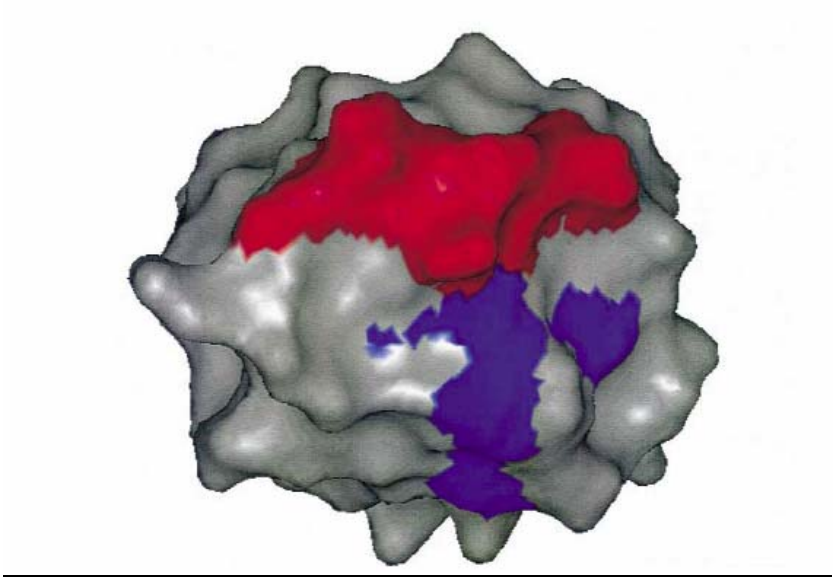


Fig. 1.5 The epitope of egg white lysozyme as recognized by mAb D1.3 (PDB code 1FDL).31 The amino acids that contribute to the N-terminal part of the epitope (18, 19, 22, 23, 24, and 27) are indicated in blue. The amino acids that contribute to the C-terminal part of the epitope (116-121, 124, and 125) are indicated in red (Bentley GA, et al., 1991).

1.3 CALCIUM-BINDING ALLERGENS

1.3.1 Calcium binding proteins

Calcium is an essential component of bone. It is also required for many intracellular functions. Maintenance of calcium homeostasis is a dynamic process involving calcium absorption and excretion in the intestine, filtration and reabsorption in the kidney, and storage and mobilization in the skeleton. Nearly all of the body calcium (98%) is stored in slowly exchangeable skeletal hydroxyapatite crystals. It is the rapidly exchangeable calcium in recently deposited bone and in the extracellular, intracellular, and vascular spaces (2% of body calcium) that plays a critical role in intracellular communication, synthesis and release of neurotransmitters and hormones, muscle contraction, development, and differentiation, as well as aging and apoptosis. Therefore it is not surprising that a number of diseases in children and adults are associated with altered Ca^{2+}-homeostasis.

Ca^{2+}-levels are controlled by release into, and removal from, the cytoplasm (Figure 1.6). Ca^{2+}-pumps in the plasma membrane and endoplasmic reticulum (ER; the Ca^{2+}-store in a cell) keep cytoplasmic Ca^{2+}-levels low (about 0.1 mmol/L in resting cells) and generate a 10,000-fold concentration gradient across membranes (because extracellular Ca^{2+} is in the millimolar range). On stimulation Ca^{2+} enters the cytosol of the cell via Ca^{2+}-channels (plasma membrane) or via Ca^{2+}-channels in the ER, leading to the activation of a great variety of Ca^{2+}-dependent processes in the cell.

A large group of diseases is associated with defects in Ca^{2+}-channels. After entering the cell, Ca^{2+} is reversibly complexed to specific Ca^{2+}-binding proteins that fulfill multiple functions, including Ca^{2+}-buffering and transport, activation of enzymes, regulation of contraction, secretion, and proliferation (Figure 1.6) (Heizmann CW, 2005).

The majority of proteins involved in transducing intracellular Ca^{2+} signals are characterized by a common helix-loop-helix structural motif in their Ca^{2+}-binding sites, termed the EF-hand (Chazin WJ et al, 1998).

A large family of Ca-binding proteins is characterized by a common structural element, the EF-hand motif (Figure 1.7). These proteins decipher the information carried by Ca^{2+} and pass it on to various targets regulating biological functions (Heizmann CW, 2005).

The canonical EF-hand consists of a 29 residue contiguous polypeptide containing helix I (helix E), a loop around the calcium ion, and helix II (helix F). Residues 1, 3, 5, 7, 9 and 12 of the canonical binding loop chelate the Ca^{2+} ion, forming a pentagonal bipyramidal array of seven oxygen ligands. Residues 1, 3 and 5 provide monodentate oxygen ligands via sidechain oxygens, usually aspartate carboxylates. Residue 12 is a bidentate oxygen ligand, and is

almost always a glutamate residue, which ligates calcium via both sidechain carboxlyate oxygens. Residue 7 directly coordinates Ca^{2+} via its mainchain oxygen. Residue 9 hydrogen bonds to a water molecule that provides the remaining Ca^{2+} ligand. Most EF-hand proteins use this canonical motif. However, members of the S100 subfamily of CaBPs have a two residue insertion and a modified coordination scheme in the first binding loop, creating a motif termed the pseudo EF-hand (Chazin WJ et al, 1998).

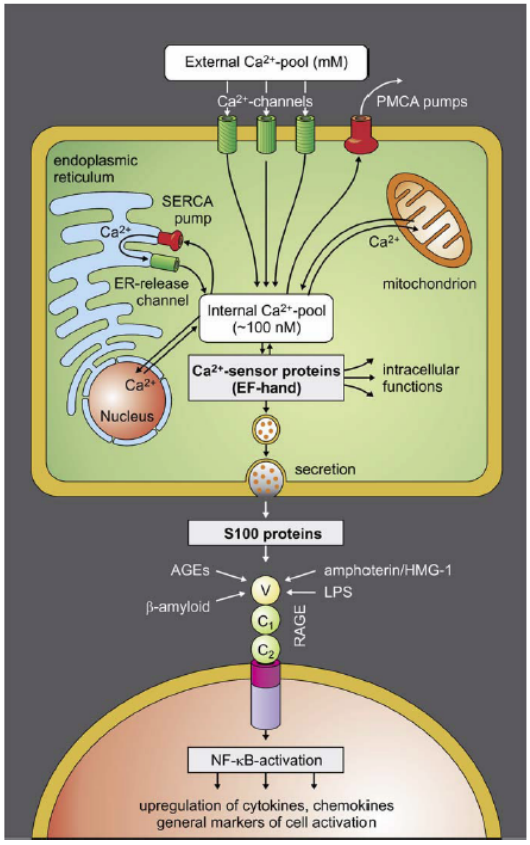


Fig. 1.6 Pathways of Ca^{2+} release and uptake. Three types of Ca^{2+}-channels (gated by ligands, voltage, emptying of Ca21-stores) mediate Ca^{2+}-entry into cells and 2 Ca^{2+}-pumps of plasma membrane and of endoplasmic reticulum remove ion from cytoplasm. Ca^{2+}-signal is mediated by EF-hand Ca^{2+}- binding proteins (including S100 protein family), which has intra- and extracellular cytokine-like functions (Heizmann CW, 2005).

Proteins of this family are closely associated with a number of human diseases such as cardiomyopathy, diabetes, brain disorders, inflammation, and cancer. They have also been proven to be valuable in the diagnosis of these diseases and as predictive markers of improving clinical management, outcome, and survival of patients. They are considered as having potential as drug targets to improve therapies.

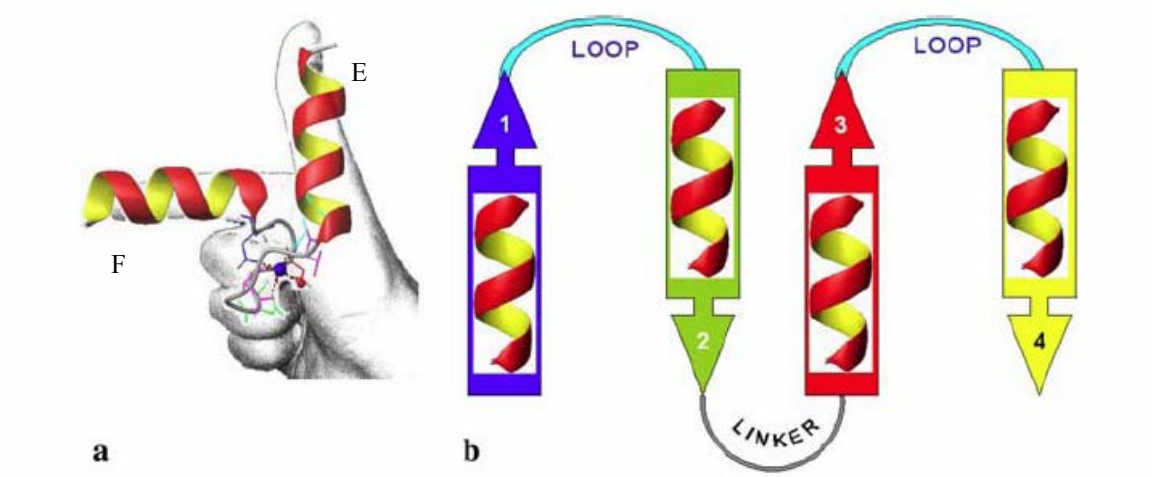


Fig. 1.7 a The EF-hand proteins are based on a motif characterized by a helix–loop–helix structure, with a 12 or 14 amino acid long interhelical loop able to bind a calcium ion. The motif's name is derived from the E and F helices, in the Ca2+ binding site of parvalbumin, resembling the thumb and the forefinger pair of the hand .**b** The minimal structural and functional unit of EF-hand proteins is a domain composed of a pair of EF-hand motifs tethered together by a linker of variable length. Helices are numbered from the N-terminal to the C-terminal of the domain (Capozzi et al., 2006).

A prominent member is troponin (subunits TN-T, TN-I, TN-C), a regulator of striated muscle contraction. Measurements of troponin I levels are routinely used in the diagnosis of myocardial infarction. In addition, mutations in the troponin I subunit are associated with familial hypertrophic cardiomyopathy.

Parvalbumin has been found to be a major fish allergen in human beings. Approximately 70% of fish allergies, mediated by parvalbumin IgE antibodies, lead to respiratory and gastrointestinal symptoms. Parvalbumin also became of interest because of its potential use in gene therapy to correct cardiomyopathy (Heizmann CW, 2005).

1.3.2 Calcium binding allergens

In the last decade, the number of available allergen sequences has rapidly increased due to the introduction of molecular biology techniques for allergen characterization. Sequence analysis of allergen-encoding cDNAs revealed the presence of typical calcium-binding motifs, termed EF-hands, within allergens from various sources. Parvalbumin represented the first calcium-binding allergen described. It represents the major allergen in fish and belongs to a subfamily of closely related calcium-binding proteins that contain two functional and one silent EF-hand motif. In countries with high fish consumption almost 1 per 1000 of the population is allergic to fish with most of the IgE antibodies directed against parvalbumin. Because of sequence and structural similarities among parvalbumins, fish allergic individuals exhibit allergic reactions after contact and/or consumption of various fish species. Next after parvalbumin, EF-hand

domains were identified in the deduced amino acid sequences of two birch pollen allergens and the allergens were subsequently identified as a three and a two EF-hand allergen designated, Bet v 3 and Bet v 4, respectively. It was found that IgE recognition of Bet v 3 was strongly influenced by the presence or absence of protein-bound calcium, a property which was shared by most of the other calcium-binding allergens described later on. Table 1 provides an overview of the calcium-binding allergens described so far. The allergens were grouped according to the number of calcium-binding domains. Allergen sources, molecular weights, accession numbers for the sequences and references are also displayed. As is evident from Table 1, pollens from various plant species turned out to be major sources of calcium-binding allergens. The latter may be related to the fact that calcium ions are important for pollen tube growth and perhaps for the self-incompatibility response in plants.

Many calcium-binding allergens are therefore abundantly expressed in pollen tissue and, due to the fact that pollen from wind pollinated plants becomes airborn in large amounts, can sensitize patients via the respiratory mucosa. In this context it is noteworthy that calcium-binding allergens, although representing intracellular proteins, are rapidly eluted when pollen becomes hydrated as it happens after contact with the mucosa of patient). Bet v 3 from birch pollen was the first described calcium-binding pollen allergen. It was isolated by immunoscreening of a birch pollen cDNA library using serum IgE from a pollen allergic patient. Within its sequence, Bet v 3 contains three typical calcium-binding motifs and was found to be highly expressed in mature pollen. Another family of calcium binding allergens representing two EF-hand proteins of 8-9 kDa was identified in tree-, weed-, and grass pollens (Table 1.2). Jun o 2 and Ole e 8 represent four EF-hand calcium-binding allergens which were isolated from cypress and olive (Valenta R et al., 2000).

Calcium-binding allergens were however also described in parasites, which are potent inducers of IgE responses, and in animal hair/dander. Recently even a calcium-binding autoallergen was isolated using serum IgE from atopic dermatitis patients for screening of a human epithelial expression cDNA library.

No	Allergen	Species - Scientific Name	Species - Common Name	Allergen Type	Allergen Description	Class
1	Aln g 4	*Alnus glutinosa*	alder	tree Fagales		non-IUIS
2	Bet v 3	*Betula verrucosa*	birch	tree Fagales		IUIS
3	Bet v 4	*Betula*	birch	tree Fagales		IUIS

		verrucosa				
4	Bos d 3	*Bos domesticus*	domestic cattle	animals	Ca-binding S100 homolog	IUIS
5	Bra n 1	*Brassica napus*	rapeseed	foods	2S albumin	IUIS
6	Bra n 2	*Brassica napus*	rapeseed	foods		non-IUIS
7	Bra r 1	*Brassica rapa*	turnip	foods	2S albumin	IUIS
8	Che a 3	*Chenopodium album*	lamb's-quarters, pigweed, white goosefoo	weed Caryophyllales	polcalcin	IUIS
9	Cyn d 7	*Cynodon dactylon*	Bermuda grass	grass Poales		IUIS
10	Cyp c 1.01	*Cyprinus carpio*	common carp	foods	parvalbumin	non-IUIS
11	Cyp c 1.02	*Cyprinus carpio*	common carp	foods	parvalbumin	non-IUIS
12	Gad c 1	*Gadus callarias*	cod	foods	allergen M, parvalbumin beta	IUIS
13	Gad m 1	*Gadus morhua*	Atlantic cod	foods	parvalbumin beta	non-IUIS
14	Hom s 4	*Homo sapiens*	human autoallergens	others		IUIS
15	Jun o 4	*Juniperus oxycedrus*	prickly juniper	tree Pinales	homolog: calmodulin; old name: Jun o 2	IUIS
16	Ole e 3	*Olea europea*	olive	tree Lamiales Oleaceae		IUIS
17	Ole e 8	*Olea europea*	olive	tree Lamiales Oleaceae	Ca2+-binding protein	IUIS
18	Phl p 7	*Phleum pratense*	timothy	grass Poales		non-IUIS
19	Ran e 1	*Rana esculenta*	edible frog	foods	parvalbumin alpha	IUIS
20	Ran e 2	*Rana esculenta*	edible frog	foods	parvalbumin beta	IUIS
21	Sal s 1	*Salmo salar*	Atlantic salmon	foods	parvalbumin	IUIS
22	Sco j 1	*Scomber japonicus*	chub mackerel	foods	parvalbumin	non-IUIS
23	Syr v 3	*Syringa vulgaris*	lilac	tree Lamiales Oleaceae		non-IUIS
24	The c 1	*Theragra chalcogramma*	Alaska pollock	foods	parvalbumin beta	non-IUIS

Tab. 1.2 Calcium-binding allergens from the Pfam A Family PF00036 (http://fermi.utmb.edu/cgi-bin/SDAP/sdap_07?dB_Type=0&Code=3&Pfam=PF00036).

1.3.3 Calcium-binding proteins as cross-reactive allergens

The classical EF-hand domain found also in calcium-binding allergens consists of an a-helix, a loop coordinating the Ca^{2+} ion and a second a-helix. In the EF-hand, calcium ions are usually bound through four carboxylate or carboxamide groups and a single backbone carbonyl oxygen placed in the loop with a specific spacing. Although EF-hand domains represent highly conserved sequence motifs also in calcium-binding allergens, it seems that crossreactivity of IgE antibodies from allergic patients who are sensitized against a particular calcium-binding allergen is rather limited to closely related proteins. There is extensive IgE crossreactivity of fish allergic individuals with parvalbumins from various fish species which explains why parvalbumin-sensitized individuals exhibit allergic symptoms after ingestion of various fish species (Table 1.2). Likewise, there is strong crossreactivity among the members of the two-EF-hand pollen allergens. Patients who are sensitized to the two-EF hand allergen from birch always exhibit IgE crossreactivity with the homologous proteins in weed and grasspollens and therefore constitute a group of allergic patients who can exhibit allergic reactions after contact with pollens from many, even unrelated plant species. While there is extensive crossreactivity to calcium-binding allergens within certain allergen families (parvalbumins, 2 EF-hand pollen allergens) there seems to be little or no crossreactivity among the various families of calcium-binding allergens in general. The latter fact indicates that the less conserved sequence motifs and, perhaps conformational epitopes which are unique to a given protein family, play an important role for IgE recognition.

Crossreactivities of IgE antibodies within certain families of calciumbinding proteins may have important implications for diagnosis as well as for therapy. If an allergic patient contains for instance IgE antibodies against a particular 2 EF-hand pollen allergen (e.g., birch: Bet v 4) these IgE antibodies will crossreact with 2 EF-hand allergens in grass-, weed- and other tree pollens leading to allergic reaction against most plant pollens. Calcium-binding allergens can thus be used as marker allergens which when used for diagnostic testing allow to predict broad sensitization of a given patient. On the other hand it seems possible to treat sensitized patients by specific immunotherapy only with a few representative members containing the crossreactive epitopes in order to cover the complete familyof crossreactive calcium-binding allergens (Valenta R et al., 2000).

1.3.4 Conformation-dependent IgE recognition of calcium-binding allergens

It is well established that calcium-binding proteins can undergo dramatic open-closed conformational changes depending on the presence or absence of protein-bound calcium. It is

also well established that these conformational changes can strongly affect the physochemical properties of calcium-binding proteins. E.g., it is known that the calcium-bound (open conformations) are more stable than the apoforms (closed conformations) and that the calcium-bound (open conformations) forms can expose hydrophobic surfaces which are not available in the closed conformation (Figure 1.8). Calcium can thus strongly influence the ability of calcium-binding proteins to interact with other ligands and thereby represents a physiological calcium-dependent regulation mechanism for ligand interactions.

Likewise it was discovered that the IgE-binding capacity of calcium-binding allergens is affected by calcium. A calcium-dependent modulation of the IgE-binding capacity was reported for calcium-binding birch pollen allergens, Bet v 3 and Bet v 4, as well as for other calcium-binding pollen allergens and the major fish allergen, parvalbumin. Loss of the protein-bound calcium triggers a large change in conformation and an associated loss of conformation-dependent IgE epitopes (Breiteneder H et al, 2005).

Most sensitized patients contain IgE antibodies which recognize preferentially the calcium-bound form and exhibit strongly reduced IgE binding to the calcium-free apoform. This observation indicates that patients who are allergic against calcium-binding allergens were preferentially sensitized against the calcium-bound forms which are also more stable than the apoforms (Valenta R et al., 2000).

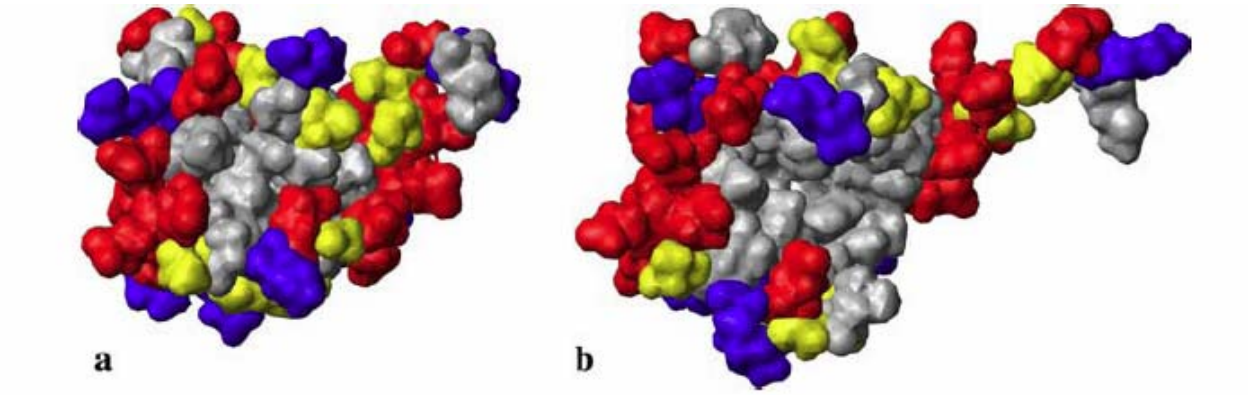


Fig. 1.8 Surface representation of the N-terminal domain of CaM, a calcium sensor protein, in the apo (**a**, 1CMF) and calciumbound (**b**, 1CMG) forms. The surface is colored according to the physicochemical properties of the underlying amino acids: r*ed* acidic, *blu*e basic, *yellow* hydrophilic, and *gray* hydrophobic. Calcium binding to CaM leads to a transition from a closed to an open state, with resulting exposure of a large hydrophobic surface, allowing the protein to interact with targets to accomplish regulatory functions (Capozzi et al., 2006).

1.3.5 Conversion of calcium-binding allergens into hypoallergenic molecules for vaccination against allergy

A major disadvantage of allergen-specific immunotherapy is that administration of allergens to patients can cause anaphylactic side effects. Therefore attempts are made to convert active allergens into molecules with reduced allergenic activity for immunotherapy with reduced rate of anaphylactic side effects. The finding that IgE recognition of calcium-binding allergens depends on protein-bound calcium has stimulated ideas to convert these allergens into allergen variants with reduced IgE binding capacity by mutation of the calcium-binding sites. It was demonstrated that mutated versions of the birch pollen allergen Bet v 4 and the 2 EF-hand allergens from rape containing amino acid exchanges within the calcium-binding domains lost their calcium-binding properties and exhibited a strongly reduced IgE-binding capacity. It was also demonstrated that deletion variants and fragments of calcium-binding allergens exhibited reduced IgE binding capacity. This opens now possibilities to produce by genetic engineering hypoallergenic versions of calcium-binding allergens which may be used for allergen-specific immunotherapy with reduced risk of anaphylactic side effects (Valenta R et al., 2000).

1.4 EuroPREVALL - THE PREVALENCE, COST AND BASIS OF FOOD ALLERGY ACROSS EUROPE

EuroPrevall is a multidisciplinary integrated project (IP) involving 15 European member-states, Bulgaria (a candidate country), Switzerland and Iceland, and Ghana. Of the 53 partners, there are 17 clinical organisations and 11 small-medium sized enterprises (SMEs). Its aim is to deliver the information and tools necessary for policy makers, regulators and the food industry to effectively manage food allergies across Europe and hence deliver an improved quality of life to food allergic consumers. Improved quality of life, including reduced risk, will be delivered by integrating information and developing tools for use by European food allergy scientists, health professionals, food and biotech industries, and consumers – those who are food allergic and those who are not. Much of these benefits will be realised through exploitation of research results by SME industries.

More specifically, EuroPREVALL will:

- Provide a library (collection) of highly characterized , well defined food allergens which will allow the development of component resolved diagnosis, to developed novel methods with improved diagnostic power for food allergy including allergen chips and measures of IgE-cross linking as an alternative skin test in vitro.
- Carry out epidemiological studies to establish the true occurrence, prevalence and distribution of food allergy and hypersensitivity in children and adults as well as influences during pregnancy and early-years, involving a birth cohort of 10,000 babies, yet to be born across Spain, Germany, Iceland, Greece, Poland and the UK.
- Investigate the role of diet, environment and infections in determining patterns of food allergy by mapping patterns of food allergy against other environmental allergies and patterns of food consumption. Determine the role of infections in allergy as well as the effectiveness of genetic markers.
- Measure the socioeconomic impact of food allergy, the cost to society and the effect on food allergic individuals and their families by questionnaires. Develop tools to assess the economic impact on individuals and their households, as well as the agro-food industry.
- Use the battery of existing methods, and those newly developed in EuroPrevall, to predict the outcome of exposure of susceptible individuals to allergens in real foods and in model systems.

- Capitalise on existing freely-available evidence-based information, developed in conjunction with allergic patient groups for the benefit of patients and the scientific community.
- Create training materials for patient groups, health professionals and the food industry.
- Train a new generation of food allergy scientists in a consensus approach.

Italy is one of the country involved in the project. The participation concerns on the assembling of a new library of natural and recombinant allergens. Allergens will be characterised with regards structure (primary to quaternary), using NMR and X-ray crystallography, and post-translational modification and the presence of different isoforms (EuroPREVALL, 2005).

1.5 HIGH THROUGHPUT PRODUCTION OF RECOMBINANT PROTEINS

The discovery of thousands of novel protein-coding sequences and the increased availability of large cDNA collections provide the opportunity to investigate protein function in a systematic manner and at an unprecedented scale. This opportunity highlights the need for development of high throughput (HT) methods for protein isolation.

Purified proteins are a key reagent for numerous assays that address fundamental questions about their structure, function and regulation.

For applications that require large amounts of purified proteins (mg) HT methods for protein isolation can be used to efficiently screen many different constructs (orthologues, tags etc.) to identify those that produce a high yield of soluble protein. For microscale applications HT protein purification provides the front end to produce proteins for various applications that require limited amount of protein per sample.

Over the past few decades several protein **expression systems** have been developed for recombinant protein expression. Different expression hosts (e.g., *Escherichia coli*, *Saccharomyces cerevisiae*, *Pichia pastoris*, insect and mammalian cell lines) have been developed to express heterologous proteins. Additionally, genomic approaches are being pursued to solve the structure of numerous proteins.

Each of these systems has its strengths and weaknesses concerning yield, proper folding, post-translational modification (PTM), cost, speed and ease of use. With respect to HT application, it is also useful to consider the 'success rate' – the fraction of proteins that can be produced in practical yields.

Escherichia coli is the simplest and by far the most widely used organism for protein expression. The most appreciated advantages of this system are its speed, ease of use and low cost. These advantages are sometimes offset by the lack of eukaryotic post-translational modification PTMs and the low solubility of some proteins.

Interest in *Saccharomyces cerevisiae* as a protein expression system has increased recently because this organism combines the advantages of an inexpensive, fast growing unicellular organism with the physiological properties of a eukaryotic cell.

Protein production in *insect* cells is commonly performed using the baculovirus system. This protein expression system is popular for large scale protein expression. Appreciated advantages of insect cells are the robust and relatively inexpensive cell culture and the fact that most eukaryotic PTMs are executed properly.

Cell-free expression systems are particularly attractive for HT applications because the absence of a cell membrane eliminates the harsh process steps associated with introducing

DNA into cells, lysing cells and clearing lysate. Expression systems from eukaryotic cell lysates have the additional benefit that most PTMs are executed properly. The most widely used open expression systems are bacterial, wheat germ and reticulocyte lysates, although lysates from other cell types have also been made (Pascal B et al., 2003).

High-throughput protein purification depends on **affinity tags** to provide a generic purification strategy. The use of affinity tags enables different proteins to be purified using a common method as opposed to highly customized procedures used in conventional chromatographic purification.

Introducing an affinity tag may have a positive effect in the biochemical properties of the target protein, especially in bacterial protein expression. Affinity tags have been observed to:

- improve protein yield,
- prevent proteolysis,
- facilitate protein refolding,
- protect the antigenicity of the fusion protein, and
- increase solubility.

On the other hand, although the advantages it is important to consider any tag can potentially negatively affect the target protein resulting in:

- a change in protein conformation,
- lower protein yields,
- inhibition of enzyme activity,
- alteration in biological activity,
- undesired flexibility in structural studies
- toxicity

Due to the somehow unpredictable changes that adding a tag may introduce in a protein and its behaviour, it is usually desirable to remove the tag. This reflects on the design of the protein fusion. Importantly, removal of the tag needs to be considered when designing a process for the production of a recombinant protein that is intended for human use to enable production of a 'native' (i.e., tagless) protein. And consequently, both the enzyme(s) used to cleave the tag and the cleaved fusion partner need to be removed from the purified protein. Affinity tags can be defined as exogenous amino acid (aa) sequences with a high affinity for a specific biological or chemical ligand. A major group of affinity tags consists of a peptide or protein that binds a small ligand linked on a solid support (e.g., his-tags bind to immobilized metals). Another group includes tags that bind to an immobilized protein partner such as an antibody or antibody purification using protein A affinity chromatography.

In Table 1 the main tags available are reported with the respective dimension (aa). Although the number of different tags available may reflect the different biochemical properties of proteins and the host cells, it is possible to discriminate tags needed for purification (e.g., his-tag) from those developed to aid, e.g., solubility and folding like maltose-binding protein (MBP), glutathione *S*-transferase (GST) or small ubiquitin modifying protein (SUMO), among others (Arnau J et al., 2006).

Tag	**Size (aa)**	**Comments**
His-tag	5-15	Purification under native or denaturing conditions
FLAG	8	Calcium-dependent, mAb-based purification
Streptag II	8	Modified streptavidin, elution with biotin analog
HA-tag	9	InXuenza virus hemagglutinin tag, Ab-based purification
Softag1, Softag 3	13-8	Recognized by polyol-responsive mAb
c-myc	10	mAb-based purification
T7-tag	11-16	mAb-based purification
S-tag	15	S-protein resin affnity purification
Elastin-like peptides	18-320	Protein aggregation by temperature shift, intein used to remove tag
Chitin-binding	52	Binds only insoluble chitin
Thioredoxin	109	Affnity purification with modified resin
Xylanase	163	Cellulose based capture, elution with glucose
Glutathione *S*-transferase	201	Glutathione or GST-Ab affnity
Maltose binding protein	396	Amylose affnity purification
NusA	495	Increased solubility in *E. coli*. Affnity tag needed for purification

Tab. 1.3. Affinity and solubility tags for recombinant proteins (Arnau J et al., 2006).

It is often reasonably assumed, although rarely formally demonstrated, that small tags, such as the His_6-tag, bear a smaller risk of steric interference than larger tags, such as glutathione-S-transferase (GST) or maltose binding protein (MBP).

The His_6-tag is a popular purification tag because of its small size, its relatively strong reversible binding and because it functions under denaturing conditions. These advantages have led many structural proteomics groups to use the His_6-tag almost exclusively for their initial studies.

For prokaryotic proteins the success rate is typically, 50%, which most likely reflects the close phylogenetic relationship between the target organisms and the expression system.

For eukaryotic proteins, the success rates are significantly lower. Proteins become progressively more difficult to purify as their molecular weight increases.

Given the poor performance of the His_6-tag for higher eukaryotic proteins, many alternative purification tags have been developed, many of which increase the solubility of the recombinant proteins. Several studies systematically examined the success rates obtained with different tags in HT purifications (Pascal B et al., 2003).

1.6 NMR IN STUCTURAL PROTEOMICS OF SMALL PROTEINS

The Magnetic Nuclear Resonance (NMR) is advised, to good reason, the prince technique for the structural determination of the biomolecules in solution, that is flanked to the already consolidated and well exploited X rays solid state crystallography. The NMR notional principles are largely treated in specialized books and, for the present work of thesis, they will be only illustrated with a distinctive attention turned to depict the type of information that it could be obtained from the spectrum of a protein in solution.

The result of a NMR experiment on a very simple molecule, *e.g.* the the water, takes the name of FID (Free Induction Decay).

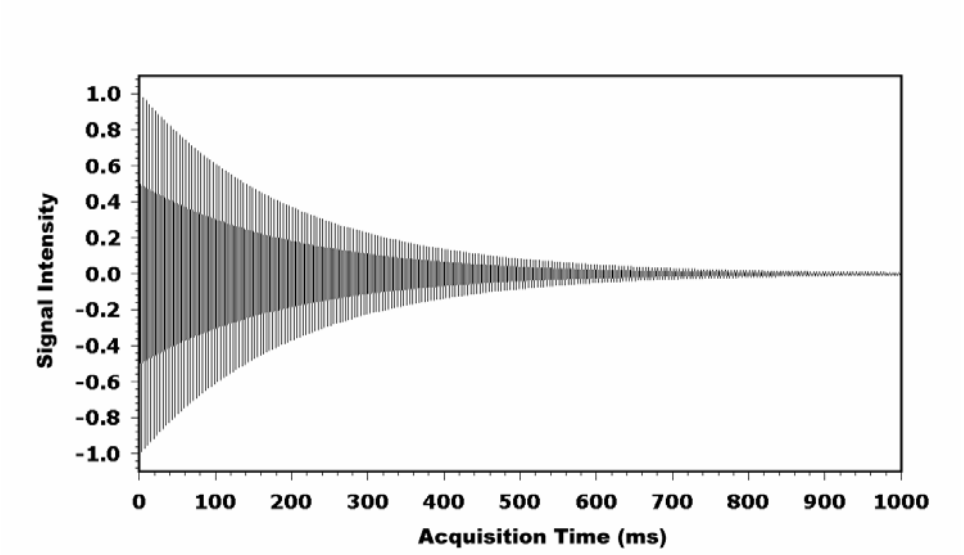


Fig. 1.9 Representation of the water signal (FID) caused from only one type of hydrogen atom.

It consists of a signal with fluctuating intensity in the time (defined time of acquisition) with an exponential decadence trend. For the water molecule, that contains only one type of hydrogen atom, the resultant FID contains three information: 1) the frequency of resonance, dependent from the chemical characteristics of the molecule; 2) the exponential decay time, defined as the time of laxness, that depends on different factors, first of all the rotation speed of the molecule around to his own axis; 3) the signal intensity, proportional to the observed nucleus concentration, in the sample.

The more complex molecules have more difficult to comprise FID and, for this Fourier's transform allows the conversion in a more comprehensible spectrum that contains the same original information.

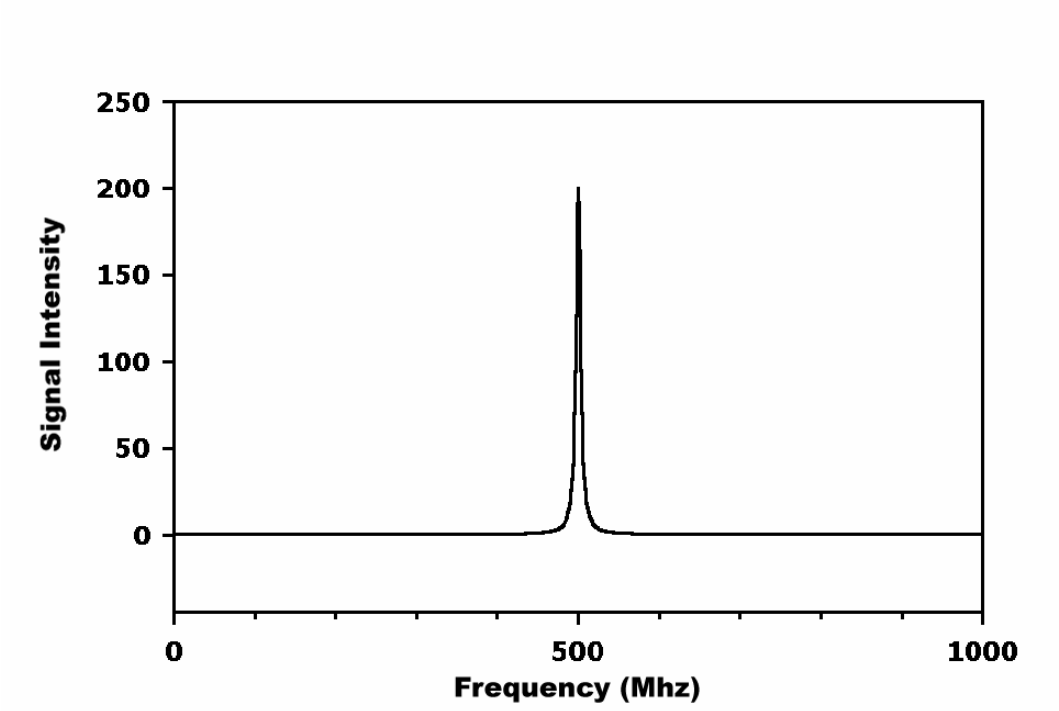


Fig. 1.10 Fourier's transform of the water FID.

The axis of the abscissas shows the frequency of resonance of the observed nucleus (in this case the atoms of the equivalent hydrogen of the water) that is different for each type of magnetically active nuclei. The position of the signal in the spectrum is commonly expressed with the term "chemical shift", that indicates the difference between resonance frequency of the observed signal and the frequency of a reference molecule. The value is generally comprised between 0 and 10 ppm, although, in many substances, could be observed values distant from these limits. For any proteins it's possible to acquire hydrogen spectra but also those relative to other atoms, for example carbon-13 and nitrogen-15: each of them resonating at completely different frequencies from the other one. The concentration of the nucleus in the sample (equal to that of the molecule multiplied for the number of identical nucleus for molecule) is proportional to the area of the signal.

The other information, concerning the transversal relaxation time (T_2), is contained in the width, to half height, of the transformed signal ($\Delta\nu$):

$$\Delta\nu = 1/ (\pi\ T_2)$$

In a protein that doesn't contain prosthetic paramagnetic groups, the T_2 is inversely proportional respect to the rotation time around to the rotation axle of the molecule (τ_R),

unless other local movements intervene to alter, also transitorily, the chemical environment of the nucleus in observation. However, in the more general case:

$$T_2 = c / \tau_R$$

According to the relation of Stokes-Einstein:

$$\tau_R = 4\pi\eta a^3 / (3\kappa T)$$

the rotation time is inversely proportional with respect to the kinetic energy of the molecule and directly proportional with respect to the solution viscosity (η) and to the volume of the molecule (proportional to the cube of the ray of the molecule, **a**, considered globular).

This indicator is, in approximation, directly proportional to the molecular molar mass. This explains the reason for which the NMR is applicable to molecules with molar mass under 30 kDa. That is the same to say that proteins (or aggregates of them), with a number of 270 amino acids, could not be studied (considering a middle molar mass for each amino acid of 110 Da). For bigger molecules, each NMR signal would be so large to result flat: its height would be equivalent to ground noise, since the signal's area has to remain constant. Large and few intense signals are absolutely to be avoid in NMR, like in any other technique, ever since the spectrum could show completely superposed signals and, then, not unequivocally identifiable.

The spectrum of the glycine, the natural smaller amino acid, dissolved in water, presents two signals: one relative to the hydrogen atoms in α position (that are two equivalent) and the other one referred to the solvent-water added to it.

$$H_2N\text{-}CH_2\text{-}COOH$$

The signals of the aminic hydrogen atoms and those of the carboxylic groups result instead grouped in those of the solvent, since they exchange chemically with it, and so they result invisible. Since the concentration of the solvent is, commonly, three orders of greatness superior than that one of the solute, its signal would be hugely predominant in the spectrum and, for reasons that won't be discussed for brevity, it could obscure the presence of the less intense signals than which of the dissolved molecules. For this reason, it is opportune to use deuterated water (D_2O) as solvent, which is spectroscopically silent. The result is still a spectrum containing a single signal (except for a residual hydrogen signal due to traces of HOD formed for chemical exchange of the aminic protons with the carboxylic one and the solvent).

The spectrum of the dissolved polypeptide poliglycin in D_2O, presents a number of signals equal to the double of that of the amino acids that constitute it, instead that the same number.

$$H_2N\text{-}CH_2\text{-}CO\text{-}(NH\text{-}CH_2\text{-}CO)_n\text{-}NH\text{-}CH_2\text{-}COOH$$

This because the molecule assumes a low energy conformation that loses every symmetry; so that all the hydrogen atoms, also those bound to the same atom of carbon, get a different chemical environment and becoming inequivalent. Each of these signals is characterized by a resonance frequency, and then a *chemical* shift very similar, determining a very crowded spectrum in the region around 3.5 ppm. For a protein with mean dimensions (10-15 kDa), in which all the natural amino acids are present, the ^{1}H-NMR spectrum results to be very complex in all its regions (from 0 to 10 ppm), resulting by the overlap of 500-800 signals derived from hydrogen atoms with different chemical characteristics.

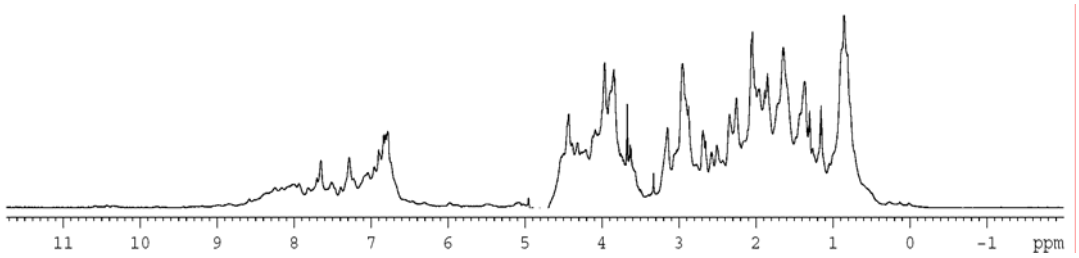


Fig. 1.11 ^{1}H-NMR spectrum of a protein with medium dimension (10-15 kDa) in which all the natural amino acids are present.

Although distinguished signals aren't visible, it's however possible divide the spectrum in characteristic regions where the different chemical types of hydrogen atoms resound. Proceeding from left (defined as *downfield* spectrum part) to right (*upfield*), the spectrum can be divided into the following regions:

a) amidic region (11-7.5 ppm): signals of amidic protons that don't exchange (not exchanging) with the solvent. This region is indicative of a molecular compact structure (protein *folding*) as only the internal amidic groups are inaccessible to the solvent and don't exchange protons with it;

b) aromatic region (7.5-6.5 ppm) is relative to the lateral chains of amino acids like phenylalanine, tyrosine and tryptophan, commonly collected in the hydrophobic "core" of the protein;

c) alfa-protonic region (4.8-3.5 ppm): it is the region where the signals of hydrogen of every amino acid in α position (αCH) resound; this area of the spectrum is diagnostic for the secondary protein structure: as more the signals are dispersed in this region as more the polypeptidic skeleton is rigid;

d) aliphatic region (3.0-1.0 ppm): here there are the hydrogen atoms signals of all the lateral chains. Being the more crowded region it's not able to give specific information in the analysis of a mono-dimensional spectrum;

e) methyl region (1.0-0.0 ppm): all the methyl protons signals of alanine, methionine, threonine, valine, leucine and isoleucine resound in this region. The presence of isolated signals under 0.5 ppm is indicative of stable hydrophobic interactions, and then of a rigid tertiary structure.

Besides the above-mentioned indications obtained from the analysis of a mono-dimensional spectrum (1D-NMR), the NMR spectroscopy is able to give information on the spatial relations of the protein molecule, allowing an indirect measure of the distances between the atoms constituting it. For this purpose it's necessary to acquire a bi-dimensional spectrum (2D-NMR), resultant from an experiment known as NOESY (acronym of Nuclear Overhauser Effect SpectroscopY). In defining the number of spectrum dimension, it needs to accentuate that it is relative to the only domain of the frequencies. For this, while the 1D-NMR spectrum is characterized by a graphic that reports on the abscissas the domain of frequencies and on the ordinates the signals intensity, the 2D-NMR spectrum is constituted by frequencies domain both in abscissas and in ordinates, since the signals intensity dimension is expressed like the level curves analogous to isohypse of a geographical map. In figure 1.12 a schematic representation of a 2D-NMR spectrum is showed. The molecule is characterized by six signals: three amidic not exchanging with the solvent and three αCH. The diagonal of a 2D-NMR spectrum coincides with the 1D-NMR spectrum, while out of the diagonal three correlation signals are present, so called as they correlate two signals staying on the diagonal. From the analysis of correlation signals it is possible to acquire two information: a) the exact chemical shift of each signals, one correspondent to the abscissas and the other one to the ordinates of the bi-dimensional graphic; b) the distance of the atoms corresponding to the signals, since the correlation peak volume is proportional to r^{-6}, where r is the distance. The decisive passage to obtain the three-dimensional molecule structure is the signals assignment, that is the attribution to each hydrogen atom of the molecule a precise position in the 1D-NMR spectrum.

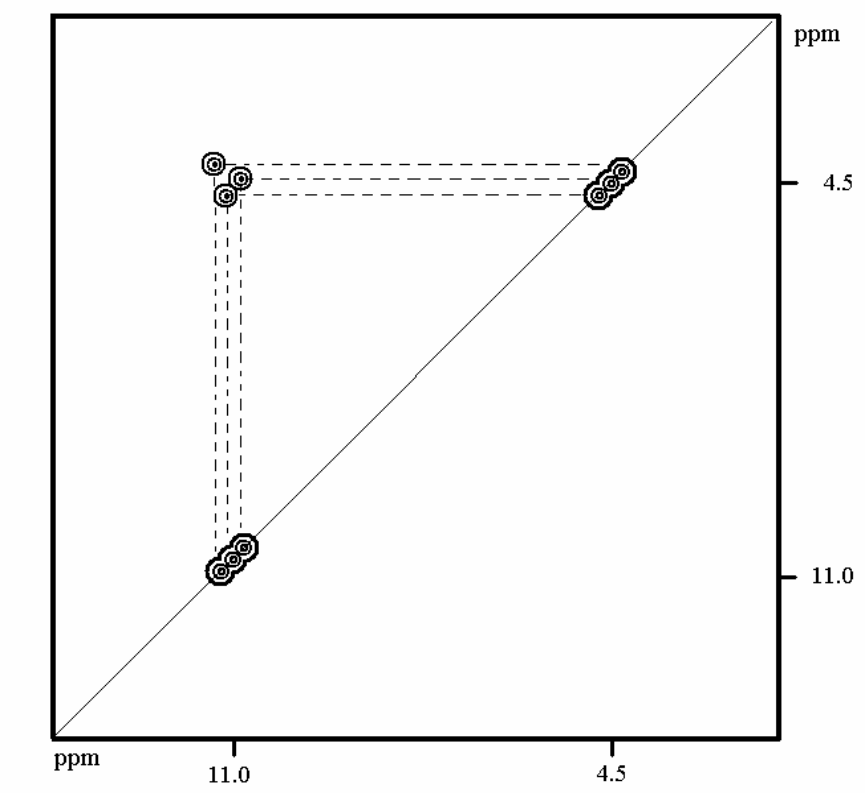


Fig. 1.12 Schematic representation of a 2D-NMR spectrum of a molecule characterized from six signals.

The 2D-NMR spectrum analysis provides important information, also without arriving to molecule structure determination involving a complex work. The form and the dispersion of correlation peaks allows to establish if the molecule forms aggregation state or if it has a rigid structure, sign of a correct tertiary structure. This information is very important in proteomic research projects aiming to the study of the relations between protein sequence, structure and function.

2 MATERIALS

2.1 EQUIPMENT

Analytical balance (SBC 21, Scaltec).

Centrifuge tubes (Beckman Coulter).

Chromatographic columns:

- metal chelate affinity chromatography (nickel chelate HiTrap, Amersham Biosciences);
- GST affinity chromatography (GSTrap, Amersham-Biosciences);
- size exclusion chromatography (Superdex 75 Hiload 16/60 column, Pharmacia Biotech);
- anion exchange chromatography (diethylaminoethyl, DEAE column, Whatman) or cation exchange chromatography (Macro Prep S supports SO_3^-, BIORAD)

Dialysis membrane (MWCO, 4000-6000) con relativi fermi (CELLU.SEP T1, NFPI).

Eectrorophoresis apparatus (Mini-Protean II, Biorad).

Franch Pressure Cells P_{max} 40000 psi (SIM-AMINCO Spectronic Instrument).

HPLC (Amersham Pharmacia Biotech).

Kier (Vapor Matic 770, Asal).

Refrigerator (FTO 220).

Regulated power supply (E-C 105, E-C Apparatus Corporation).

Shaker (Universal table Shaker 709, Asal srl).

Thermostated agitator: (Climatic Hood 810, Asal srl).

Thermostated centrifuge with rotors: JA 10 RPM_{max} 10000, JA 14 RPM_{max}. 14000, JA 25.50 RPM_{max} 25000 (Avanti J-25, Beckman Coulter).

Thermostated magnetic agitator (T_{max} 300 °C, RPM_{max} 9, FALC).

Microcentrifuge, RPM_{max} 14000 (Beckman Coulter).

Micropipets P2, P10, P20, P100, P150, P200, P1000 (Gilson).

pH meter Basic 20 (CRISON).

Plates (Sterilin).

Ultrafiltration cells, 10-50 ml (8010, 8050 Amicon)

Ultrafiltration regenerated cellulose membrane MW>5000 (PLCC, MILLEPORE).

Vortex.

2.2 REAGENTS

Acetic acid, glacial (C.E., 99% purity)

Acrylamide (Fluka)

Agar

Ampicillin (Fluka)

Bromophenol Blue (Merk-Darmastad)

Calcium carbonate (Rectapur, 99.9% purity)

Commassie Blue (Fluka)

EDTA (C.E., 100% purity)

Ethanol (C.E., 99.9% purity)

Ethanol, 70% (C.E., 69% purity)

Ethidium bromide

Ferrous sulfate epta-hydrate (C.E., 99.5% purity)

Glycerol

Glycin (C.E., 99% purity)

Guanidine hydrochloride (Sigma, 99% purity)

Hydrochloric acid, 37% (C.E., 36.5% purity)

Imidazole (Fluka, 69% purity)

IPTG (Sigma)

Methyl alcohol (C.E., 69% purity)

Magnesium chloride (C.E., 99.5% purity)

Manganous sulfate epta-hydrate (C.E., 99.5% purity)

β-mercaptoethanol (Sigma, 98% purity)

Nichel chloride (C.E., 99.5% purity)

Phosphoric acid (C.E., 99.9% purity)

PMSF (Sigma)

Potassium chloride (C.E., 99.5% purity)

Potassium phosphate bibasic (Carlo Erba)

Protein Molecular Weight marker (Biorad)

SDS-PAGE standard, Low range (Biorad)

Sodium acetate (C.E., 99.9% purity)

Sodium chloride (C.E., 99.5% purity)

Sodio dodecil-solfato (Boehringer Mannheim)

Sodium hydroxide (C.E., 97.5% purity)

Temed (Sigma)

Thiamine (Sigma)

Tryptone (Difco)

Tris-base (C.E., 99% purity)

Urea (Carlo Erba)

Yeast extract (Difco)

2.3 ENZYME

Actev protease (Invitrogen)

DNasi liofilizzata (Boehinger Mannhein)

BR Clonase enzime (Invitrogen)

FactorXa (Amersham)

LR Clonase enzime (Invitrogen)

Proteinase K (Invitrogen)

Trombin

2.4 SOLUTIONS

1 M solutions were prepared by dissolving 1 mole of substance in distilled water. To adjust the pH at a specific value, the acid or the base was added under agitation. The final volume of 1 liter was adjusted in a volumetric cylinder. The 1M solution utilized along the thesis work are the following (the adopted molar masses are reported in parenthesis):

1 M calcium chloride (111 g/mol)

1 M nichel chloride (237.7 g/mol)

1 M potassium chloride (74.55 g/mol)

1 M sodium chloride (58.44 g/mol)

1 M IPTG (200g/mol)

1 M Sodium hydroxide (40g/mol)

1 M Imidazol (68.08 g/mol)

1 M Tris-HCl pH 6.8, 7.5, 8.0, 8.5 (748.6 g/mol)

2.5 LIQUID MEDIA

LB medium was prepared dissolving 10 g of tryptone, 5 g of yeast extract and 10 g of sodium chloride in 800 ml of water. The final volume was adjusted at 1 litre in a volumetric cylinder. The solution was transferred in flasks and autoclaved at 121 °C for 20 minutes. At 60 °C the specific antibiotic was added to the solution.

2xY medium was prepared dissolving 16 g of tryptone, 10g of yeast extract and 5 g of sodium chloride in 800 ml of water. The final volume was adjusted at 1 litre in a volumetric cylinder. The solution was transferred in flasks and autoclaved at 121 °C for 20 minutes. At 60 °C the specific antibiotic was added to the solution.

2.6 SOLID MEDIA

For the growths on plates LB and 2xY medium were used. The media were prepared using the same reagents for the liquid media and 2% agar. The solution was transferred in flasks and autoclaved at 121 °C for 20 minutes. At 60 °C the specific antibiotic was added to the solution. Before the polymerization, the solution was poured in the plates leaved to cool. The plates were stored at 4 ºC until needed.

2.7 SELECTABLE MARKER

The antibiotic Ampicillin was used as selectable marker during the bacterial growths. To prepare an ampicillin stock solution 100 mg/ml (1000x), the below reported procedure was followed.

1 g of ampicillin was dissolved in 8 ml of distilled water. The pH was adjusted at 9 by adding 2N Sodium hydroxide under agitation. The final volume was adjusted at 10 ml in a volumetric cylinder. The solution was stored as 1 ml aliquots at -20ºC.

The antibiotic was added to the media as over described, using a dilution ratio of 1:1000.

3 METHODS

3.1 GATEWAY TECHNOLOGY

The *Gateway* technology is a universal cloning method based on the site-specific recombination proprieties of bacteriophage lambda that provides a rapid and highly efficient way to move DNA sequences into multiple vector systems for functional analysis and protein expression. This method allows using multiple types of DNA sequences (PCR products, cDNA clones, restriction fragments) and multiple destination vectors. Moreover, *Gateway* technology is suitable for adaptation to high-throughput formats giving the opportunity to investigate the more effective methods to isolate protein targets, in a systemic way. In particular, this technology permits to test rapidly more vector-host combinations to increase the successful probabilities in protein expression strategies (Figure 3.1).

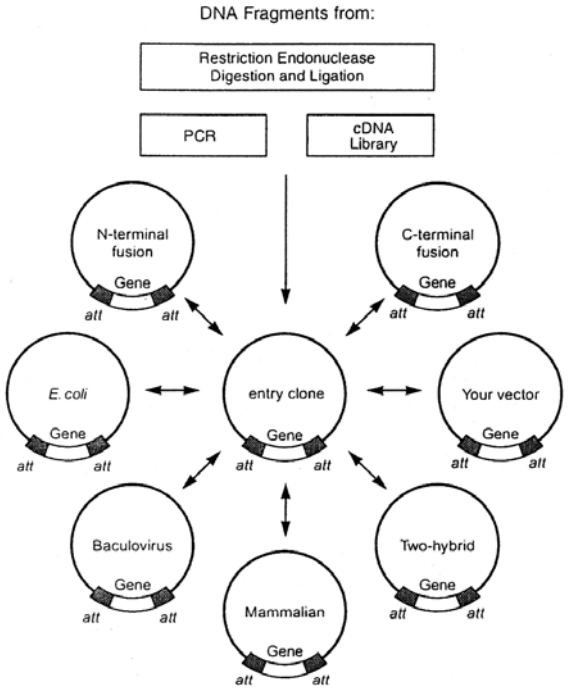


Fig. 3.1 The *Gateway* technology provides a rapid and highly efficient way to move DNA sequences into multiple vector systems for functional analysis and protein expression.

3.1.1 Recombination reactions

A recombination reaction consists of a DNA fragment transfer from a donor to a receiver vector. This reaction is catalyzed by a mixture of restriction enzymes that bind to specific nucleotidic attachment sequences (*att* sites), bring together the target sites, cleave them, and covalently attach the DNA. The *Gateway* technology is based on the bacteriophage lambda

site-specific recombination system which facilitates the integration of lambda into the *E. coli* chromosome. The integration occurs via intermolecular DNA that involves two major components: the DNA recombination sequences (*att* sites) and the *Clonase enzyme mix*, a mixture of lambda and *E. coli* encoded recombination proteins. The *att* sites serve as the binding sites for recombination proteins and have been well-characterized: *att*B on the *E. coli* chromosome and *att*P on the lambda chromosome. The λ integration occurs between *att*B and *att*P sites to give rise to *att*L and *att*R sites. Two recombination reactions constitute the basis of the *Gateway* technology:

- BP reaction catalyzed by BP *Clonase enzyme mix*, facilitates recombination of an *att*B-substrate (ex. *att*B-PCR product) with an *att*P substrate (donor vector) to create an *att*L entry clone (Figure 3.2).

Fig. 3.2: The DONOR vector (pDONOR) contains attP sites compatible with the PCR product extremities (attB sites). The PCR product attach occurs in these nucleotidic sequences (att sites). The entry clone attL sites are composed of sequences from attB sites of PCR product and attP sites from donor vector.

- LR reaction catalyzed by LR *Clonase enzyme mix*, facilitates recombination of an *att*L-entry clone with an *att*R substrate (destination vector) to create an *att*B expression clone (Figure 3.2).

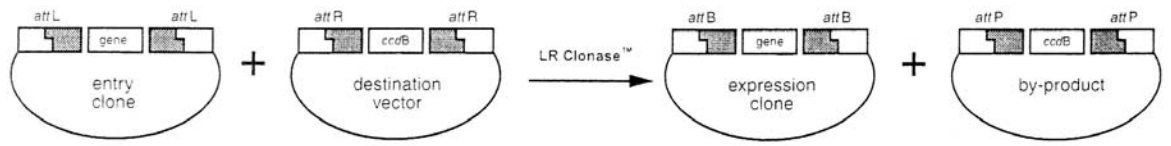


Fig. 3.3: The Entry clone (pENTR) contains attL sites compatible with the attR sites of the destination vector. The gene attach occurs in these compatible nucleotidic sequences (att sites). The expression clone attB sites are composed of sequences from attL sites of the entry clone and attR sites from destination vector. The destination vector contains elements necessary to express the gene of interest in appropriate system (i.e. E.coli, mammalian, yeast, insect).

3.2 AMPLIFICATION OF THE GENES BY PCR

The nucleotide sequences coding for the selected proteins were amplified from cDNA by PCR, using the specific primers, reported in annex 1. In a 0.5 ml microcentrifuge tube, the following 50 µl PCR reaction mixture were prepared:

cDNA template (5ng/µl)	5 µl
dNTP mix (10mM)	1 µl
Primer Forward (10pmol/µl)	10 µl
Primer Reverse (10pmol/µl)	10 µl
Pfu DNA Polymerase	1 µl
Pfu Buffer 10x	5 µl
H_20	18 µl
Final volume	**50 µl**

The amplification was performed using the following cycling parameters:

Step	**Time**	**Temperature**	**Cycles**
Initial Denaturation	5 minutes	94 °C	1x
Denaturation	1 minutes	94 °C	35x
Annealing	1 minutes	54°C	
Extension	2 minutes	72°C	
	10 minutes	72°C	1x

The PCR products were analysed by agarose gel electrophoresis as described at paragraph 3.10.

3.3 BP RECOMBINATION REACTION

According to *Gateway* technology (Invitrogene, 22 September 2003), the ENTRY CLONES were constructed using the PCR products. The reaction mixtures were incubated 1 hour at 25°C with BP Clonase enzyme mix and the specific BP Clonase Reaction Buffer (Invitrogene). Each reaction mixture was composed as follows showed:

5X BP Clonase reaction buffer	4 µl
attB-PCR product	300 ng
pDONR vector	300 ng
TE buffer pH 8.0	to 16 µl
LR Clonase	4 µl

After incubation, 2 µl of Proteinase K solution (2µg/µl) was added to the mixture and incubated at 37°C for 10minutes. Each reaction mixture was used to transform competent TOPO10 *E.coli* host strain as described in paragraph 3.4 and the entry clones obtained were selected on plates using Kanamycin 50µg/ml LB medium.

3.4 TRANSFORMING COMPETENT CELLS

To transform competent cells the following protocol was used. For each transformation, 2 µl of the BP recombinant reaction were added into 1.5 ml microcentrifuge tube containing 50 µl of competent TOPO10 *E.coli* cells. After a gentle shaking, the mixture was incubated on ice for 30 minutes. The cells were heat-shocked for 30 seconds at 42°C, without shaking. Immediately the tubes were transferred to ice. 450 µl of room temperature S.O.C. medium were added to the cells and the tubes were shaken horizontally at 37°C for 1hour. 20µl, 50µl and 100µl of the mixture were spread on pre-warmed selective plates (Kanamycin 50µg/ml) and incubated overnight at 37°C.

3.5 ENTRY CLONES SEQUENCING

The entry clones (pENTR/D-TOPO) generated by BP recombinations were sequenced to confirm the DNA sequences. The plasmid were extracted from 3 ml of growth culture using the Q*uantum prep Plasmid Mini Prep Kit* (BIORAD). The sequencing was performed by PRIMM (Centro Servizi Sequenzaimento DNA, Milano) using the M13 Sequencing Primers whose sequence is here reported:

M13 [forward]	5′-GTAAAACGACGGCCAG- 3′
M13[reverse]	5′-CAGGAAACAGCTATGAC- 3′

The sequencing results were compared with the actual sequences of the gene fragments.

3.6 LR RECOMBINATION REACTIONS.

To produce expression clones, a recombinational cloning of genes with destination vectors was performed. To obtain every gene cloned into different expression systems, all the pENTR/D-TOPO (Invitrogen), containing the specific gene, were recombined with an equimolar mixture of destination vectors, carrying five different specific tags, described at paragraph 4.4: $(His)_6$-tag for pDEST17, GST for pDEST15, Trx+$(His)_6$-tag for pDEST20A, GST+$(His)_6$-tag for pDEST30A, NusA+$(His)_6$-tag for pDEST60A. The reaction mixtures

were incubated for 1 hour at 25°C with LR Clonase enzyme and the specific LR Clonase Reaction Buffer (Invitrogen). Each reaction mixture was composed as follows:

5X LR Clonase reaction buffer	4 µl
Destination vector mix	300 ng
Entry clone pENTER/D-TOPO (gene)	300 ng
TE buffer pH 8.0	to 16 µl
LR Clonase	4 µl

After incubation, 2 µl of Proteinase K solution (2 µg/µl) were added to the mixture and incubated at 37°C for 10 minutes. Each reaction mixture was used to transform three competent *E.coli* host strains (BL21 *Gold, Codon plus, Plys*), to obtain for each strain at least one colony (CFU) for every destination vector.

To transform 50 µl of competent cells 2 µl of LR reaction mixture were used as described at the paragraph 3.4. The transformed cells were isolated on plates with LB medium using Ampicillin (100µg/ml) as selectable marker.

3.7 PROTEIN EXPRESSION SCREENING

After transformation, the isolated colonies were grown in 5 ml of 2xYT medium and Ampicillin (100µg/ml) at 37°C with continuous shaking at 210 rpm. To induce protein expression, Isopropyl β-thiogalactopyranoside (IPTG) was added to the culture when the absorbance at 600 nm reached the value 0.5. Cells were harvested after 20 hours from induction and protein expression was checked by SDS-page described at paragraph

3.8 PROTEIN SOLUBILITY SCREENING

The cells of the samples found positive to expression, were harvested from 2.5 ml of culture. The cells were broken with liquid nitrogen and the cellular debris were centrifuged. The proteins solubility was checked by loading on SDS-page both the soluble fractions and the inclusion bodies obtained by centrifugation.

3.9 EXPRESSION CLONES SEQUENCING

Every destination vector allows to express the protein with a different peptide or fusion protein (tags), encoded from a different nucleotidic sequence with specific dimension (bp). To confirm that the DNA sequences were correct, the expression clones generated by LR recombinations and showing a high expression yield were sequenced. The plasmid were

extracted from 3 ml of growth culture using the Q*uantum prep Plasmid Mini Prep Kit* (BIORAD). The sequencing was performed by PRIMM (Centro Servizi Sequenzaimento DNA, Milano) using the T7 Sequencing Primers whose sequence is here reported:

T7[forward]	5′-TAATACGACTCACTATAGGG- 3′
T7[reverse]	5′-GCTAGTTATTGCTCAGCGG - 3′

The sequencing results were compared with the actual sequences of the gene fragments.

3.10 AGOROSE AND SDS-PAGE ELECTROPHORESIS

Electrophoresis is the migration of charged molecules in solution in response to an electric field. Their rate of migration depends on the strength of the field, on the net charge, size and shape of the molecules and also on the ionic strength, viscosity and temperature of the medium in which the molecules are moving. As an analytical tool, electrophoresis is simple, rapid and highly sensitive. It is used analytically to study the properties of a single charged species, and as a separation technique.

Generally the sample is run in a support matrix such as paper, cellulose acetate, starch gel, agarose or polyacrylamide gel. The matrix inhibits convective mixing caused by heating and provides a record of the electrophoretic run: at the end of the run, the matrix can be stained and used for scanning, autoradiography or storage. In addition, the most commonly used support matrices - agarose and polyacrylamide - provide a means of separating molecules by size, in that they are porous gels. A porous gel may act as a sieve by retarding, or in some cases completely obstructing, the movement of large macromolecules while allowing smaller molecules to migrate freely. Because diluted agarose gels are generally more rigid and easy to handle than polyacrylamide of the same concentration, agarose is used to separate larger macromolecules such as nucleic acids, large proteins and protein complexes. Polyacrylamide, which is easy to handle and to make at higher concentrations, is used to separate most proteins and small oligonucleotides that require a small gel pore size for retardation.

Proteins are amphoteric compounds; their net charge therefore is determined by the pH of the medium in which they are suspended. In a solution with a pH above its isoelectric point, a protein has a net negative charge and migrates towards the anode in an electrical field. Below its isoelectric point, the protein is positively charged and migrates towards the cathode. The net charge carried by a protein is in addition independent of its size - ie: the charge carried per unit mass (or length, given proteins and nucleic acids are linear macromolecules) of molecule

differs from protein to protein. At a given pH therefore, and under non-denaturing conditions, the electrophoretic separation of proteins is determined by both size and charge of the molecules.

Nucleic acids however, remain negative at any pH used for electrophoresis and in addition carry a fixed negative charge per unit length of molecule, provided by the phosphate group of each nucleotide of the nucleic acid. Electrophoretic separation of nucleic acids therefore is strictly according to size.

3.10.1 SDS-PAGE of Proteins

Sodium dodecyl sulphate (SDS) is an anionic detergent which denatures proteins by "wrapping around" the polypeptide backbone and SDS binds to proteins fairly specifically in a mass ratio of 1.4:1. In so doing, SDS confers a negative charge to the polypeptide in proportion to its length. It is usually necessary to reduce disulphide bridges in proteins before they adopt the random-coil configuration necessary for separation by size: this is done with 2-mercaptoethanol or dithiothreitol. In denaturing SDS-PAGE separations therefore, migration is determined not by intrinsic electrical charge of the polypeptide, but by molecular weight.

This is done by SDS-PAGE of proteins or agarose gel electrophoresis of nucleic acids of known molecular weight along with the protein or nucleic acid to be characterised. A linear relationship exists between the logarithm of the molecular weight of an SDS-denatured polypeptide, or native nucleic acid, and its Rf, the ratio of the distance that the protein has moved from its point of origin (the beginning of the separating gel) relative to the distance the tracking dye has moved (the gel front). The Rf is calculated as the ratio of the distance migrated by the molecule to that migrated by a marker dye-front. A simple way of determining relative molecular weight by electrophoresis (Mr) is to plot a standard curve of distance migrated vs. log10MW for known samples, and read off the logMr of the sample after measuring distance migrated on the same gel.

There are two types of buffer systems in electrophoresis, continuous and discontinuous. A continuous system has only a single separating gel and uses the same buffer in the tanks and the gel. In a discontinuous system, a non-restrictive large pore gel, called a stacking gel, is layered on top of a separating gel called a resolving gel. Each gel is made with a different buffer, and the tank buffers are different from the gel buffers. The resolution obtained in a discontinuous system is much greater than that obtained with a continuous system.

Gel preparation

Two glass plates were assembled with two side spacers and clamps as shown by instructions. The resolving and stacking gels were prepared on ice to avoid the rapid polymerization. The specific compositions were reported below:

Resolving Gels: Acrylamide concentration of 16% in 0.25 M Tris-HCl pH 8.8	
Reagents	**Volume (ml to make 10 ml)**
40% Acrylamide	4
water (distilled)	2,05
1 M Tris-HCl pH 8.8	3,75
10% SDS	0,1
10%APS	0,1
TEMED	0,01
Stacking Gels: Acrylamide concentration of 4.5% in 0.125 M Tris-HCl pH 6.8	
Reagents	**Volume (ml to make 5 ml)**
40% Acrylamide	0,66
water (distilled)	2,8
1 M Tris-HCl pH 6.8	0,5
10% SDS	0,04
10%APS	0,04
TEMED	0,01

Tab. 3.3 In the table the composition of Resolving and Stacking Gels are reported.

The ingredients were added to the reaction tube in the order shown above and gently mixed, ensuring that no air bubbles were formed. The *Resolving Gel* solution was poured into glass plate carefully assembled. The gel was overlaid with isopropanol to ensure a flat surface and to exclude air. The isopropanol was washed off after gel had set (±15 min). After mixing the *Stacking Gel* solution was poured onto top of set. The comb was inserted before and removed after polymerization. The gels were assembled in the tank filled with electrophoresis buffer composed of 196mM glycine, 0.1% SDS, 50mM Tris-HCl pH 8.3. The liquid samples were diluted with the solution composed of 125mM Tris-HCl pH 6.8, 10% 2-mercaptoethanol, 10% SDS, 10% glycerol, containing a little bromophenol blue, while the pellet were resuspended in the same solution and then diluted with water. The samples were boiled for 5 min at 95°C in a waterbath and injected by a micro-syringe (Hamilton) buffer on stacking gels. The apparatus was connected with a voltage of 60V and, after the tracer was run through the Stacking Gel, the voltage was increased to100V.At the end of the run, the gels were covered with staining solution composed of 0.2% Coomassie Brilliant Blue, 45:45:10 %

methanol:water:acetic acid. The gels were sealed in a plastic box and leave 2 hours at 37°C on shaker. The gels were destained with 25% 65% 10% methanol water acetic acid mix, with shaking.

3.11 LARGE SCALE PROTEIN PRODUCTION

The proteins were produced on a large scale to obtain sufficient amount for NMR analysis. The protein sample production followed the steps described in the diagram reported in figure 3.4.

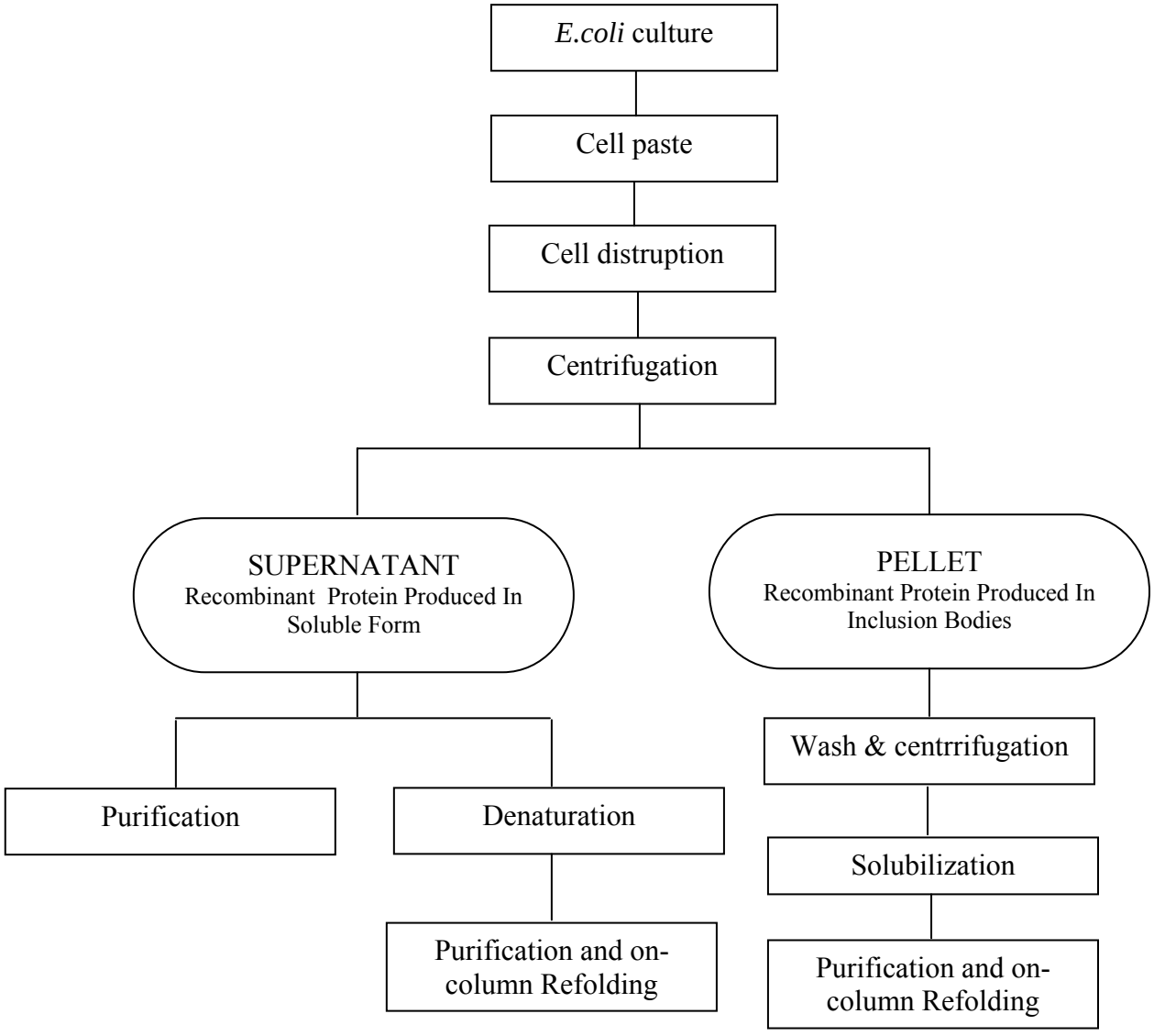


Fig. 3.4 Protein production steps.

The growths were performed in 3 litres of medium and ampicillin (100μg/ml) using the same conditions of the expression screening (paragraph 3.7). The cell were harvested by

centrifugation and stored at -80°C. The cellular pellet was resuspended in lisys buffer and cells were disrupted by three passages through French press (SIM Aminco) at 10000 psi. The cell lysate was separated from cell debris and unbroken cells by centrifugation for 30 min at 14000 rpm in Beckman JA 25.50 centrifuge. If the recombinant proteins were produced in a soluble form, the supernatant containing the protein was applied on cromatography columns (paragraph 3.12) to isolate the protein sample.

Purification of the soluble proteins, was made using a standard protocol for Histidine-tagged (or GST-tagged) proteins. The purification techniques are described in paragraph 3.12. The following steps were adopted:

- metal chelate (or GST) affinity chromatography (nickel chelate HiTrap and GStrap, Amersham-Biosciences);
- cleavage with Factor Xa (Amersham) for Histidine-tagged proteins or with Actev Protease (Invitrogene) for the others tags;
- metal chelate and GST affinity chromatography to remove the tag and uncleaved protein;
- size exclusion chromatography (Superdex 75 Hiload 16/60 column, Pharmacia Biotech);
- anion exchange chromatography (diethylaminoethyl, DEAE column, Whatman) or cation exchange chromatography with Macro Prep S supports (SO_3^-, BIORAD).

Purification of **insoluble** and **unfolded proteins** required a denaturation refolding step on the nickel chelate column. Briefly, after loading the His_6-tagged protein sample on the Hitrap column, a denaturation solution (50mM Tris pH 8.0, 200mM NaCl, 7M Guanidinium chloride) was passed through the column, for four hours at flow 0.5 ml/min. Guanidinium was slowly removed from the column by a 16 hours continuous gradient against (50mM Tris pH 8.0, 200mM NaCl), to allow protein refolding. After elution of the protein from the Hitrap column the proteins were digested and isolated as just described.
All purification steps were performed at 4°C and checked by SDS-PAGE.

The concentration of purified protein samples were determined by measuring the absorbance at 276 nm, using the specific extinction coefficients as calculated by ProtPar parameters tools (http://us.expasy.org/tools/protparam.html).

3.12 PROTEIN PURIFICATION TECHNIQUES

Biomolecules were purified by using techniques that separate according to differences in specific properties as summarized in table 3.4 and showed in figure 3.5.

Property	Technique
Biorecognition (ligand specificity)	Affinity chromatography
Charge	Ion exchange chromatography
Size	Gel filtration (size exclusion)
Hydrophobicity	Hydrophobic interaction chromatography Reverse phase chromatography

Tab. 3.4 Properties of the purification techniques.

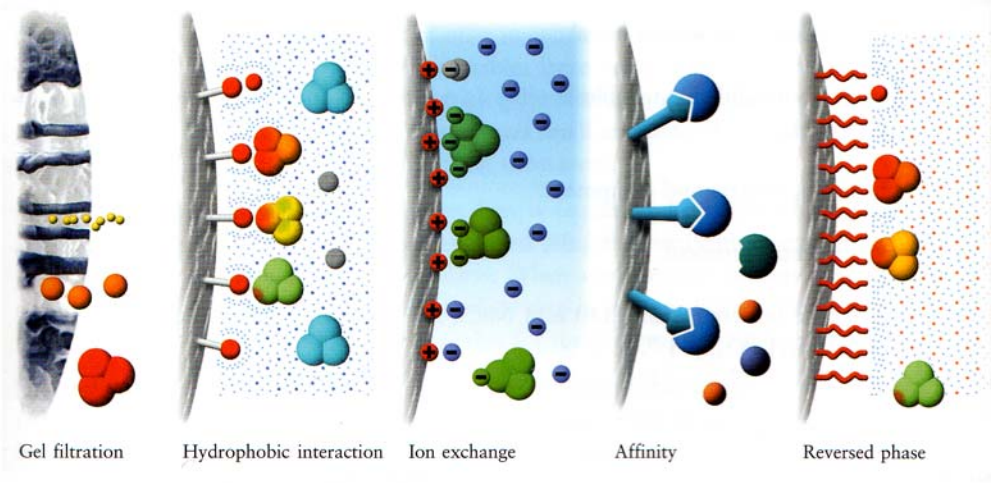


Fig. 3.5 Separation principles in chromatographic purification (Affinity Chromatography Principles and Methods, 2001).

The techniques used for the purification of recombinant proteins during this thesis work, are below illustrated.

3.12.1 Gel filtration

By gel filtration, the molecules mixed in a solution are separated according to differences in their sizes as they pass through a column packed with a chromatographic medium which is a gel. It is also called size exclusion or gel permeation chromatography. In gel filtration chromatography, the stationary phase consists of porous beads with a well-defined range of pore sizes. The stationary phase for gel filtration is said to have a fractionation range, meaning that molecules within that molecular weight range can be separated.

A **gel** is a heterogeneous phase system in which a continuous liquid phase, usually aqueous, is contained within the pores of a continuous solid phase, the gel matrix. In gel made for gel filtration, the pores have a carefully controlled range of sizes, and the matrix is chosen for its chemical and physical stability, and inertness (lack of adsorptive properties). Gels may be formed from polymers by cross-linking dextran. Some polymers, like agarose, form gels spontaneously under the appropriate conditions. For example, composite gels, such as Superdex, may be prepared by grafting a second polymer onto a pre-formed matrix. Dextran chains are covalently bonded to a highly cross-linked agarose gel matrix. Composite gels are of interest since they can combine valuable properties from more than one gel-forming system.

The pores in the gel matrix which are filled by the liquid phase are comparable in size to the molecules we may wish to separate. Relatively to small molecules can diffuse into the gel from a surrounding solution, whereas relatively large molecules will be prevented by their size from diffusing into the gel to the same degree. Sufficiently large molecules are completely unable to diffuse into the gel and are thus confined to the solution outside. In a gel filtration, particles in bead form are packed to form a separation bed through which a buffer solution, the eluent, is passed (Gel filtration Principles and Methods,1998).

Sample molecules which are to be separated are added in solution as a zone to the top of the bed. The sample zone moves down the bed as eluent is added to the top. The small molecules which diffuse into the gel beads are delayed in their passage down the column compared with the large molecules which cannot diffuse into the gel and move continuously down the column in the flowing eluent. The large molecules thus leave the column first, followed by the smaller molecules in the order of their sizes.

Characteristics of the **sample** which are important for the results include its volume and viscosity. The volume of the sample will influence the size of column which will be needed, and the viscosity must not be so large as to cause hydrodynamic instability. It is the viscosity which places an upper limit on the sample concentration. pH, ionic strength and composition are not significant as long as they don't affect the size or stability of the molecules to be separated and are not outside the stability range of the gel filtration medium.

The most important characteristic of a gel filtration **column** is the way in which the gel is packed. If the column is evenly packed so the sample zone is not unnecessarily broadened as it passes down the column then good results can be obtained. If the column is packed unevenly then good results will never be obtained from it. The length of the column, cm, is significant since it affects both the resolution and the time taken to eluite it.

The composition of the **eluent** is unimportant for the separation mechanism. The eluent can thus be whatever is convenient for, with regards to the overall requirements of the experiments. The rate at which the eluents flows through the column affects not only the speed at which the separation is obtained but also the resolution that can be achieved. At least for large molecules, lower is the flow rate, better is the resolution.

Results in gel filtration are typically expressed in the form of an elution diagram showing the variation of solute concentration in the eluent with the volume of eluent passed through the column. For protein and nucleic acid work and in many other applications continuous detection using a UV-monitor and a recorder gives an immediate permanent record, a chromatogram. Solutes normally give symmetrical peaks.

The use of gel filtration for the determination of molecular weight or size, particularly of proteins is well documented.

Several models have been proposed to describe the behaviour of solutes during gel filtration. Most have regarded the partition of solute molecules between the gels particles and surrounding fluid as an entirely steric effect. Larger are the molecular dimension of the solute and greater is the gel forbidden proportion. Proteins that are small enough can fit inside all the pores in the beads and are said to be included. These small proteins have access to the mobile phase inside the beads as well as the mobile phase between beads and elute last in a gel filtration separation. Proteins that are too large to fit inside any of the pores are said to be excluded. They have access only to the mobile phase between the beads and, therefore, elute first.

Proteins of intermediate size are partially included - meaning they can fit inside some but not all of the pores in the beads. These proteins will then elute between the large ("excluded") and small ("totally included") proteins.

In practice, gel filtration can be used to separate proteins by molecular weight at any point in a purification of a protein. It can also be used for buffer exchange - a protein dissolved in a sodium acetate buffer, pH 4.8, can be applied to a gel filtration column that has been equilibrated with tris buffer, pH 8.0. Using the tris buffer, pH 8.0, as the mobile phase, the protein moves into the tris mobile phase as it travels down the column, while the much smaller sodium acetate buffer molecules are totally included in the porous beads and travels much more slowly than the protein (Gel filtration Principles and Methods,1998).

3.12.2 Affinity chromatography

Affinity chromatography separates proteins on the basis of a reversible interaction between a protein (or group of proteins) and a specific ligand coupled to a chromatographic matrix. The technique is ideal for a capture or intermediate step in a purification protocol and can be used whenever a suitable ligand is available for the protein of interest. With high selectivity , hence high resolution, and high capacity for the protein of interest, purification levels in the order of several thousand-fold with high recovery of active material are achievable. Target protein are collected in a purified, concentrated form.Biological interactions between ligand and target molecule can be a result of electrostatic or hydrophobic interactions, van der Waals' forces and hydrogen bonding. To elute the target molecule from the affinity medium the interaction can be reversed, either specifically using a competitive ligand, or non-specifically, by changing the pH, ionic strength or polarity.In a single step, target molecules can be purified from complex biological mixtures, native forms can be separated from denatured forms of the same substance and small amounts of biological material can be purified from high level of contaminating substances.Successful affinity purification requires a biospecific ligand that can be covalently attached to a chromatographic matrix. The coupled ligand must retain its specific binding affinity for the target molecules and, after washing away unbound material, the binding between the ligand and target molecule must be reversible to allow the target molecules to be removed in an active form. Any component can be used as a ligand to purify its respective binding partner (Affinity Chromatography Principles and Methods, 2001). Some typical biological interactions, frequently used in affinity chromatography, are listed in the table below:

Ligand	**Target molecule**
Glutathione	Glutathione-S-transferase or GST fusion protein
Metal ions	Poly (His) fusion protein, native proteins with histidine, cysteine and/or tryptophan residues on their surfaces
Enzyme	Substrate analogue, inhibotor, cofactor
Antibody	Antigen, virus, cell
Lectin	Polysaccharide, glycoprotein, cell surface receptor, cell
Nucleic acid	Complementary base sequence, histones, nucleic acid polymerase, nucleic acid binding protein
Hormone, vitamin	Receptor, carrier protein

Tab. 3.5 Typical biological interactions, frequently used in affinity chromatography.

The purification of recombinant proteins can often be simplified by incorporating a tag of known size into the protein. As well as providing a marker for expression and facilitating detection of the recombinant protein, an important role for the tag is to enable a simple purification by affinity chromatography.

In this work, the affinity columns used for recombinant protein purifications are GSTrap Fast Flow 5ml (Amersham Pharmacia Biotech) and HiTrap chelating HP column (Amersham Pharmacia Biotech). In figure 3.6 the principal steps of the affinity chromatography purification are shown.

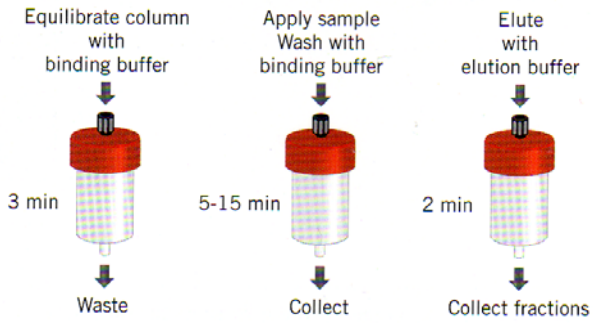


Fig. 3.6 Steps of a poly (His) or GST fusion protein purification using a pre-packed column.

In the table below (3.6) the composition of the specific buffers for GSTrap and HiTrap columns are reported.

	Binding buffer	**Wash buffer**	**Elution buffer**
GSTrap	PBS pH 7.3 (140mM NaCl; 2,7mM KCl; 10mM Na_2HPO_4; 1,8mM KH_2PO_4)	PBS pH 7.3	50mM Tris HCl pH 8 10mM Reduced Glutathione
HiTrap	20mM Tris HClpH 8	-20mM Tris HCl pH 8 - 20mM Tris HClpH 6 100mM KCl	20mM Tris HCl pH 8 300mM Imidazol

Tab. 3. 6 Composition of the specific buffers for GSTrap and HiTrap columns.

3.12.3 Ion exchange chromatography

Separation in ion exchange chromatography depends upon the reversible adsorption of charged solute molecules to immobilized ion exchange groups of opposite charge. Most ion exchange experiments are performed in five main stages. These steps are illustrated schematically in figure 3.7. The first stage is equilibration in which the ion exchanger is brought to a starting state, in terms of pH and ionic strength, which allows the binding of the desired solute molecules. The exchanger groups are associated at this time with exchangeable

counter-ions (usually simple anions or cations, such as chloride or sodium). The second stage is sample application and adsorption, in which solute molecules, carrying the appropriate charge, displace counter-ions and bind reversibly to the gel. Unbound substances can be washed out from the exchanger bed using starting buffer. In the third stage, substances are removed from the column by changing to elution conditions unfavourable for ionic bonding of the solute molecules. This normally involves increasing the ionic strength of the eluting buffer or changing its pH. In Figure 3.7 desorption is achieved by the introduction of an increasing salt concentration gradient and solute molecules are released from the column in the order of their strengths of binding, the most weakly bound substances being eluted first. The fourth and fifth stages are the removal from the column of substances not eluted under the previous experimental conditions and re-equilibration at the starting conditions for the next purification (Ion Exchange Chromatography Principles and Methods, 2001).

.

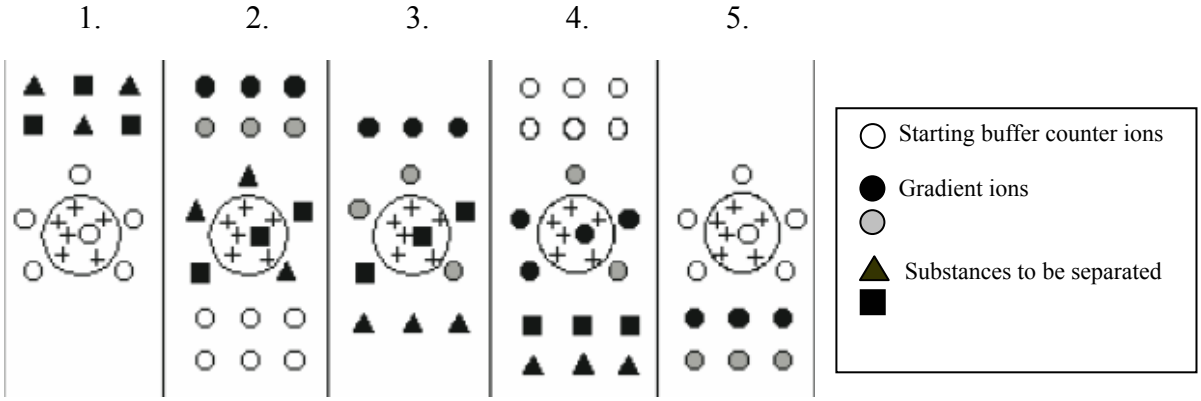


Fig. 3.7 Ion exchange chromatography steps: 1. Starting conditions; 2. Adsorption of sample substances; 3. Start of desorption; 4. End of desorption; 5. Regeneration (Ion Exchange Chromatography Principles and Methods, 2001).

Separation is obtained since different substances have different degrees of interaction with the ion exchanger due to differences in their charges, charge densities and distribution of charge on their surfaces. These interactions can be controlled by varying conditions such as ionic strength and pH. The differences in charge properties of biological compounds are often considerable, and since ion exchange chromatography is capable of separating species with very minor differences in properties, e.g. two proteins differing by only one charged amino acid, it is a very powerful separation technique.

In ion exchange chromatography one can choose whether to bind the substances of interest and allow the contaminants to pass through the column, or to bind the contaminants and allow the substance of interest to pass through. Generally, the first method is more useful since it allows a greater degree of fractionation and concentrates the substances of interest. In addition to the ion exchange effect, other types of binding may occur. These effects are small and are mainly due to van der Waals forces and non-polar interactions.

An ion exchanger consists of an insoluble matrix to which charged groups have been covalently bound. The charged groups are associated with mobile counterions. These counter-ions can be reversibly exchanged with other ions of the same charge without altering the matrix. It is possible to have both positively and negatively charged exchangers.

Positively charged exchangers have negatively charged counter-ions (anions) available for exchange and are called anion exchangers. **Negatively** charged exchangers have positively charged counter-ions (cations) and are termed cation exchangers.

The matrix may be based on inorganic compounds, synthetic resins or polysaccharides.

The characteristics of the matrix determine its chromatographic properties such as efficiency, capacity and recovery as well as its chemical stability, mechanical strength and flow properties. The nature of the matrix will also affect its behaviour towards biological substances and the maintenance of biological activity.

The presence of charged groups is a fundamental property of an ion exchanger. The type of group determines the type and strength of the ion exchanger; their total number and availability determines the capacity (Ion Exchange Chromatography Principles and Methods, 2001). There is a variety of groups which have been chosen for use in ion exchangers; some of these are shown in Table 3.7.

Anion exchangers	**Functional group**
Diethylaminoethyl (DEAE)	-O-CH2-CH2-N+H(CH2CH3)2
Quaternary aminoethyl (QAE)	-O-CH2-CH2-N+(C2H5)2-CH2-CHOH-CH3
Quaternary ammonium (Q)	-O-CH2-CHOH-CH2-O-CH2-CHOH-CH2-N+(CH3)3
Cation exchangers	**Functional group**
Carboxymethyl (CM)	-O-CH2-COO
Sulphopropyl (SP)	-O-CH2-CHOH-CH2-O-CH2-CH2-CH2SO3
Methyl sulphonate (S)	-O-CH2-CHOH-CH2-O-CH2-CHOH-CH2SO3

Tab. 3.7 Functional groups used on ion exchangers.

Some properties of strong ion exchangers are:

• Sample loading capacity does not decrease at high or low pH values due to loss of charge from the ion exchanger.

• A very simple mechanism of interaction exists between the ion exchanger and the solute.

• Ion exchange experiments are more controllable since the charge characteristics of the media do not change with changes in pH. This makes strong exchangers ideal for working with data derived from electrophoretic titration curves.

3.11 PRELIMINARY SCREENING OF PROTEIN FOLDING

The purified samples were analysed by NMR spectroscopy to verify the folding. The folding analysis of the protein samples was performed at CERM (Resonance Magnetic European Centre) in Florence by qualified staff.

NMR samples were prepared by dissolving the protein in 400 μl of the solution containing 100 mM NaCl and 50 μl of D_2O. The final concentration was ranged between 0.2 and 5 mM. The pH of the protein solution was adjusted at 7.

The NMR experiments were acquired on AVANCE 900, AVANCE 800, AVANCE 700 and DRX 500 spectrometers operating at the nominal proton frequencies of 900.13, 800.13, 700.13 and 500.13 MHz. 900, 800 and 500 MHz spectrometers are equipped with a triple resonance (TXI) 5 mm cryoprobe with a z-axis pulse field gradient. All NMR experiments but that used to determine the temperature dependence of the HSQC signals, were performed at 298 K. All the spectra were processed with the Bruker XWINNMR software package and analyzed by the program CARA (ETH Zurich). The samples were left at 4° C between the measurements.

4 RESULTS

4.1 PROTEIN SELECTION

This work was faced to the production of recombinant proteins belonging to the calcium-binding family and their targets. The aim was to obtain pure recombinant protein samples correctly folded for structural characterization by NMR.

Protein category	I.N.	Protein Name	aa	MW (Da)
calmodulin targets	1a)	Neuromodulin	238	24802
	1b)	Neuromodulin GAP-43 5-226	222	22959
	1c)	Neuromodulin GAP-43 41-226	186	18812
	2a)	Endothelial differentiation-related factor 1 ISO-α	148	16369
	2b)	Endothelial differentiation-related factor 1 ISO-β	139	15480
	3)	Brain-specific polypeptide (PEP-19)	62	6660
	4a)	Cyclin-dependent Kinase inhibitor1 (P21)	164	18119
	4b)	Cyclin-dependent Kinase inhibitor1(P21) fragment 1-160 aa	160	17610
	4c)	Cyclin-dependent Kinase inhibitor1 (P21) fragment 25-164 aa	139	15560
	5a)	RNA-binding protein EWS Q01844	656	68478
	5b)	RNA binding protein EWS, fragment 256-447aa	191	20273
	6a)	similar to CAVP-target protein (IQ)	316	34985
	6b)	similar to CAVP-target protein (IQ) splicing variant of CAVP-target	264	28687
	7)	SH3 adapter protein SPIN 90: 722 aa, fragment 1-81 aa	81	9109
	8)	Tuberous sclerosis TSC2 C-terminale domain, 1531-1758 aa	228	56890
EF-Hands	9)	Calcineurin B Homologous Protein (CHP1)	195	22325
	10)	Calcineurin B Homologous Protein 2 (CHP2) O43545 196 aa	196	22321
	11)	Tescalcin	214	24750
	12)	Visinin-Like Protein 1	191	22011
	13)	Visinin-Like Protein 3 (Hippocalcin-Like Protein 1)	193	22182
	14)	Calcyphosine	189	20967
	15)	Calcyphosine Isoform 2	162	24837
	16a)	Calneuron Q	219	24837
	16b)	Calneuron Q N-Terminal Domain 33-104 aa	72	8170
	17)	Guanylyl Cyclase Activating Protein 1 P43080	201	22912
	18)	Guanylyl Cyclase Activating Protein 2 Q	200	23478
	19)	Guanylyl Cyclase Activating Protein 3 O95843	209	23822
	20)	S100A2	97	10985
	21)	S100A16	103	11801
	22a)	Peflin	284	30381
	22b)	Peflin ESP1 (EFH1_2),116-182 aa	67	7814
	22c)	Peflin ESP2 (EFH3_4) ,183 – 250 aa	66	8005
	23a)	ALG-2 ESP1 (EFH1_2)	67	7746
	23b)	ALG-2 ESP2 (EFH3_4)	66	7773
	24)	EP15_HUMAN - ESP1 (EH1) 15-85 aa	71	7759
	26)	S100A10	97	11203
	27)	Calmodulin Like to Skin Protein (CLSP)	146	15920
	28)	Hippocalcin	193	22153
Hippocalcin target	29)	BIR3-NAIP fragment 278-344 aa	68	7610.6

Tab. 4.1 Proteins selected for the project. aa: number of aminoacidics; MW: molecular weight (Da); I.N.: identification number. In red the proteins for which the Gateway entry clone was not obtained.

The protein targets, with known sequence but unknown structure, were selected from a genomic databank, on the base of the relevance of their human cellular function. Moreover, to face one of the essential parameter for NMR structural determination, only proteins with size less than 30kDa were chosen. In the table 4.1 the proteins selected are listed with the respective aminoacidic number (aa) and molecular weight (MW). An Identification Number (I.N.) was assigned to each protein as showed in the table 4.1.

4.2 CLONING OF THE GENES CODING FOR THE SELECTED PROTEINS, IN THE *GATEWAY* pENTR/D-TOPO VECTOR

The nucleotide sequences coding for the proteins reported in table 4.1 were amplified from cDNA by PCR using specific primers. The sequence of the genes and primers is reported in Annex 1.

Following the cloning protocol of the Gateway, the genes coding for the recombinant proteins were extracted from agarose gel and cloned into the pENTR/ D-TOPO by **BP recombination** reaction as describe at par 3.3.

The BP recombination reaction mixtures were used to transform competent TOP10 *E.coli* cells. For each transformation, six well-spaced colonies were selected to extract the plasmids and verify the correct insertion of the genes. This check was performed by PCR using Universal M13 Forward and Reverse primers. In the table 4.2 the expected dimensions of the amplification fragments are reported for each protein.

The PCR products were analysed by agarose gel electrophoresis. The agarose gels of the samples correctly amplified are reported in figure 4.1. (i.e. which gave DNA fragments with right size, see Table 4.2).

I.N. proteins	Amplified fragment dimension (bp)	I.N. proteins	Amplified fragment dimension (bp)
1a	1077	**15**	849
1b	1028	**16a**	1019
1c	920	**16b**	578
2a	806	**17**	965
2b	779	**18**	962
3	548	**19**	989
4a	854	**20**	653
4b	842	**21**	671
4c	782	**22a**	1214
5b	938	**22b**	563
6a	1310	**22c**	566
6b	1159	**23a**	563
7	605	**23b**	560
8	1046	**24**	575
9	952	**26**	653
10	950	**27**	800
11	1004	**28**	941
12	935	**29**	566
13	941		
14	929		

Tab. 4.2 Expected dimensions of the amplified fragments obtained using M13 Universal Primers.

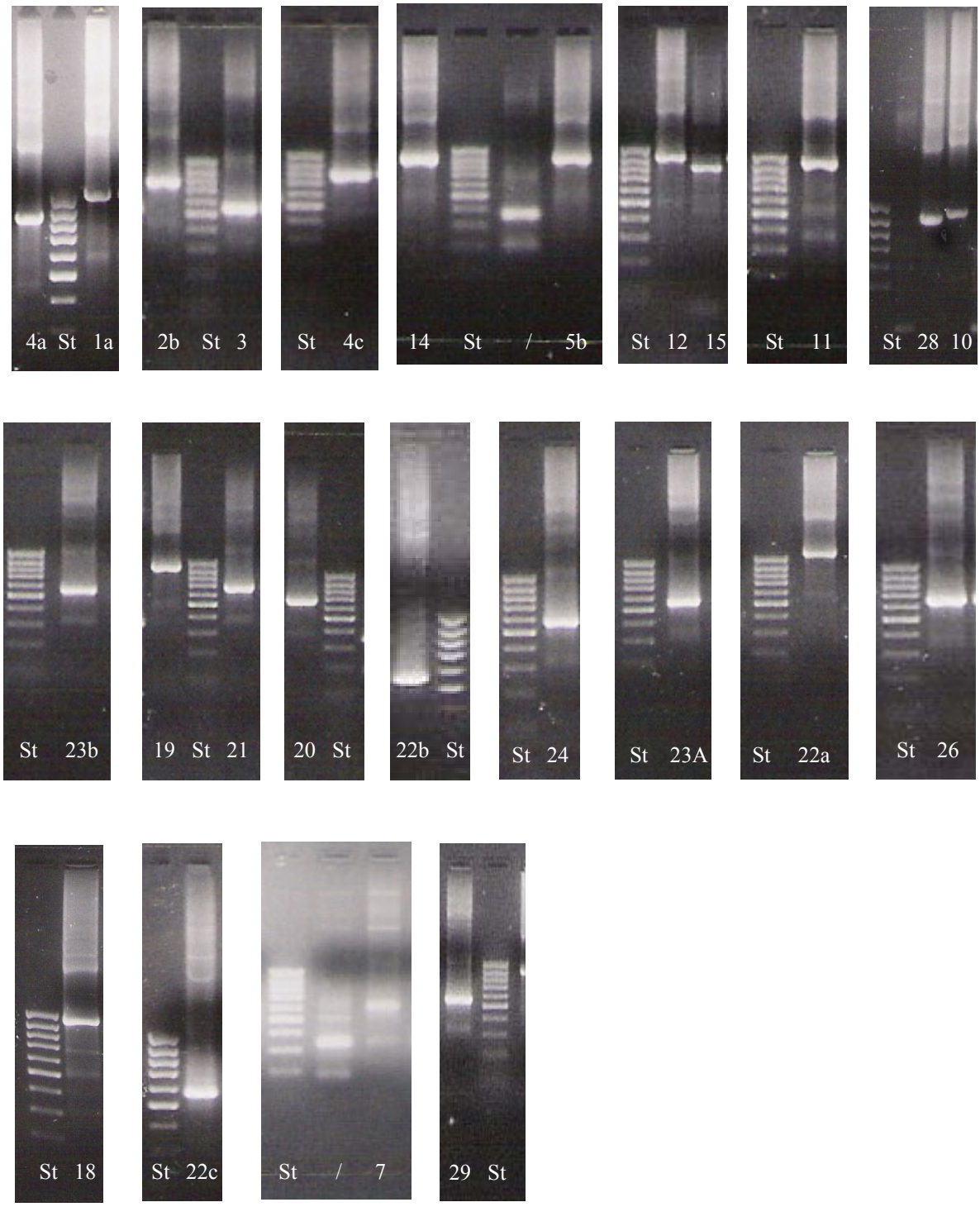


Fig. 4.1. Agarose gels with thw DNA fragments obtained by amplification of the genes cloned in pENTR/D-TOPO using M13 Universal Primers. The standard (St) is composed of ten fragments having known number of base pairs (bp): 1031 bp; 900 bp; 800 bp; 700 bp; 600; 500 bp; 400 bp; 300 bp; 200 bp; 100 bp and 80 bp.

The sequences of the entry clones (pENTR/D-TOPO) were analyzed by single run sequencing using the Universal Forward and Reverse M13 primers, which was performed at Primm Biotech. The sequencing results confirmed that the genetic sequences of the entry clones were

corrected for all the samples, with the exception of those coding for the proteins evidenced with red colour in table 4.1.

4.3 CLONING OF THE GENES CODING FOR THE SELECTED PROTEINS, IN THE *GATEWAY* DESTINATION VECTORS

For this and the next operative steps, the thesis' work was limited to the proteins reported in table 4.3 for time reason. Indeed, along a PhD thesis work the preliminary list of proteins of interest was too long to be completed in areasonable time. The selection of subset was performed by taking into account the highest representativeness of cases. The genes coding for the proteins of table 4.3, were transferred from the pENTR/D-TOPO to five destination vector described in paragraph 3.6.

Protein category	I.N.	Protein Name	aa
calmodulin targets	2a)	Endothelial differentiation-related factor 1 ISO-α	148
	2b)	Endothelial differentiation-related factor 1 ISO-β	139
	4b)	Cyclin-dependent Kinase inhibitor1(P21) fragment 1-160 aa	62
	4c)	Cyclin-dependent Kinase inhibitor1 (P21) fragment 25-164 aa	164
	7)	SH3 adapter protein SPIN 90: 722 aa, fragment 1-81 aa	81
	8)	Tuberous sclerosis TSC2 C-terminale domain, 1531-1758 aa	228
EF-Hands	9)	S100A2	195
	10)	S100A16	103
	26)	S100A10	97
	27)	Calmodulin Like to Skin Protein (CLSP)	146
	28)	Hippocalcin	193
Hippocalcin target	29)	BIR3-NAIP fragment 278-344 aa	68

Tab. 4.3. Proteins selected to continue the work of thesis.

4.3.1 Destination vectors

The destination vectors reported in figure 4.2 were used to perform the LR recombination reactions. The **pDEST-15** is composed of 7013 base pairs (bp) and contains the gene for the expression of the GST-tag; the **pDEST-17** is composed of 6354 bp and contains the gene for the expression of the $(His)_6$-tag; the **pETG-20A** is composed of 7449 bp and contains the gene for the expression $(His)_6$ + TrxA- tag; the **pETG-30A** is composed of 7784 [bp] and contains the gene for the expression of the $(His)_6$+ GST-tag; the **pETG-60A** is composed of 8597 bp and contains the gene for the expression of the$(His)_6$+ NusA-tag. During the recombination, the fragment included between the *att*R1 and *att*R2 sites, and containing the sequence coding for chloramphenicol resistance, is exchanged. The destination vectors are

ampicillin-resistant in an external region from the *att* sites. T7 *promoter* and T7 *term* identify the sites from which the amplification of the genetic sequences starts, during the PCR reaction.

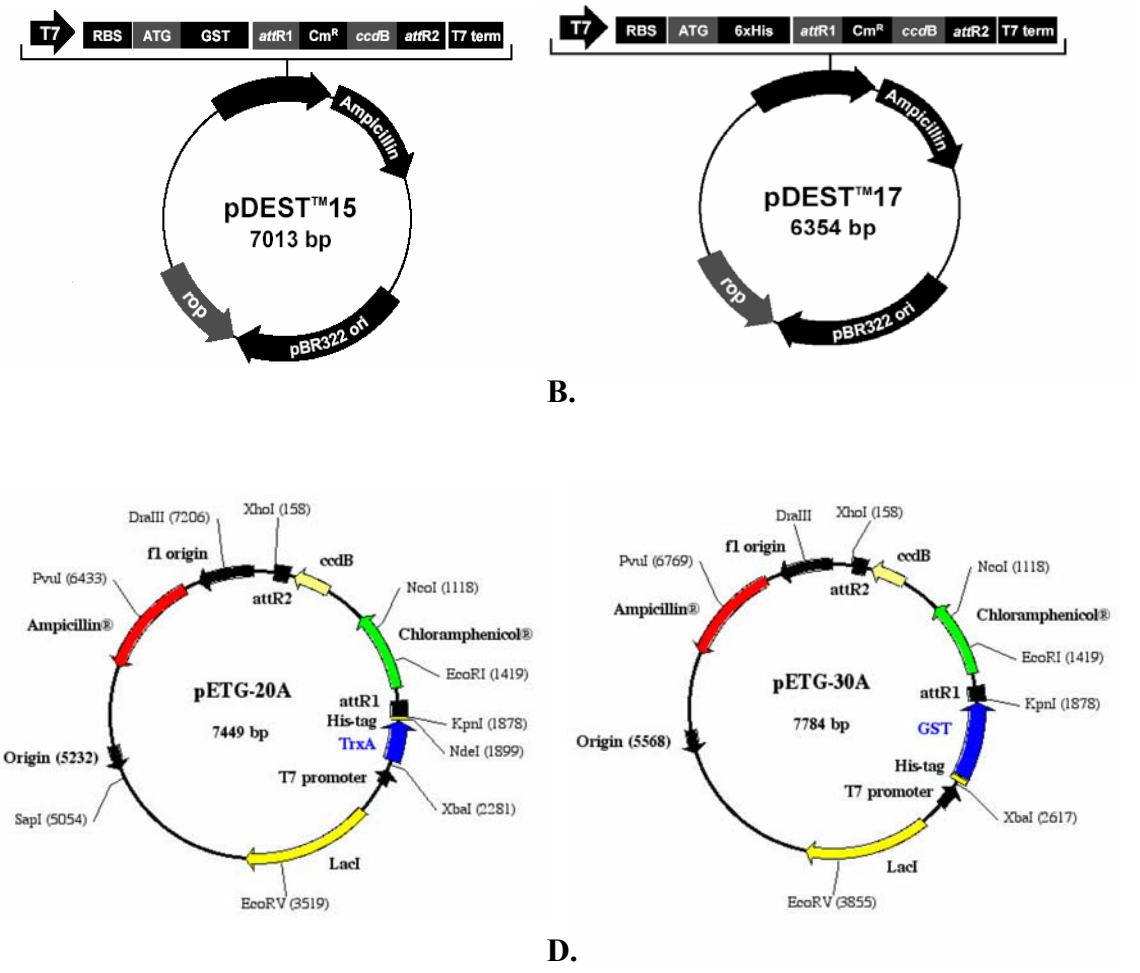


Fig. 4.2 Schematic representations of the destination vectors: **A**=pDEST15; **B**=pDEST17; **C**=pETG-20A; **D**=pETG-30A; **E**=pETG-60A.

It was chosen to clone the proteins in five different destination vectors to test rapidly more vector-host combinations to increase the successful probabilities in protein expression strategies. Additionally, certain affinity tags have a beneficial effect on protein solubility especially in bacterial protein expression. Although the advantages of affinity tags greatly outweigh their potential problems, it is important to consider that any tag can potentially interfere with folding, function or crystallization of the target proteins (Pascal et al,. 2003).

4.3.2. LR recombination reaction and transformation to competent *E.coli* host strains

The **LR recombination** reaction was performed in a way slightly different from that suggested by the Gateway protocol. In fact, each gene coding for the selected proteins, was transferred from the pENTR/ D-TOPO not to a single pDEST, but directly in a single step, to a mixture of 5 destination vectors: pDEST-15; pDEST-17; pETG-20A; pETG-30A; pETG-60A (paragraph4.3.1). This strategy allows a notable reduction of time and consumption of reagents, which is of particular concern considering the high cost of the LR Clonase enzyme. In each reaction mix a quantitatively casual recombination between the pENTR/ D-TOPO and each destination vector was expected to happen. To identify the pDEST containing the genes, the reaction mix of each genes was transformed in three different *E.coli* strains, (BL21GOLD, plys and Codon plus) and plated on selective media containing ampicillin. As the destination vectors contain the gene for the ampicillin-resistance, only the cells able to grow in the presence of this antibiotic were defined “correctly transformed”, having acquired one of the 5 destination vectors. All the transformations in the three *E.coli* host strains produced colonies on selective LB agar plates. 25 colonies were selected from each plate and used for the protein expression screening (paragraph 4.4) by which the identity of the pDEST, acquired from each colony, was established on the basis of the size of the protein produced.

4.4 PROTEIN EXPRESSION SCREENING

To characterize the identity of the plasmid acquired from each transformed colony, the cells were grown in 5 ml of LB medium and after induction with IPTG, the protein expression was analyzed by SDS-page. With the Gateway recombination system, the proteins were expressed fused with other proteins or polypeptides called "tag", to simplify the successive purification steps. Every fusion tag has a specific molecular weight reported in table 4.4.

Destination vectors	pDEST-17	pDEST-15	pETG-20A	pETG-30A	pETG-60A
Fusion tags	$(His)_6$-tag	GST	TrxA+$(His)_6$-tag	GST+$(His)_6$-tag	NusA+$(His)_6$-tag
MW (Dalton)	4300	25000	16650	28977	58857

Tab. 4.4 Molecular Weight, in Dalton, of the fusion tag expressed by the destination vectors pDEST-15; pDEST-17; pETG-20A; pETG-30A; pETG-60A.

The molecular weight of each recombinant protein, fused to each tag, was obtained adding the MW of the fusion tag with the MW of the protein and is reported in tab.4.5.

Proteins	**Fusion TAG**				
	$(His)_6$-tag MM=4300 Da	GST MM=25000 Da	TrxA+$(His)_6$-tag MM=16650 Da	GST+$(His)_6$-tag MM=28977 Da	NusA+$(His)_6$-tag MM=58857 Da
S100A2 MW=10985.2 Da	**15285.2 Da**	**35985.2 Da**	**27635.2 Da**	**39962.2 Da**	**69842.2 Da**
S100A16 MW=11801.4 Da	**16101.4 Da**	**36801.4 Da**	**28451.4 Da**	**40778.4 Da**	**70658.4 Da**
S100A10 MW=11203.1 Da	**15503.1 Da**	**36203.1 Da**	**27853.1 Da**	**40180.1 Da**	**70060.1 Da**
Endothelial differentiation-related factor 1 – isoform α (EDF-1 iso α) MW=16368.6 Da	**20668.6 Da**	**41368.6 Da**	**33018.6 Da**	**45345.6 Da**	**75225.6 Da**
Endothelial differentiation-	**19780.5 Da**	**41480.5 Da**	**32130.5 Da**	**44457.5 Da**	**74337.5 Da**

related factor 1 – isoform β (EDF-1 iso β) MW=15480.5 Da					
Cyclin-dependent kinase inhibitor 1 (P21) - fragment 1-160 (P21-160) MW=17609.6 Da	**21909.6 Da**	**22609.6 Da**	**34259.6 Da**	**46586.6 Da**	**76466.6 Da**
Cyclin-dependent kinase inhibitor 1 (P21) - fragment 1-140 (P21-140) MW=15560.3 Da	**19860 Da**	**41560.3 Da**	**32210.3 Da**	**44537.3 Da**	**74417.3 Da**
SH3 adapter protein SPIN90 - fragment 1-81 (SH3-81) MW=9109.3 Da	**13409.3 Da**	**35109.3 Da**	**25759.3 Da**	**38086.3 Da**	**67966.3 Da**
Tuberous sclerosis 2 protein TSC2_HUMAN C-terminal domain 1531-1758 (TSC2-228) MW=26011.6 Da	**30311.6 Da**	**52011.6 Da**	**42661.6 Da**	**54988.6 Da**	**84868.6 Da**
Hippocalcin MW=22296.2 Da	**26596.2 Da**	**47296.2 Da**	**38946.2 Da**	**51273.2 Da**	**81153.2 Da**
BIR3-NAIP fragment 278-344 aa MW=7610.6 Da	**11910.6 Da**	**32610.6 Da**	**24260.6 Da**	**36587.6 Da**	**66467.6 Da**

Tab. 4.5 Molecular Weigh (expressed in Dalton) of the recombinant proteins fused with each of the tags expressed by destination vectors: pDEST-15; pDEST-17; pETG-20A; pETG-30A; pETG-60A.

The SDS-page protein profile of each culture of the screening was compared to that of protein MW standards. Figure 4.3 shows the SDS PAGE of the two standards used for this expression screening . From this comparative analysis, it was possible to obtain the MW of the protein overexpressed by each culture and then to recognize the pDEST plasmid carrying the gene , of each colony.

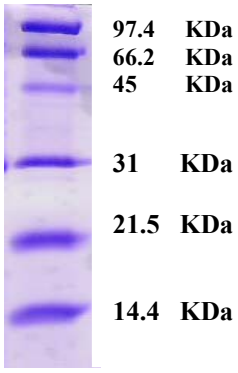
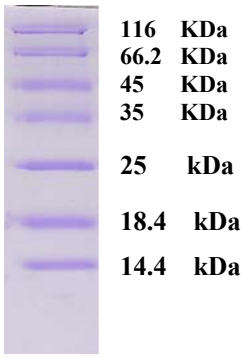
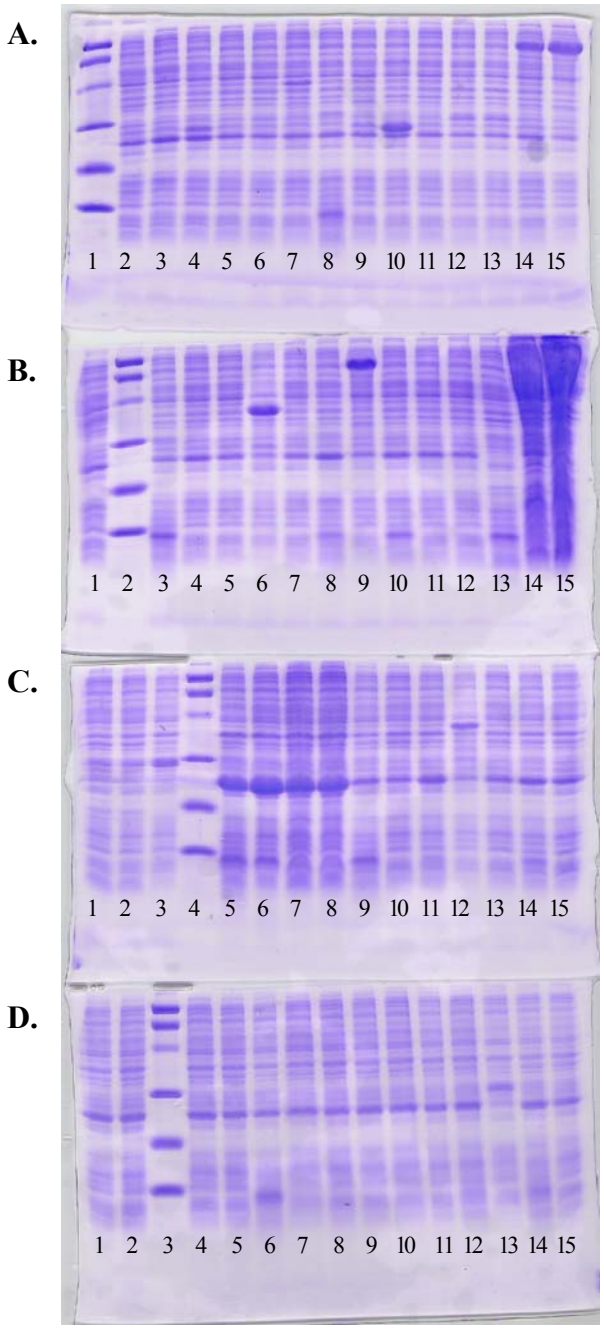
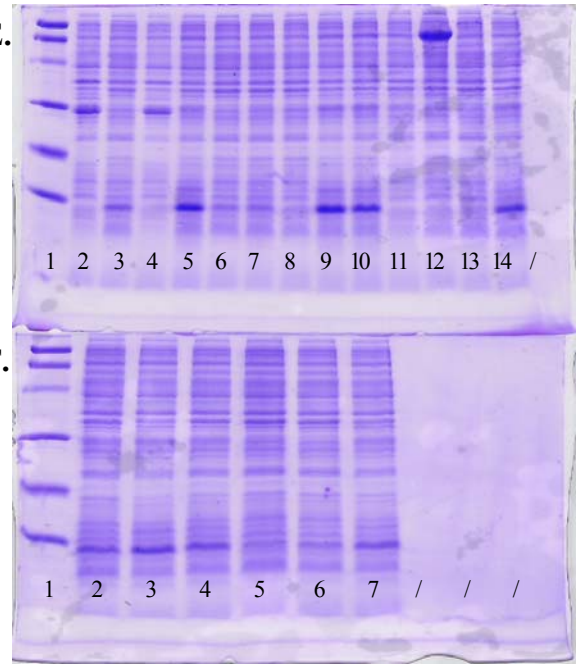
STANDARD 1 Low range	STANDARD 2 Protein Molecular Weight Marker
	

Fig. 4.3 The standards are made of proteins with known molecular weight. The proteins of Low range standard (1) have the following MW: 97.4 kDa, 66.2 kDa, 45 kDa, 31 kDa, 21.5 kDa, 14.4 kDa. The proteins of Protein Molecular Weight Marker (2) have the following MW: 116 kDa, 66.2 kDa, 45 kDa, 35 kDa, 25 kDa, 18.4 kDa, 14.4 kDa.

4.4.1 Expression screening of the protein *S100A2*

The expression of the **S100A2** protein in three *E.coli* BL21 host strains (Codon plus, Plys and Gold) was verified by SDS-PAGE as described in paragraph 3.7. In the figure 4.4 the gels concerning the expression of S100A2 are reported.





	A	B	C	D	E	F
1	st	p15	G2	C+12	st	st
2	p1	st	G3	C+13	G5	G19
3	p2	p16	G4	St	G6	G20
4	p3	p17	st	C+14	G7	G21
5	p4	p18	C+1	C+15	G8	G22
6	p5	p19	C+2	C+16	G9	G23
7	p6	p20	C+3	C+17	G10	G24
8	p7	p21	C+4	C+18	G11	
9	p8	p22	C+5	C+19	G12	
10	p9	p23	C+6	C+20	G14	
11	p10	p24	C+7	C+21	G15	
12	p11	p25	C+8	C+22	G17	
13	p12	G1	C+9	C+23	G18	
14	p13	/	C+10	C+24	G13	
15	p14	/	C+11	C+25	/	

Fig.4.4. SDS-PAGE of the expression of **S100A2** protein in different *E.coli* strains. Page A (line 2-15) and B (lines1, 3-12): expression in *Plys*. Page B (lines 13), C (lines 1-3), D (lines 2-13): expression in *Gold*. In page C (lines 5-15) and D (lines 1-2 and 4-15) is reported the expression in *Codon plus.* The samples loaded in each line are classified with a letter, corresponding to the initial of the host strain (C+= Codon plus, G= Gold, p= pLys), and a number, given to each colony, as showed in table G. The *Low Range* standard in all gels, contains six proteins with known molecular weight (97.4KDa, 66.2KDa, 45KDa, 31KDa, 21.5KDa, 14 KDa).

From the observation of the gels in figure 4.4, it is possible to notice that many samples show an over-expressed band with a specific molecular weight. Comparing the MW of this band, with the hypothetical MW of the recombinant protein S100A2 calculated as described in table 4.5, it is possible to deduce which plasmid has been acquired from each colony. The results of this screening can be summarized as follow:

- the pDEST17 plasmid has been acquired from the colonies *Codon plus* BL21 numbers 16, BL21 *GOLD* numbers 1, 2, 12, 19, 20, 21, 24, 6, 8, 14 and 13 and finally from the colonies BL21 *pLys* numbers 7, 16, 21 and 23. In fact all these colonies present an over-expressed band of around 15000 Da;
- the pETG20A plasmid has been acquired from the colonies BL21*Codon plus* 23, BL21 *GOLD* 5 and 7, and finally from the colonies BL21 *pLys* 9, presenting an over-expressed band of around 27600 Da;
- the pETG30A plasmid has been acquired only from the colony BL21 pLys numbers 19, showing an over-expressed band of around 39900 disposed between the two bands of the standard with MM respectively of 31 kDa and 45 kDa;
- the pETG60A plasmid, instead, has been acquired from the colonies BL21 *pLys* 13, 14 and 22 and from the colony BL21 *GOLD* 16, showing an over-expressed band of around 69800 Da disposed between the two bands of the standard with MM respectively of 66.2 kDa and 97.4 kDa;
- finally, any colony has acquired the pDEST15, since any band of around 36 KDa can be observed on the gel.

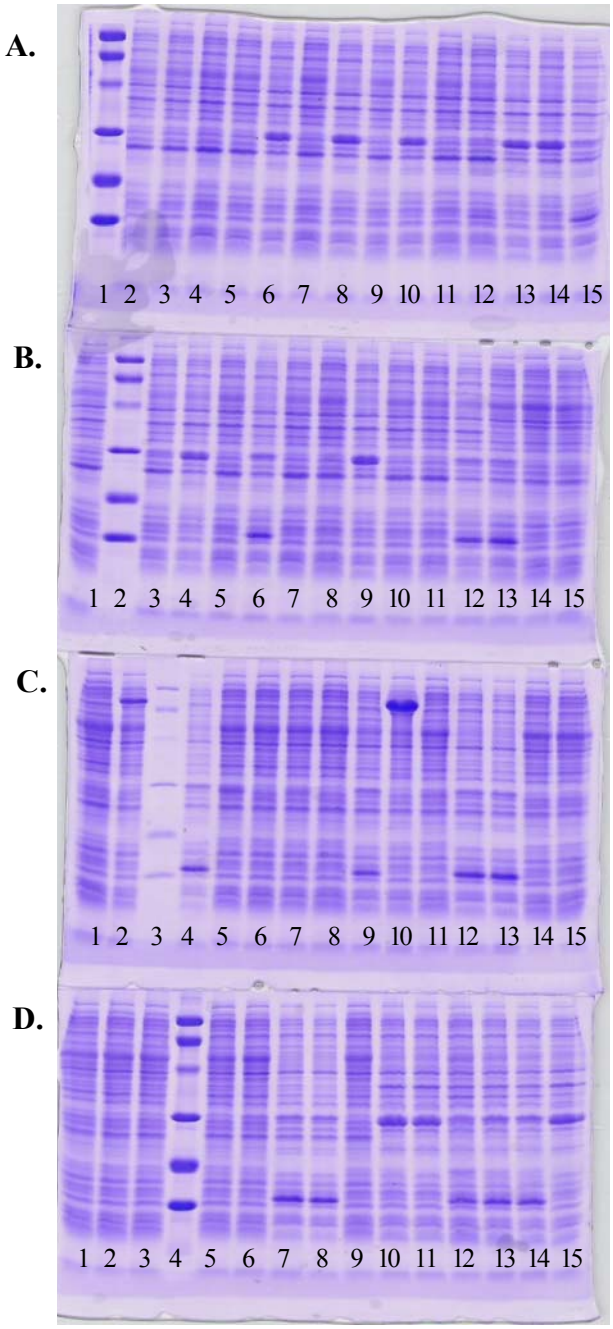
The results of the expression screening of S100A2 protein are summarized in table 4.6.

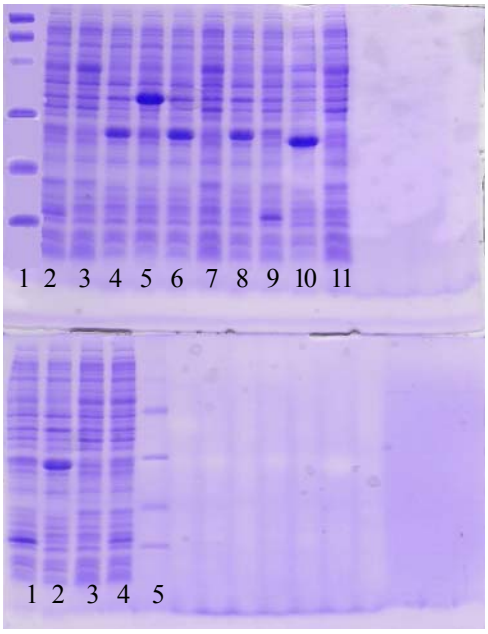
	pDEST15	**pDEST17**	**pETG20A**	**pETG30A**	**pETG60A**
pLys	/	7,16,21,23	9	19	13,14,22
Codon plus	/	16	23	/	/
GOLD	/	1,2,12,19,20,21, 24,6,8,14,13	5,7	/	16

Tab. 4.6 Results of the S100A2 expression screening. The number in red refer to the colonies that were selected to perform the solubility screening (paragraph 4.5).

4.4.2 Expression screening of the protein *S100A16*

The expression of the **S100A16** protein in three *E.coli* BL21 host strains (Codon plus, Plys and Gold) was verified by SDS-PAGE as described in paragraph 3.7. In the figure 4.5 the gels concerning the expression of S100A16 are reported.





G.

	A	B	C	D	E	F
1	st	C+15	p4	p18	st	G20
2	C+1	st	p5	p19	G7	G21
3	C+2	C+16	st	p20	G8	G23
4	C+3	C+17	p6	St	G9	G24
5	C+4	C+18	p7	p21	G10	st
6	C+5	C+19	p8	p22	G11	
7	C+6	C+20	p9	p23	G12	
8	C+7	C+21	p10	p24	G14	
9	C+8	C+22	p11	p25	G15	
10	C+9	C+23	p12	G1	G17	
11	C+10	C+24	p13	G2	G18	
12	C+11	C+25	p14	G3	/	
13	C+12	p1	p15	G4	/	
14	C+13	p2	p16	G5	/	
15	C+14	p3	p17	G6	/	

Fig.4.5 SDS-PAGE of the expression of **S100A16** protein in different *E.coli* strains. Page A (line 2-15) and B (lines1, 3-12): expression in *Codon plus.* Page B (lines 13-15), C (lines 1-2, 4-15), D (lines 1-3, 5-9): expression in *Plys*. In page D (lines 10-15), E (lines 2-11) and F(lines 1-4) is reported the expression in *GOLD.* The samples loaded in each line are classified with a letter, corresponding to the initial of the host strain (C+= Codon plus, G=

Gold, p= pLys), and a number, given to each colony, as showed in table G. The *Low Range* standard in all gels, contains six proteins with known molecular weight (97.4KDa, 66.2KDa, 45KDa, 31KDa, 21.5KDa, 14 KDa).

From the observation of the gels in figure 4.5, it is possible to notice that many samples show an over-expressed band with a specific molecular weight. Comparing the MW of this band, with the hypothetical MW of the recombinant protein S100A16 calculated as described in table 4.5, it is possible to deduce which plasmid has been acquired from each colony. The results of this screening can be summarized as follow:

- the pDEST17 has been acquired from the colonies BL21*Codon plus* numbers 19 and 25, BL21 *GOLD* numbers 3, 4, 5, 20 and finally from the colonies BL21 *pLys* numbers 1, 6, 11, 14, 15, 23, 24, showing an over-expressed band of around 16000 Da;
- the pETG20A plasmid has been acquired from the colonies BL21*Codon plus* 5, 7, 9, 12, 13, 17, 22, BL21 *GOLD* 1, 2, 6, 11, 14, 16 and 21, showing an over-expressed band of around 28400 Da;
- the pETG30A plasmid has been acquired only from the colony BL21 *GOLD* numbers 10, presenting a band of MM of around 39900 disposed between the two bands of the standard with MM respectively of 31 kDa and 45 kDa;
- the pETG60A plasmid, instead, has been acquired from the colonies BL21 *pLys* 13, 14 and 22 and from the colony BL21 *GOLD* 16, showing an over-expressed band of around 69800 Da disposed between the two bands of the standard with MM respectively of 66.2 kDa and 97.4 kDa;
- finally, any colony has acquired the pDEST15, since any band of around 36 KDa can be observed on the gel.

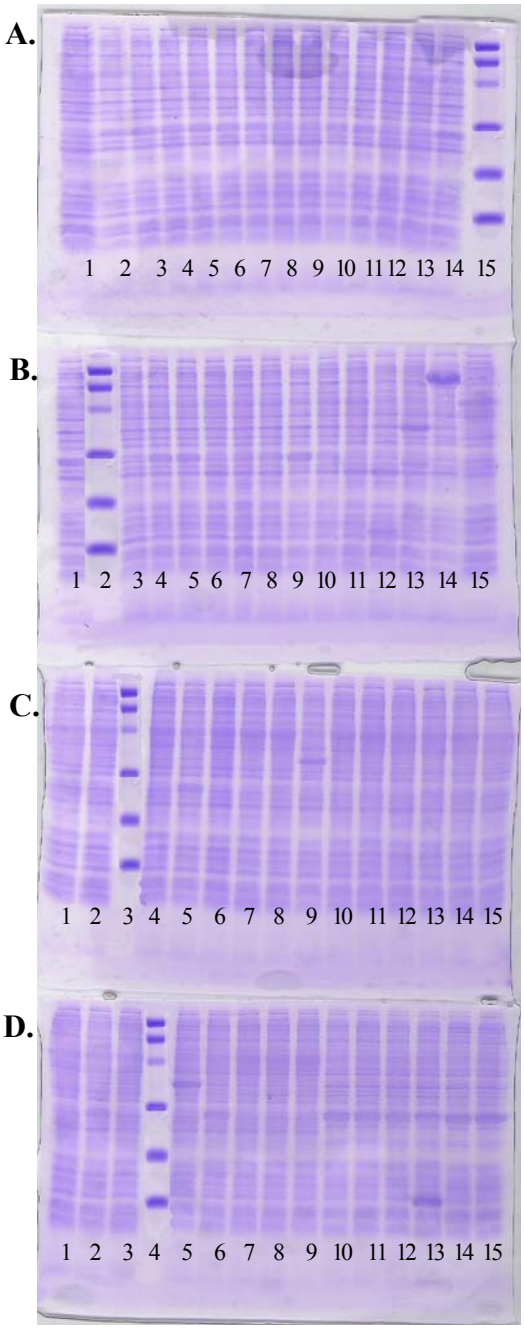
The results of the expression screening of S100A16 protein are summarized in table 4.7.

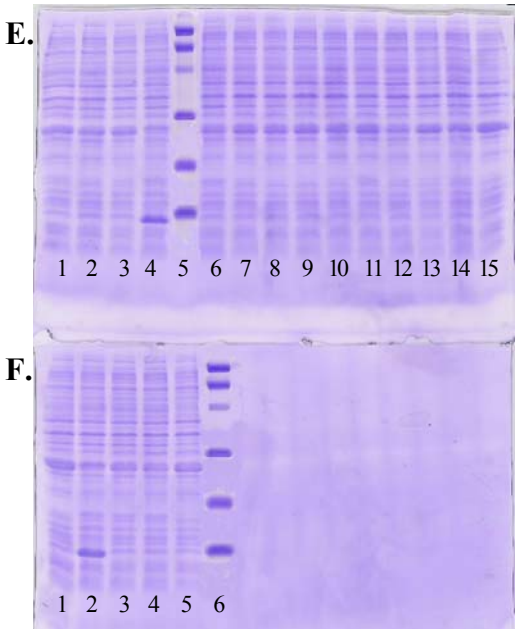
	pDEST15	**pDEST17**	**pETG20A**	**pETG30A**	**pETG60A**
Codon plus	/	19,25	5,7,9,12,13,17,22	/	/
pLys	/	1,6,11,14,15,23,24	/	/	5,12
GOLD	/	3,4,5,20	1,2,6,11,14,16,21	10	/

Tab. 4.7 Results of the S100A16 expression screening. The number in red refer to the colonies that were selected to perform the solubility screening (paragraph 4.5).

4.4.3 Expression screening of the protein *S100A10*

The expression of the **S100A10** protein in three *E.coli* BL21 host strains (Codon plus, Plys and Gold) was verified by SDS-PAGE as described in paragraph 3.7. In the figure 4.6 the gels concerning the expression of S100A10 are reported.





G.

	A	B	C	D	E	F
1	C+1	C+15	p4	p18	G7	G21
2	C+2	st	p5	p19	G8	G22
3	C+3	C+16	st	p20	G9	G23
4	C+4	C+17	p6	St	G10	G24
5	C+5	C+18	p7	p21	st	G25
6	C+6	C+19	p8	p22	G11	st
7	C+7	C+20	p9	p23	G12	
8	C+8	C+21	p10	p24	G13	
9	C+9	C+22	p11	p25	G14	
10	C+10	C+23	p12	G1	G15	
11	C+11	C+24	p13	G2	G16	
12	C+12	C+25	p14	G3	G17	
13	C+13	p1	p15	G4	G18	
14	C+14	p2	p16	G5	G19	
15	st	p3	p17	G6	G20	

Fig. 4.6 SDS-PAGE of the expression of **S100A10** protein in different *E.coli* strains. Page A (line 2-15) and B (lines1, 3-12): expression in *Codon plus.* Page B (lines 13-15), C (lines 1-2, 4-15), D (lines 1-3, 5-9): expression in *Plys*. In page D (lines 10-15), E (lines 1-4, 6-15) and F(lines 1-5) is reported the expression in *GOLD.* The samples loaded in each line are classified with a letter, corresponding to the initial of the host strain (C+= Codon plus, G= Gold, p= pLys), and a number, given to each colony, as showed in table G. The *Low Range* standard in

all gels, contains six proteins with known molecular weight (97.4KDa, 66.2KDa, 45KDa, 31KDa, 21.5KDa, 14 KDa).

From the observation of the gels in figure 4.6, it is possible to notice that many samples show an over-expressed band with a specific molecular weight. Comparing the MW of this band, with the hypothetical MW of the recombinant protein S100A10 calculated as described in table 4.5, it is possible to deduce which plasmid has been acquired from each colony. The results of this screening can be summarized as follow:

- The pDEST17 has been acquired from the colonies BL21*Codon plus* numbers 4, 10 and 22 showing an overexpressed band of around 15500 Da;
- the pETG20A plasmid has been acquired from the colonies BL21*Codon plus* 22, BL21 and BL21 pLys 7 and 22, showing an over-expressed band of around 27800 Da;
- the pETG30A plasmid has been acquired only from the colony BL21 *pLys* 1, 11 and 21, BL21 *Codon plus* 8 and 9, showing an over-expressed band of around 40100 Da disposed between the two bands of the standard with MM respectively of 31 kDa and 45 kDa;
- the pETG60A plasmid, instead, has been acquired from the colonies BL21 *pLys* 2, showing an over-expressed band of around 70000 Da disposed between the two bands of the standard with MM respectively of 66.2 kDa and 97.4 kDa;
- Finally, any colony has acquired the pDEST15, since any band of around 36 KDa can be observed on the gel.

The results of the expression screening of S100A10 protein are summarized in table 4.8.

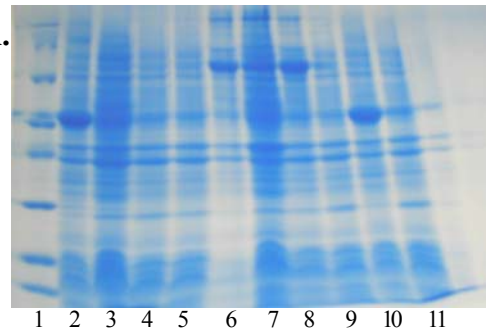
	pDEST15	**pDEST17**	**pETG20A**	**pETG30A**	**pETG60A**
Codon plus	/	/	22	8, 9	/
pLys	/	/	7, 22	1, 11, 21	2
GOLD	/	4, 10, 22	1	/	/

Tab. 4.8 Results of the S100A10 expression screening. The number in red refer to the colonies that were selected to perform the solubility screening (paragraph 4.5).

4.4.4 Expression screening of the isoform α of the Endothelial differentiation-related factor (EDF-1 iso α) protein

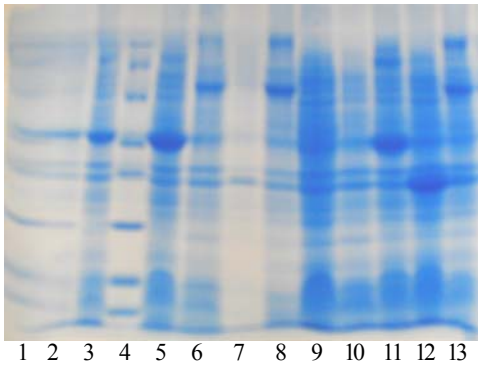
The expression of the **EDF-1 iso α** protein in three *E.coli* BL21 host strains (Codon plus, Plys and Gold) was verified by SDS-PAGE as described in paragraph 3.7. In the figure 4.7 the gels concerning the expression are reported.

A.



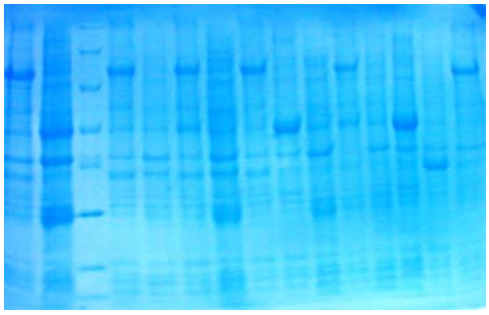
1 2 3 4 5 6 7 8 9 10 11

B.



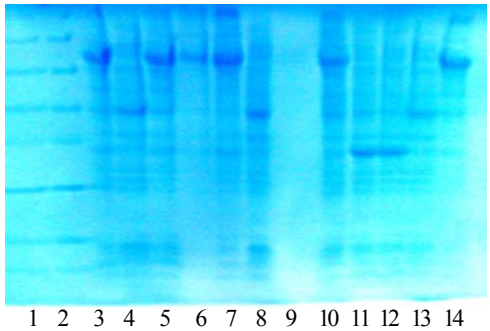
1 2 3 4 5 6 7 8 9 10 11 12 13

C.



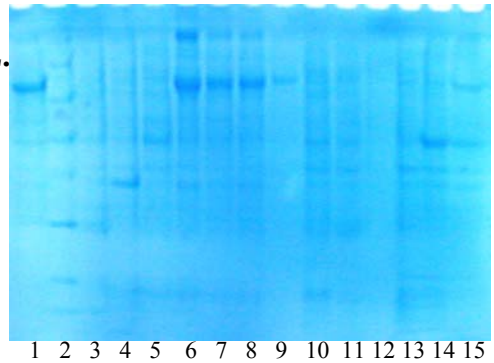
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

D.



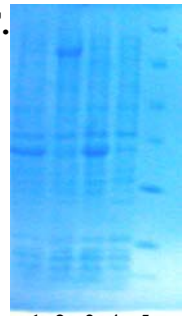
1 2 3 4 5 6 7 8 9 10 11 12 13 14

E.



1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

F.



1 2 3 4 5

G.

	A	B	C	D	E	F
1	st	G11	C+4	St	p13	C+18
2	G1	st	C+5	St	St	C+19
3	G2	G12	st	p1	p14	C+20
4	G3	G13	C+6	p2	p15	C+21
5	G4	G14	C+7	p3	p16	st
6	G5	G15	C+8	p4	p17	
7	G6	G16	C+9	p5	p18	
8	G7	G17	C+10	p6	p19	
9	G8	G18	C+11	p7	p20	
10	G9	G19	C+12	p8		
11	G10	G20	C+13	p9		
12	/	/	C+14	p10		
13	/	/	C+15	p11		
14	/	/	C+16	p12		
15	/	/	C+17	/		

Fig. 4.7 SDS-PAGE of the expression of **EDF-1 iso α** protein in different *E.coli* strains. Page A and B: expression in *GOLD*. Page C and F: expression in *Codon plus*. In page D and E is reported the expression in *Plys*. The samples loaded in each line are classified with a letter, corresponding to the initial of the host strain (C+= Codon plus, G= Gold, p= pLys), and a number, given to each colony, as showed in table G. The Protein Molecular Weight Marker Standard in all gels, contains seven proteins with known molecular weight (11.6KDa, 66.2KDa, 45KDa, 35KDa, 25KDa, 18.4 KDa, 14.4KDa).

From the observation of the gels in figure 4.7, it is possible to notice that many samples show an over-expressed band with a specific molecular weight. Comparing the MW of this band, with the hypothetical MW of the recombinant protein EDF-1 iso α calculated as described in table 4.5, it is possible to deduce which plasmid has been acquired from each colony. The results of this screening can be summarized as follow:

- any colony has acquired the pDEST17, since it's not observed any bands of around 18400 Da;
- the pETG20A plasmid has been acquired from the colonies BL21 Codon plus 5,9,12,16,18,20, from the colonies BL21 pLys 9,10,15 and from the colonies BL21 Glold 10 and 19, showing an over-expressed band of around 33 kDa;
- the pETG30A plasmid has been acquired from the colony Codon plus 3, 11, 15 from the colonies BL21 pLys 2, 6, 11 and from the colonies BL21 Glold 1, 9, 11, 12, 18, showing an over-expressed band of around 45300 Da;
- the pETG60A plasmid has been acquired from the colonies BL21 Codon plus 3, 4, 6, 8, 10, 13, 17, 19 from the colonies BL21 pLys 1, 3, 4, 5, 8,12,13,17 and from the colonies BL21 Glold 5, 6, 7, 13, 15 and 20, showing an over-expressed band of around 75 kDa, disposed between the two bands of the standard with MM respectively of 116 kDa and 66.2 kDa;
- finally, any colony has acquired the pDEST15, since any band of around 41 KDa can be observed on the gel.

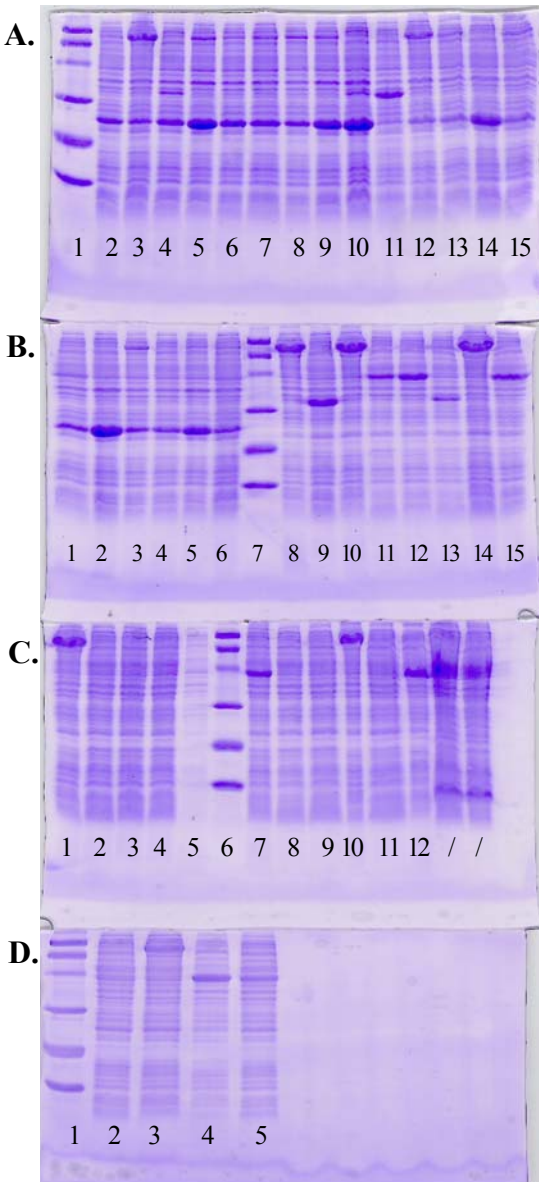
The results of the expression screening of **EDF-1 iso α** protein are summarized in table 4.9.

	pDEST15	**pDEST17**	**pETG20A**	**pETG30A**	**pETG60A**
Codon plus	/	/	5,9,12,16, 18,20	3,11,15	3,4,6,8,10,13,17,19
GOLD	/	/	10,19	1,9,11,12,18	5,6,7,13,15,20
pLys	/	/	9,10,15	2,6,11	1,3,4,5,8,12,13,17 18,19,20

Tab. 4.9 Results of the S100A10 expression screening. The number in red refer to the colonies that were selected to perform the solubility screening (paragraph 4.5).

4.4.5 Expression screening of the isoform β Endothelial differentiation-related factor 1 (EDF-1 iso β) protein

The expression of the **EDF-1 iso β** protein in three *E.coli* BL21 host strains (Codon plus, Plys and Gold) was verified by SDS-PAGE as described in paragraph3.7. In the figure 4.8 the gels concerning the expression are reported.



E.

	A	B	C	D
1	St	C+15	G9	St
2	C+1	C+16	G10	P8
3	C+2	C+17	G11	P12
4	C+3	C+18	G12	P13
5	C+4	C+19	G13	P18
6	C+5	C+20	St	
7	C+6	St	G15	
8	C+7	G1	G16	
9	C+8	G2	G17	
10	C+9	G3	G18	
11	C+10	G4	G19	
12	C+11	G5	G20	
13	C+12	G6	/	
14	C+13	G7	/	
15	C+14	G8	/	

Fig. 4.8 SDS-PAGE of the expression of EDF-1 iso β protein in different *E.coli* strains. Page A and B (lines 1-6): expression in *Codon plus.* Page B (lines 8-15) and C (lines 1-5): expression in *GOLD*. In page D is reported the expression in *Plys.* The samples loaded in each line are classified with a letter, corresponding to the initial of the host strain (C+= Codon plus, G= Gold, p= pLys), and a number, given to each colony, as showed in table E. The *Low Range* standard in all gels, contains six proteins with known molecular weight (97.4KDa, 66.2KDa, 45KDa, 31KDa, 21.5KDa, 14 KDa).

From the observation of the gels in figure 4.8, it is possible to notice that many samples show an over-expressed band with a specific molecular weight. Comparing the MW of this band, with the hypothetical MW of the recombinant protein EDF-1 iso β calculated as described in table 4.5, it is possible to deduce which plasmid has been acquired from each colony. The results of this screening can be summarized as follow:

- any colony has acquired the pDEST15 and pDEST17, since any bands with MW compatible with the recombinant protein can be observed one;
- the pETG20A plasmid has been acquired from the colonies BL21 Codon plus 10 and from the colonies BL21 Gold 2 and 6, showing an ove-expressed band of around 32 kDa;
- the pETG30A plasmid has been acquired from the colonies BL21 pLys 13 and 18 and from the colonies BL21 Gold 4, 5, 8, 15, 20 showing an over-expressed band of around 44450 Da;
- the pETG60A plasmid has been acquired from the colonies BL21 Codon plus 2, 11, 17, from the colony BL21 pLys 12 and from the colonies BL21 Glold 1, 3, 7, 9, and 18, showing an over-expressed band of around 75 kDa, disposed between the two bands of the standard with MW respectively of 116 kDa and 66.2 kDa.

The results of the expression screening of **EDF-1 iso β** protein are summarized in table 4.10.

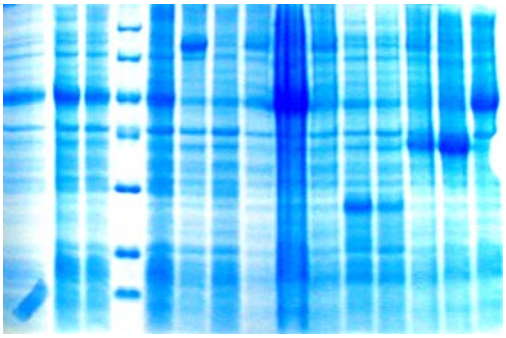
	pDEST15	**pDEST17**	**pETG20A**	**pETG30A**	**pETG60A**
Codon plus	/	/	10	/	2,11,17
GOLD	/	/	2,6	4,5,8 15,20	1,3,7 9,18
pLys	/	/	/	13,18	12

Tab. 4.10 Results of the EDF-1 iso β expression screening. The number in red refer to the colonies that were selected to perform the solubility screening (paragraph 4.5).

4.4.6 Expression screening of a fragment of Cyclin-dependent kinase inhibitor 1 (P21), aa 1-160, (P21-160)

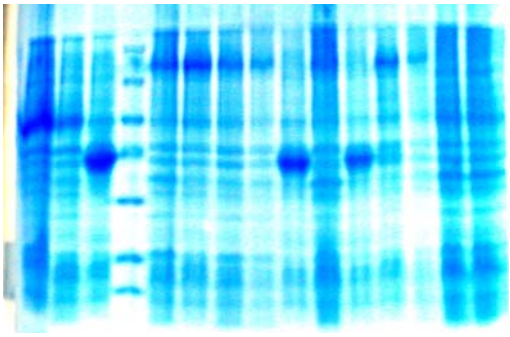
The expression of the **P21-160** protein in three *E.coli* BL21 host strains (Codon plus, Plys and Gold) was verified by SDS-PAGE as described in paragraph3.7. In the figure 4.9 the gels concerning the expression are reported

A.



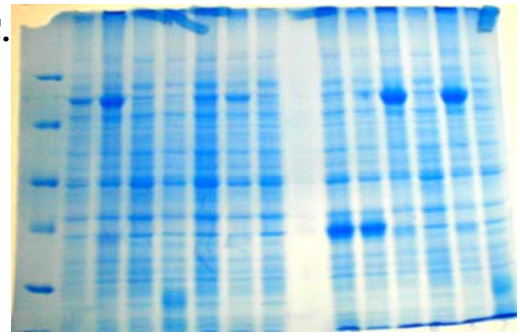
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

B.



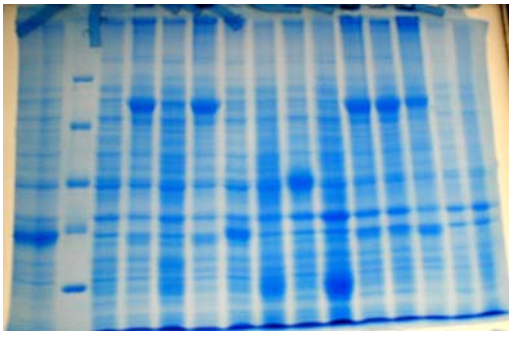
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

C.



1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

D.



1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

E.

	A	B	C	D
1	C+5	G9	St	p15
2	C+6	G10	p1	St
3	C+7	G11	p2	p16
4	st	St	p3	p17
5	C+8	G12	p4	p18
6	C+9	G13	p5	p19
7	C+17	G14	p6	p20
8	G1	G15	p7	p21
9	G2	G16	p8	C+10
10	G3	G17	p9	C+11
11	G4	G18	p10	C+12
12	G5	G19	p11	C+13
13	G6	G20	p12	C+14
14	G7	G21	p13	C+15
15	G8	G22	p14	C+16

Fig. 4.9 SDS-PAGE of the expression of **P21-160** protein in different *E.coli* strains. Page A (lines 1-3, 5-7) and D (lines 9-15): expression in *Codon plus*. Page A (lines 8-15) and B (lines 1-3, 5-15): expression in *GOLD*. In page C and D (lines 1, 3-8) is reported the expression in *Plys*. The samples loaded in each line are classified with a letter, corresponding to the initial of the host strain (C+= Codon plus, G= Gold, p= pLys), and a number, given to each colony, as showed in table E. The Protein Molecular Weight Marker Standard in all gels, contains seven proteins with known molecular weight (11.6KDa, 66.2KDa, 45KDa, 35KDa, 25KDa, 18.4 KDa, 14.4KDa).

From the observation of the gels in figure 4.9, it is possible to notice that many samples show an over-expressed band with a specific molecular weight. Comparing the MW of this band, with the hypothetical MW of the recombinant protein P21-160 calculated as described in table 4.5, it is possible to deduce which plasmid has been acquired from each colony. The results of this screening can be summarized as follow:

- the pDEST17 plasmid has been acquired from the colonies BL21 Codon plus 11, from the colonies BL21 Gold 4, 5, and from the colony BL21 pLys 14, showing an over-expressed band of around 22 kDa;
- the pETG20A plasmid has been acquired from the colonies BL21 Gold 6, 7, 11, 16 and 18, and from the colony BL21 pLys 9, 10, 15 and 20, showing an over-expressed band of around 34.2 kDa;
- the pETG30A plasmid has been acquired from the colonies BL21 Codon plus 6, 7, 8, 10 and from the colonies BL21 Gold 2, 8, 9 and 10, showing an over-expressed band of around 47 kDa;
- the pETG60A plasmid has been acquired from the colonies BL21 Codon plus 9, 12, 13 14, from the colony BL21 pLys 1, 2, 5, 6, 11, 13, 17, 19 and from the colonies BL21 Glold 1, 3, 12,13, 14 and 15, showing an over-expressed band of around 76 kDa, disposed between the two bands of the standard with MW respectively of 116 kDa and 66.2 kDa.

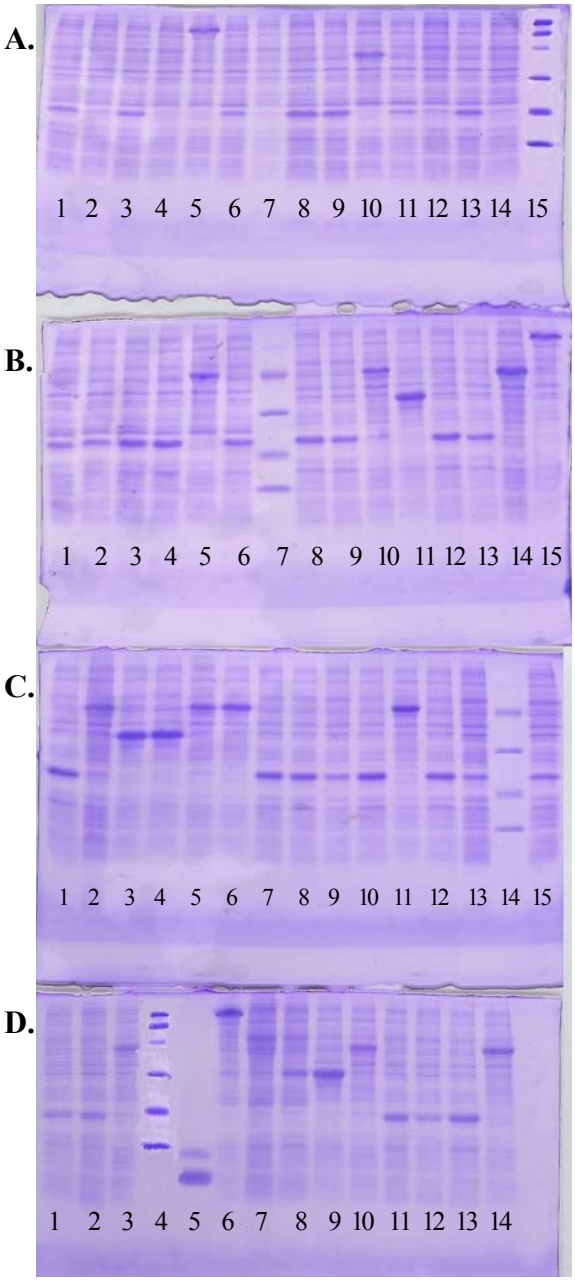
The results of the expression screening of **P21-160** protein are summarized in table 4.11.

	pDEST17	**pETG20A**	**pETG30A**	**pETG60A**
Codon plus	11	/	6,7,8,10	9,12,13,14
GOLD	4,5	6,7,11,16,18	2,8,9,10	1,3,12,13,14 15
pLys	14	9,10,15,20	/	1,2,5,6,11 13,17,19

Tab. 4.11 Results of the P21-160 expression screening. The number in red refer to the colonies that were selected to perform the solubility screening (paragraph 4.5).

4.4.7 Expression screening of a fragment of Cyclin-dependent kinase inhibitor 1 (P21), aa 25-164, (P21-140)

The expression of the **P21-140** protein in three *E.coli* BL21 host strains (Codon plus, Plys and Gold) was verified by SDS-PAGE as described in paragraph3.7. In the figure 4.10 the gels concerning the expression are reported.



E.

	A	B	C	D
1	C+1	C+15	G9	G23
2	C+2	C+16	G10	G24
3	C+3	C+17	G11	G25
4	C+4	C+18	G12	St
5	C+5	C+19	G13	/
6	C+6	C+20	G14	p1
7	C+7	St	G15	p2
8	C+8	G1	G16	p3
9	C+9	G2	G17	G12
10	C+10	G3	G18	G13
11	C+11	G4	G19	G16
12	C+12	G5	G20	G22
13	C+13	G6	G21	G17
14	C+14	G7	St	G18
15	St	G8	G22	/

Fig. 4.10. SDS-PAGE of the expression of P21-140 protein in different *E.coli* strains. Page A and B (lines 1-6): expression in *Codon plus*. Page B (lines 8-15), page C and D (lines 1-3, 9-14): expression in *GOLD*. In page D is reported the expression in *Plys* (lines 6-8). The samples loaded in each line are classified with a letter, corresponding to the initial of the host strain (C+= Codon plus, G= Gold, p= pLys), and a number, given to each colony, as showed in table E. The *Low Range* standard in all gels, contains six proteins with known molecular weight (97.4KDa, 66.2KDa, 45KDa, 31KDa, 21.5KDa, 14 KDa).

From the observation of the gels figure 4.10, it is possible to notice that many samples show an over-expressed band with a specific molecular weight. Comparing the MW of this band, with the hypothetical MW of the recombinant protein P21-140 calculated as described in table 4.5, it is possible to deduce which plasmid has been acquired from each colony. The results of this screening can be summarized as follow:

- any colony has acquired the pDEST17 plasmid, since any band with MW compatible with the recombinant protein can be observed on the gel.
- the pDEST15 plasmid has been acquired from the colonies BL21 Gold 4, 11, 12, and from the colony BL21 pLys 3, showing an over-expressed band of around 41.5 kDa;
- the pETG20A plasmid has been acquired from the colonies BL21 Codon plus 1, 3, 6, 8, 9, 11, 13, 15, 16,17 18, 20 and from the colonies BL21 GOLD 1, 2, 5, 6, 9, 15, 16, 17, 19, 20, 21, 22, 23, 24, showing an over-expressed band of around 32.2 kDa;
- the pETG30A plasmid has been acquired from the colonies BL21 Codon plus 10 and 19, and from the colonies BL21 GOLD 3, 7, 8, 10, 13, 14, 18 and 25, showing an over-expressed band of around 45 kDa;
- at last, the pETG60A plasmid has been acquired from the colonies BL21 Codon plus 5, from the colony BL21 pLys 1, and from the colonies BL21 Glold 8, showing an over-expressed band of around 74.4 kDa, disposed between the two bands of the standard with MW respectively of 116 kDa and 66.2 kDa.

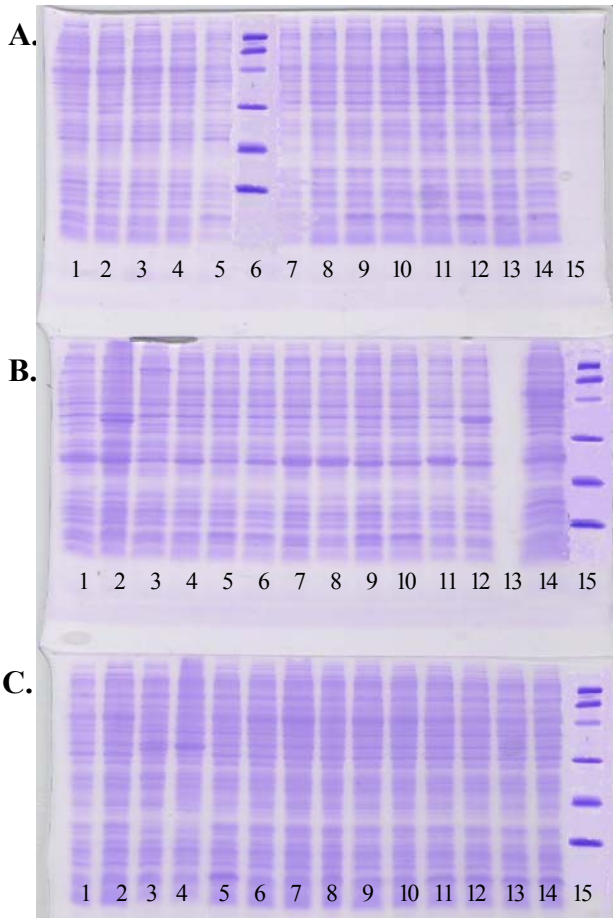
The results of the expression screening of **P21-140** protein are summarized in table 4.12.

	pDEST15	**pDEST17**	**pETG20A**	**pETG30A**	**pETG60A**
Codon plus	/	/	1,3,6,8,9,11,13 15,16,17,18,20	10,19	5
GOLD	4,11,12	/	1,2,5,6,9,15,16,17 19,20,21,22,23,24	3,7,8,10 13,14,18,25	8
pLys	3	/	/	/	1

Tab. 4.12 Results of the P21-140 expression screening. The number in red refer to the colonies that were selected to perform the solubility screening (paragraph 4.5).

4.4.8 Expression screening of a fragment of the SH3 adapter protein SPIN90, aa 1- 81, (SH3-81)

The expression of the **SH3 adapter protein SPIN90** protein in three *E.coli* BL21 host strains (Codon plus, Plys and Gold) was verified by SDS-PAGE as described in paragraph 3.7. In the figure 4.11 the gels concerning the expression are reported.



D.

	A	B	C
1	P3	C+1	G10
2	P4	C+2	G11
3	P5	C+3	G12
4	P6	C+4	G13
5	P7	C+5	G14
6	St	C+6	G15
7	G1	C+7	G16
8	G2	C+8	G17
9	G3	C+9	G18
10	G4	C+10	G19
11	G5	C+11	G20
12	G7	C+12	G21
13	G8	/	G22
14	G9	P1	G23
15	P2	St	St

Fig. 4.11. SDS-PAGE of the expression of SH3 adapter protein SPIN90 protein in different *E.coli* strains. Page A and B (lines 1-6): expression in *Codon plus.* Page B (lines 8-15), page C and D (lines 1-3, 9-14): expression in *GOLD*. In page D is reported the expression in *Plys* (lines 6-8 The samples loaded in each line are classified with a letter, corresponding to the initial of the host strain (C+= Codon plus, G= Gold, p= pLys), and a number, given to each colony, as showed in table D. The *Low Range* standard in all gels, contains six proteins with known molecular weight (97.4KDa, 66.2KDa, 45KDa, 31KDa, 21.5KDa, 14 KDa).

From the observation of the gels in figure 4.11, it is possible to notice that many samples show an over-expressed band with a specific molecular weight. Comparing the MW of this band, with the hypothetical MW of the recombinant SH3 adapter protein SPIN90 calculated as described in table 4.5, it is possible to deduce which plasmid has been acquired from each colony. The results of this screening can be summarized as follow:

- the pDEST17 plasmid, has been acquired from the colonies BL21 Codon plus 110, from the colonies BL21 GOLD 7, 14, 23, and from the colony BL21 pLys 7, showing an over-expressed band of around 13.4 kDa;
- the pDEST15 plasmid has been acquired from the colonies BL21 Codon plus 2, 12 and from the colonies BL21 GOLD 11, 12, 13, showing an over-expressed band of around 35 kDa;
- the pETG20A plasmid has been acquired from the colonies BL21 Codon plus 7, 8, 11, showing an over-expressed band of around 26 kDa;
- the pETG30A plasmid has been acquired from the colony BL21 pLys 2, showing an over-expressed band of around 38 kDa;
- finally, the pETG60A plasmid has been acquired from the colony BL21 Codon plus 3, showing an over-expressed band of around 68 kDa, disposed between the two bands of the standard with MW respectively of 116 kDa and 66.2 kDa.

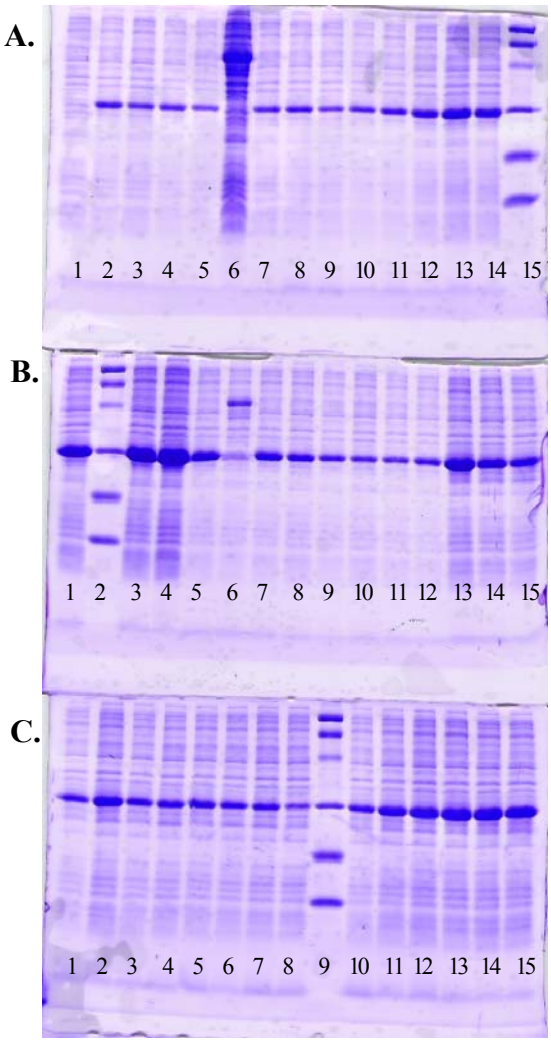
The results of the expression screening of **SH3 adapter protein SPIN90** protein are summarized in table 4.13.

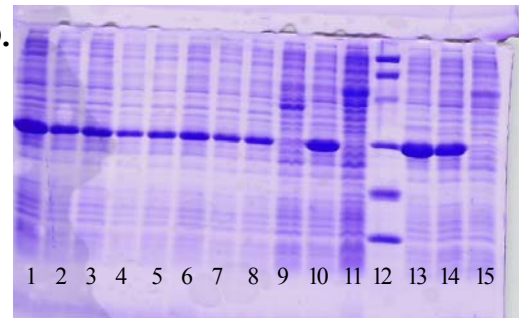
	pDEST15	**pDEST17**	**pETG20A**	**pETG30A**	**pETG60A**
Codon plus	2,12	10	7,8,11	/	3
GOLD	11,12,13	7,14,23	/	/	/
pLys	/	7	/	2	/

Tab. 4.13 Results of the SH3 adapter protein SPIN90 expression screening. The number in red refer to the colonies that were selected to perform the solubility screening (paragraph 4.5).

4.4.9 Expression screening of the C-terminal domain Tuberous sclerosis 2 protein TSC2_HUMAN, aa 1531 – 1758, (TSC2-228)

The expression of the TSC2-228 protein in three *E.coli* BL21 host strains (Codon plus, Plys and Gold) was verified by SDS-PAGE as described in paragraph3.7. In the figure 4.12 the gels concerning the expression are reported.





E.

	A	B	C	D
1	G1	G15	C+17	C+18
2	G2	St	C+4	C+19
3	G3	G16	C+5	C+20
4	G4	G17	C+6	C+21
5	G5	G18	C+7	C+22
6	G6	G19	C+8	C+23
7	G7	G20	C+9	C+24
8	G8	G21	C+10	C+25
9	G9	G22	St	p1
10	G10	G23	C+11	p2
11	G11	G24	C+12	p3
12	G12	G25	C+13	St
13	G13	C+1	C+14	P4
14	G14	C+2	C+15	P5
15	St	C+3	C+16	P6

Fig. 4.12 SDS-PAGE of the expression of TSC2-228 protein in different *E.coli* strains. Page B (lines 13-15), C and D (lines 1-8): expression in *Codon plus.* Page A and B (lines 1, 3-12): expression in *GOLD*. In page D is reported the expression in *Plys* (lines 9-11, 13-15). The samples loaded in each line are classified with a letter, corresponding to the initial of the host strain (C+= Codon plus, G= Gold, p= pLys), and a number, given to each colony, as showed in table E. The *Low Range* standard in all gels, contains six proteins with known molecular weight (97.4KDa, 66.2KDa, 45KDa, 31KDa, 21.5KDa, 14 KDa).

From the observation of the gels in figure 4.12, it is possible to notice that many samples show an over-expressed band with a specific molecular weight. Comparing the MW of this band, with the hypothetical MW of the recombinant protein **TSC2** calculated as described in

table 4.5, it is possible to deduce which plasmid has been acquired from each colony. The results of this screening can be summarized as follow:

- any colony has acquired the pDEST15 plasmid, since any bands with MW compatible with the recombinant protein can be observed on the gel.
- the pDEST17 plasmid has been acquired from all the colonies BL21 Codon plus, from the colony BL21 GOLD 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, and from the colonies BL21 pLys 4 and 5, showing an over-expressed band of around 30 kDa;
- the pETG20A plasmid has been acquired from the colony BL21 GOLD 6 and from the colony BL21 plys 1, showing an over-expressed band of around 45 kDa;
- the pETG30A plasmid has been acquired from the colony BL21 GOLD 19 and from the colonies BL21 plys 3 and 6, showing an over-expressed band of around 45 kDa;
- finally, any colony has acquired the pETG-60A plasmid, since any bands with MW compatible with the recombinant protein can be observed on the gel.

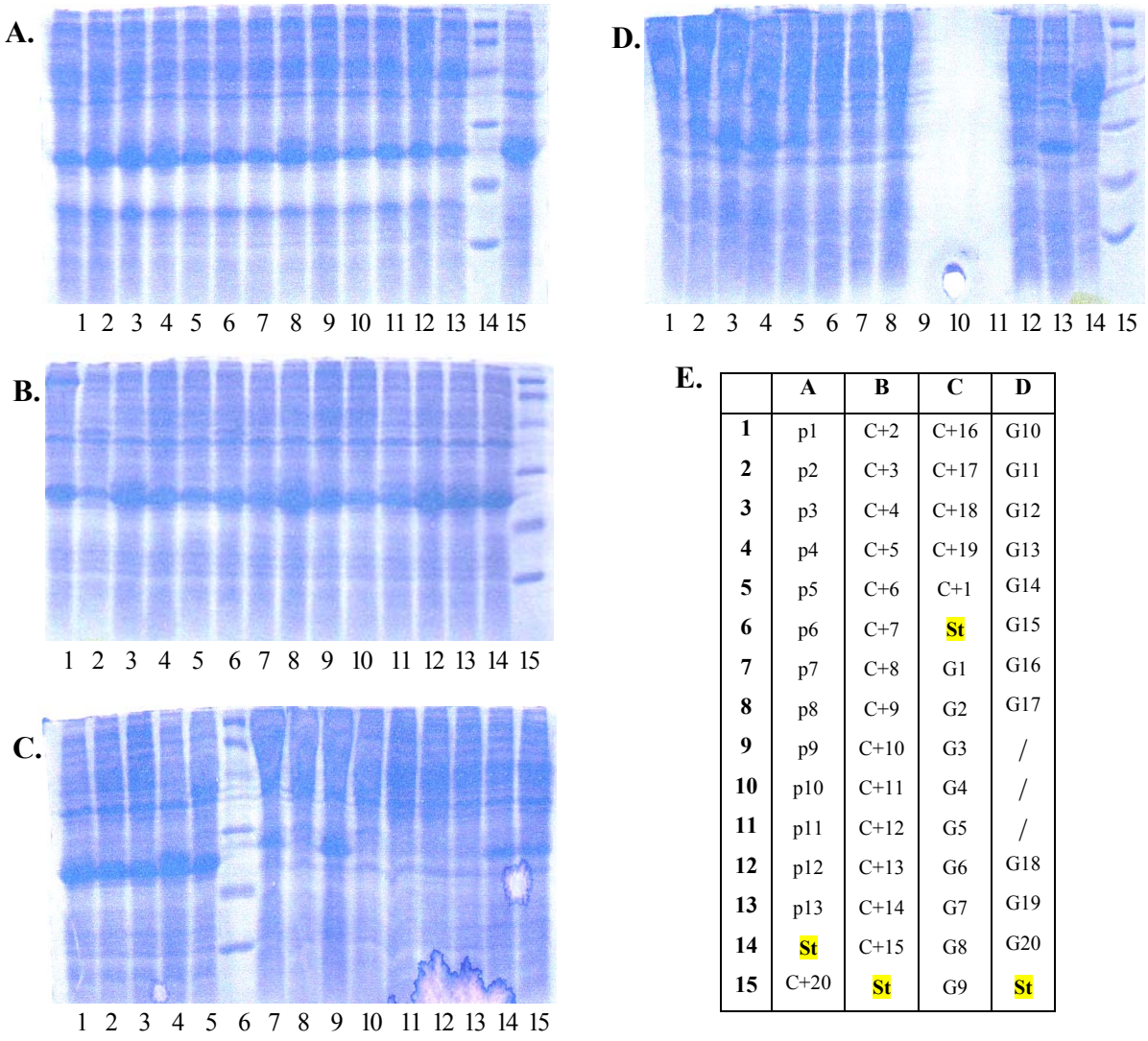
The results of the expression screening of TSC2-228 protein are summarized in table 4.14.

	pDEST15	pDEST17	pETG20A	pETG30A	pETG60A
Codon plus	/	1,2,3,4,5,6,7,8,9,10,11,12 ,13,14,15,16,17,18 19,20,21,22,23,24,25	/	/	/
GOLD	/	2,3,4,5,7,8,9,1011, 12,13,14,15	6	19	/
pLys	/	4,5	1	3,6	/

Tab. 4.14 Results of the TSC2-228 expression screening. The number in red refer to the colonies that were selected to perform the solubility screening (paragraph 4.5).

4.4.10 Expression screening of the protein Hippocalcin

The expression of the **Hippocalcin** protein in three *E.coli* BL21 host strains (Codon plus, Plys and Gold) was verified by SDS-PAGE as described in paragraph 3.7. In the figure 4.13 the gels concerning the expression are reported.



E.

	A	B	C	D
1	p1	C+2	C+16	G10
2	p2	C+3	C+17	G11
3	p3	C+4	C+18	G12
4	p4	C+5	C+19	G13
5	p5	C+6	C+1	G14
6	p6	C+7	**St**	G15
7	p7	C+8	G1	G16
8	p8	C+9	G2	G17
9	p9	C+10	G3	/
10	p10	C+11	G4	/
11	p11	C+12	G5	/
12	p12	C+13	G6	G18
13	p13	C+14	G7	G19
14	**St**	C+15	G8	G20
15	C+20	**St**	G9	**St**

Fig. 4.13 SDS-PAGE of the expression of **Hippocalcin** protein in different *E.coli* strains. Page A (lines 1-13),: expression in *Plys.* Page A (line 15), B (lines 1-14) and C(lines 1-5): expression in *Codon plus*. In page C (lines 7-15) and D (lines 1-8, 12-14): is reported the expression in *GOLD*. The samples loaded in each line are classified with a letter, corresponding to the initial of the host strain (C+= Codon plus, G= Gold, p= pLys), and a number, given to each colony, as showed in table E. The *Low Range* standard in all gels, contains six proteins with known molecular weight (97.4KDa, 66.2KDa, 45KDa, 31KDa, 21.5KDa, 14 KDa).

From the observation of the gels in figure 4.13, it is possible to notice that many samples show an over-expressed band with a specific molecular weight. Comparing the MW of this band, with the hypothetical MW of the recombinant protein **Hippocalcin** calculated as

described in table 4.5, it is possible to deduce which plasmid has been acquired from each colony. The results of this screening can be summarized as follow:

- the pDEST17 plasmid has been acquired from the colonies BL21 Codon plus 4, 8, 9, 10, 12,13, 14, 19, from the colonies BL21 GOLD 1, 3, 8, 9, 12, 13, 14, 19 and from the colonies BL21 pLys 8, 4, 5 and10, showing an over-expressed band of around 27 kDa;
- the pETG20A plasmid has been acquired from the colonies BL21 Codon plus 3, 7, 1 and from the colony BL21 GOLD 20, showing an over-expressed band of around 40 kDa;
- the pETG30A plasmid has been acquired from the colonies BL21 GOLD 2, 4, 11, showing an over-expressed band of around 51 kDa;
- the pETG-60A plasmid, has been acquired from the colony Codon plus 2, showing an over-expressed band of around 80 kDa;

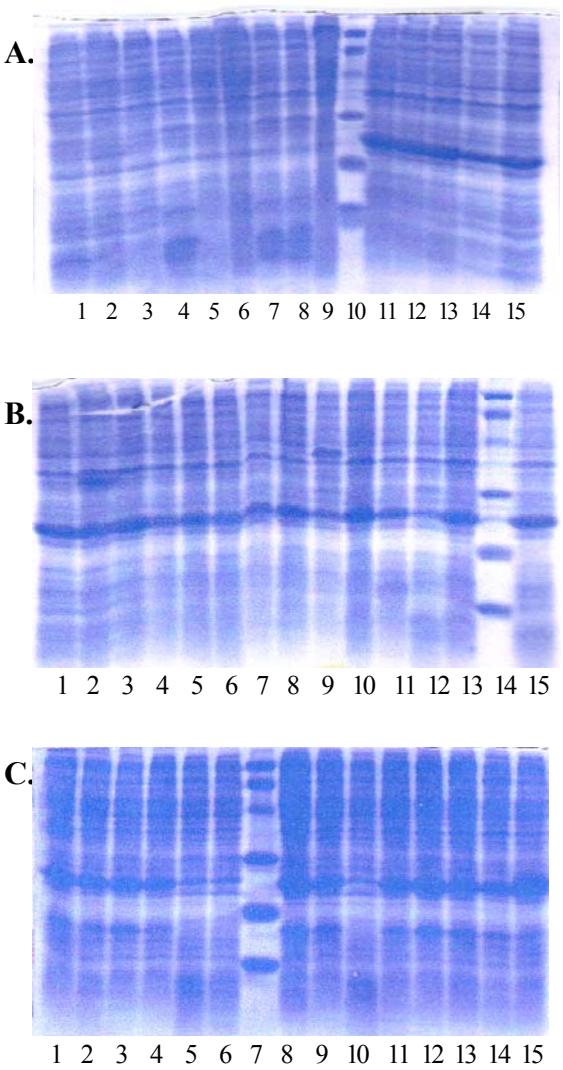
The results of the expression screening of Hippocalcin protein are summarized in table 4.15.

	pDEST17	**pETG20A**	**pETG30A**	**pETG60A**
Codon plus	4,8,9,10,12,13,14, 19	3,7,1	/	2
GOLD	1,3,8,9,12,13,14,19	20	2,4,11	/
pLys	8,4,5,10	/	/	/

Tab. 4.15 Results of the Hippocalcin expression screening. The number in red refer to the colonies that were selected to perform the solubility screening (paragraph 4.5).

4.4.11 Expression screening of a fragment of the BIR3-NAIP protein, aa 278-344

The expression of the **Bir 3** protein in three *E.coli* BL21 host strains (Codon plus, Plys and Gold) was verified by SDS-PAGE as described in paragraph 3.7. In the figures 4.14 the gels concerning the expression are reported.



E.

	A	B	C	D
1	G12	C+6	p1	P15
2	G13	C+7	p2	P16
3	G14	C+8	p3	P17
4	G15	C+9	p4	P18
5	G16	C+10	p5	P19
6	G17	C+11	p6	P20
7	G18	C+12	St	St
8	G19	C+13	p7	C+20
9	G20	C+14	p8	
10	St	C+15	p9	
11	C+1	C+16	p10	
12	C+2	C+17	p11	
13	C+3	C+18	p12	
14	C+4	St	p13	
15	C+5	C+19	p14	

Fig. 4.14 SDS-PAGE of the expression of BIR3-NAIP fragment protein in different *E.coli* strains. Page A (lines 1-9): expression in *gold.* Page A (line 11-15), B (lines 1-13, 15): expression in *Codon plus*. In page C (lines 1-6, 8-15) and D (lines 1-6, 8): is reported the expression in *plys*. The samples loaded in each line are classified with a letter, corresponding to the initial of the host strain (C+= Codon plus, G= Gold, p= pLys), and a number, given to each colony, as showed in table E. The *Low Range* standard in all gels, contains six proteins with known molecular weight (97.4KDa, 66.2KDa, 45KDa, 31KDa, 21.5KDa, 14 KDa).

From the observation of the gels in figure 4.14, it is possible to notice that many samples show an over-expressed band with a specific molecular weight. Comparing the MW of this band, with the hypothetical MW of the recombinant protein **Bir3** calculated as described in table 4.5, it is possible to deduce which plasmid has been acquired from each colony. The results of this screening can be summarized as follow:

The results of the expression screening of BIR3-NAIP fragment protein are summarized in table 4.16.

	pDEST17	pETG20A	pETG30A	pETG60A
Codon plus	20,17	7	14	/
GOLD	9,10,12,15,18,19	13	8	/
pLys	15,20,5,6,9	/	/	/

Tab. 4.16 Results of the BIR3-NAIP fragment expression screening. The number in red refer to the colonies that were selected to perform the solubility screening (paragraph 4.5).

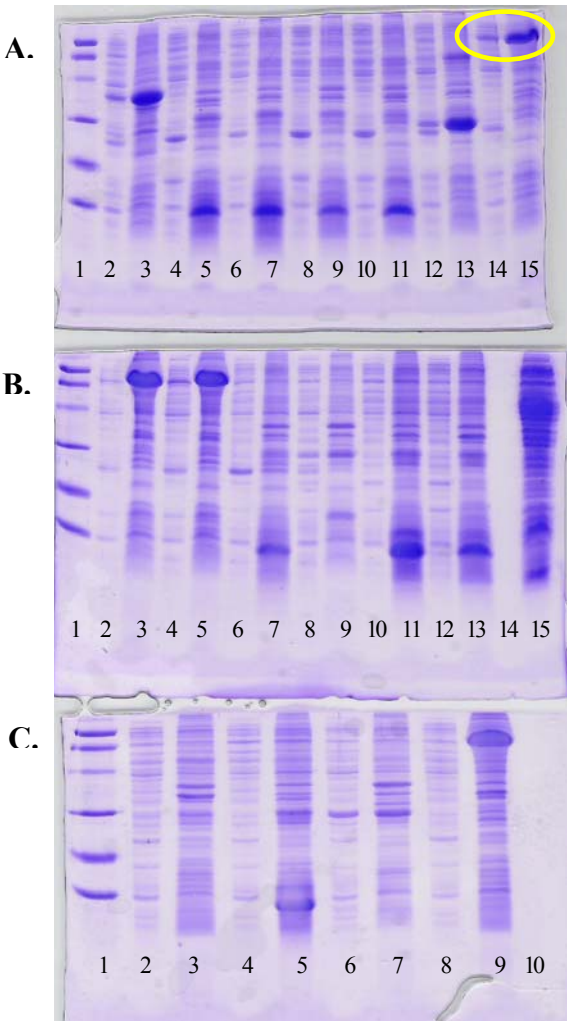
4.5 PROTEIN SOLUBILITY SCREENING

After the analysis of the protein expression on SDS-PAGEs, we have made a solubility test for each colony to verify if the proteins were expressed in inclusion bodies or in soluble form (paragraph 3.8). For each recombinant protein, at least a colony for type of host strain and for type of destination vector were selected. Such colonies are highlighted in red in the table 4.6,7,8,9,10,11,12,13,14,15,16.

The proteins solubility was checked by SDS-page verifying the presence of the protein in the soluble fraction (S.F.) or in the inclusion bodies (I.B.), loaded in adjacent lines.

4.5.1 Solubility screening of the protein S100A2

The figure 4.15 shows the SDS-page concerning the solubility screening of the protein S100A2.



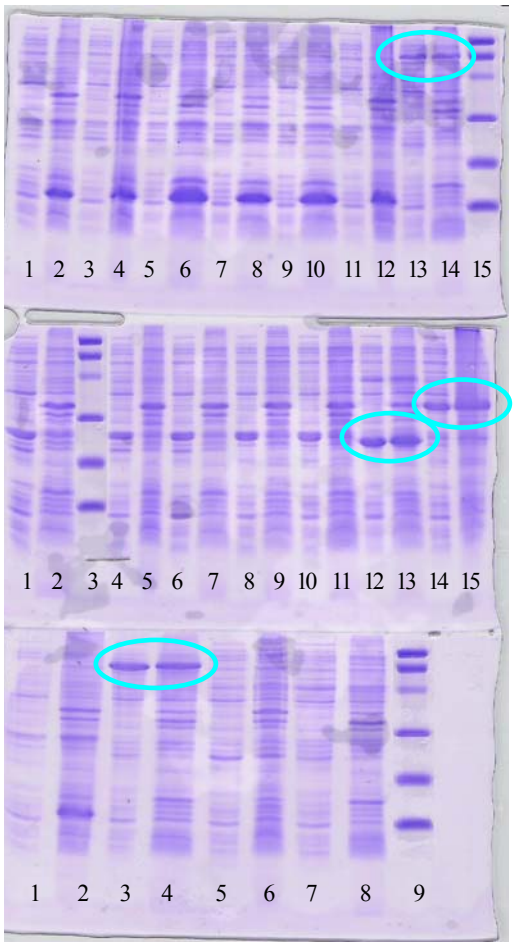
	A	B	C
1	st	st	st
2	P19 (S.F .)	P14 (S.F .)	G2 (S.F.)
3	P19 (I.B.)	P14 (I.B.)	G2 (I.B.)
4	P7 (S.F.)	P22 (S.F.)	G12 (S.F.)
5	P7 (I.B.)	P22 (I.B.)	G12 (I.B.)
6	P16 (S.F.)	C+16 (S.F.)	G5 (S.F.)
7	P16 (I.B.)	C+16 (I.B.)	G5 (I.B.)
8	P21 (S.F.)	C+23 (S.F.)	G16 (S.F.)
9	P21 (I.B.)	C+23 (I.B.)	G16 (I.B.)
10	P23 (S.F.)	G8 (S.F.)	/
11	P23 (I.B.)	G8 (I.B.)	/
12	P9 (S.F.)	G1 (S.F.)	/
13	P9 (I.B)	G1 (I.B.)	/
14	P13 (S.F.)	/	/
15	P13 (I.B.)	/	/

Fig. 4.15 Solubility screening of the protein S100A2. For each sample, the soluble fraction (S.F.) and the inclusion bodies (I.B.) were loaded in adjacent lines. The *Low Range* standard in all gels, contains six proteins with known molecular weight (97.4KDa, 66.2KDa, 45KDa, 31KDa, 21.5KDa, 14 KDa).

The S100A2 protein was expressed in an insoluble form by all the colonies. Only the colony pLys 13 (figure 4.15 page A, line 14-15 selected in yellow) expressed the 30% of the protein in a soluble form, fused with NusA+$(His)_6$ tag.

4.5.2 Solubility screening of the protein S100A16

The figure 4.16 hows the SDS-page concerning the solubility screening of the protein S100A16.



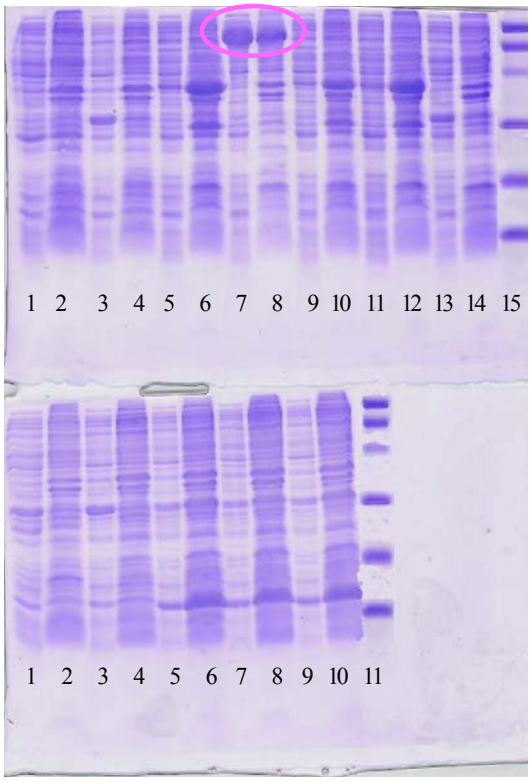
	A	B	C
1	C+19 (S.F.)	C+7 (S.F.)	G5 (S.F.)
2	C+19 (I.B.)	C+7 (I.B.)	G5 (I.B.)
3	C+25 (S.F.)	st	p12 (S.F.)
4	C+25 (I.B.)	C+17 (S.F.)	p12 (I.B.)
5	P6 (S.F.)	C+17 (I.B.)	C+3 (S.F.)
6	P6 (I.B.)	C+22 (S.F.)	C+3 (I.B.)
7	P14 (S.F.)	C+22 (I.B.)	p22 (S.F.)
8	P14 (I.B.)	G6 (S.F.)	p22 (I.B.)
9	P15 (S.F.)	G6 (I.B)	st
10	P15 (I.B.)	G11 (S.F.)	/
11	G4 (S.F.)	G11 (I.B.)	/
12	G4 (I.B.)	G16 (S.F.)	/
13	P5 (S.F.)	G16 (I.B.)	/
14	P5 (I.B..)	G10 (S.F.)	/
15	st	G10 (I.B.)	/

Fig. 4.16 Solubility screening of the protein S100A16. For each sample, the soluble fraction (S.F.) and the inclusion bodies (I.B.) were loaded in adjacent lines. The *Low Range* standard in all gels, contains six proteins with known molecular weight (97.4KDa, 66.2KDa, 45KDa, 31KDa, 21.5KDa, 14 KDa).

Observing the gels in figure 4.16, it's possible to note that the colonies pLys 12, GOLD 16, GOLD 10 and *pLys* 5 express the protein partially in a soluble form (lines selected in blue). In fact, the protein's band is visible in both the fractions. In particular, the colonies pLys 12 and pLys 5 express the protein fused with the NusA+$(His)_6$ fusion tag (line 13-14 page A, line 3-4 page C), the colony GOLD 16 expresses the protein fused with the TrxA+$(His)_6$tag (line 12-13 page B), while the colony GOLD 10 expresses the protein fused with the GST+ $(His)_6$tag (line 14-15 page B).

4.5.3 Solubility screening of the protein S100A10

The figure 4.17 shows the SDS-page concerning the solubility screening of the protein S100A10.



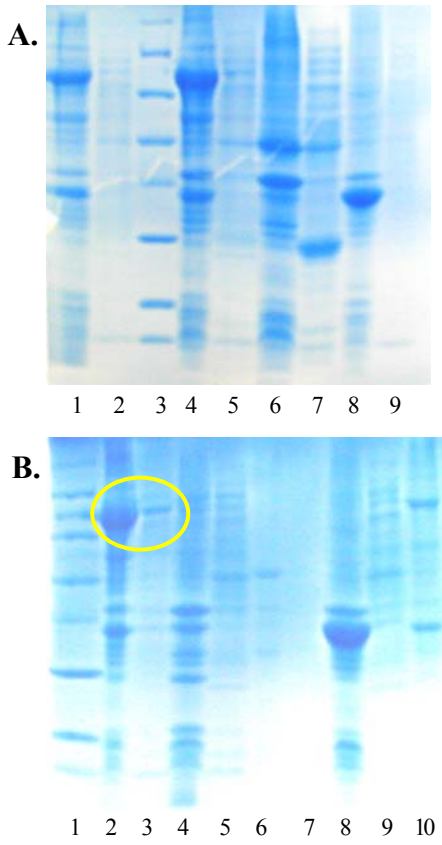
	A	B
1	C+8(S.F.)	P22 (S.F.)
2	C+8(I.B.)	P22 (I.B.)
3	C+22 (S.F.)	G1 (S.F.)
4	C+22 (I.B.)	G1 (I.B.)
5	P1 (S.F.)	G4 (S.F.)
6	P1 (I.B.)	G4 (I.B.)
7	P2(S.F.)	G10 (S.F.)
8	P2 (I.B.)	G10 (I.B.)
9	P11 (S.F.)	G22 (S.F.)
10	P11 (I.B.)	G22 (I.B.)
11	P21 (S.F.)	**st**
12	P21 (I.B.)	/
13	P7 (S.F.)	/
14	P7 (I.B.)	/
15	**st**	/

Fig. 4.17 Solubility screening of the protein S100A10. For each sample, the soluble fraction (S.F.) and the inclusion bodies (I.B.) were loaded in adjacent lines. The *Low Range* standard in all gels, contains six proteins with known molecular weight (97.4KDa, 66.2KDa, 45KDa, 31KDa, 21.5KDa, 14 KDa).

Only the colony pLys 2 (figure 4.17 page A, line 7-8, selected in pink) express the protein, partially in a soluble form, fused with NusA+$(His)_6$ tag.

4.5.4 Solubility screening of the isoform α of the Endothelial differentiation-related factor 1 (EDF-1 iso α)

The figure 4.18 shows the SDS-page concerning the solubility screening of the protein EDF-1 iso α.



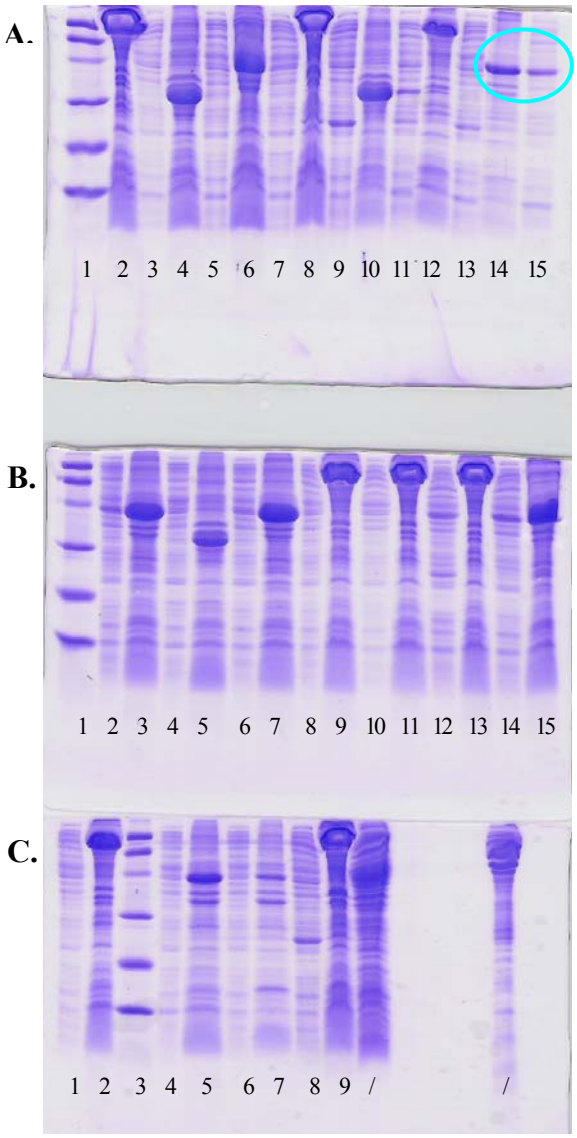
	A	B
1	C+3 (I.B.)	st
2	C+3 (S.F.)	G5 (I.B.)
3	st	G5 (S.F.)
4	C+4 (I.B.)	G10 (I.B.)
5	C+4 (S.F.)	G10 (S.F.)
6	C+5 (I.B.)	G12 (I.B.)
7	C+5 (S.F.)	G12 (S.F.)
8	C+16 (I.B.)	G19 (I.B.)
9	C+16 (S.F.)	G19 (S.F.)
10		P5 (I.B.)
11		P5 (S.F.)
12		P6 (I.B.)
13		P6 (S.F.)
14		P9 (I.B.)
15		P9 (S.F.)

Fig. 4.18 Solubility screening of the protein EDF-1 iso α. For each sample, the soluble fraction (S.F.) and the inclusion bodies (I.B.) were loaded in adjacent lines. The Protein Molecular Weight Marker Standard in all gels, contains seven proteins with known molecular weight (11.6KDa, 66.2KDa, 45KDa, 35KDa, 25KDa, 18.4 KDa, 14.4KDa).

Observing the gels in figure 4.18 , it's possible to note that all the colonies express the protein in inclusion bodies unless the colony GOLD 5, selected in yellow, expressing the protein partially in a soluble form fused with NusA+$(His)_6$ tag.

4.5.5 Solubility screening of the isoform β of theEndothelial differentiation-related factor 1 (EDF-1 isoβ)

The figure 4.19 shows the SDS-page concerning the solubility screening of the protein EDF-1 isoβ.



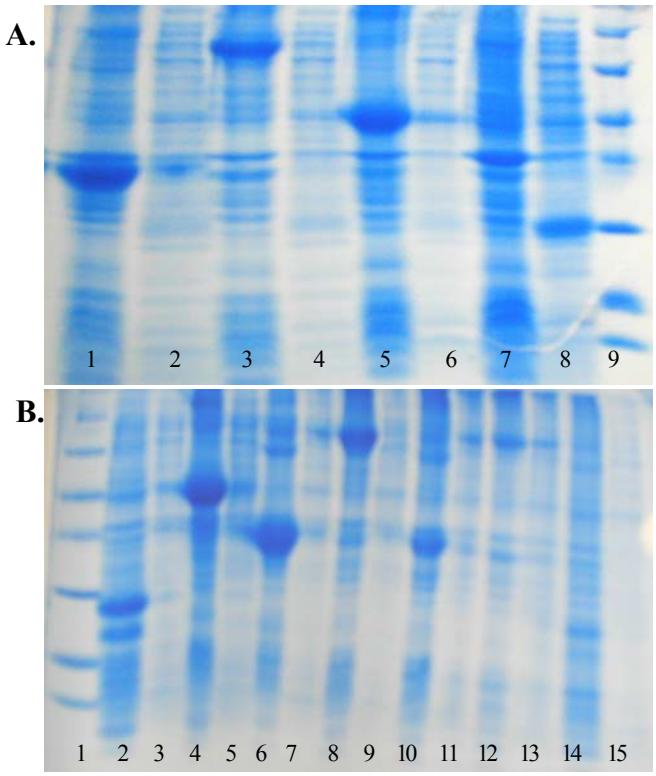
	A	B	C
1	st	st	G7 (S.F.)
2	G1 (I.B.)	G5 (S.F.)	G7 (I.B.)
3	G1 (S.F.)	G5 (I.B)	st
4	G2 (I.B.)	G6 (S.F.)	G15 (S.F.)
5	G2 (S.F.)	G6 (I.B)	G15 (I.B.)
6	G4 (I.B.)	G8 (S.F.)	P18 (S.F.)
7	G4 (S.F.)	G8 (I.B)	P18 (I.B.)
8	C+2 (I.B.)	G9 (S.F.)	C+11 (S.F.)
9	C+2 (S.F.)	G9 (I.B)	C+11 (I.B.)
10	C+10 (I.B.)	G3 (S.F.)	/
11	C+10 (S.F.)	G3 (I.B)	/
12	P12 (I.B.)	G18 (S.F.)	/
13	P12 (S.F.)	G18 (I.B.)	/
14	P13 (I.B.)	G20 (S.F.)	/
15	P13 (S.F.)	G20 (I.B.)	/

Fig. 4.19 Solubility screening of the EDF-1 isoβ protein. For each sample, the soluble fraction (S.F.) and the inclusion bodies (I.B.) were loaded in adjacent lines. The *Low Range* standard in all gels, contains six proteins with known molecular weight (97.4KDa, 66.2KDa, 45KDa, 31KDa, 21.5KDa, 14 KDa).

The figure 4 shows that all the colonies express the protein in inclusion bodies with the exception of the colony pLys 13 expressing the protein partially in a soluble, form fused with GST+$(His)_6$ tag.

4.5.6 Solubility screening of a fragment of the Cyclin-dependent kinase inhibitor, aa 1-160, (P21-160)

The figure 4.20 shows the SDS-page concerning the solubility screening of the P21-160 protein.



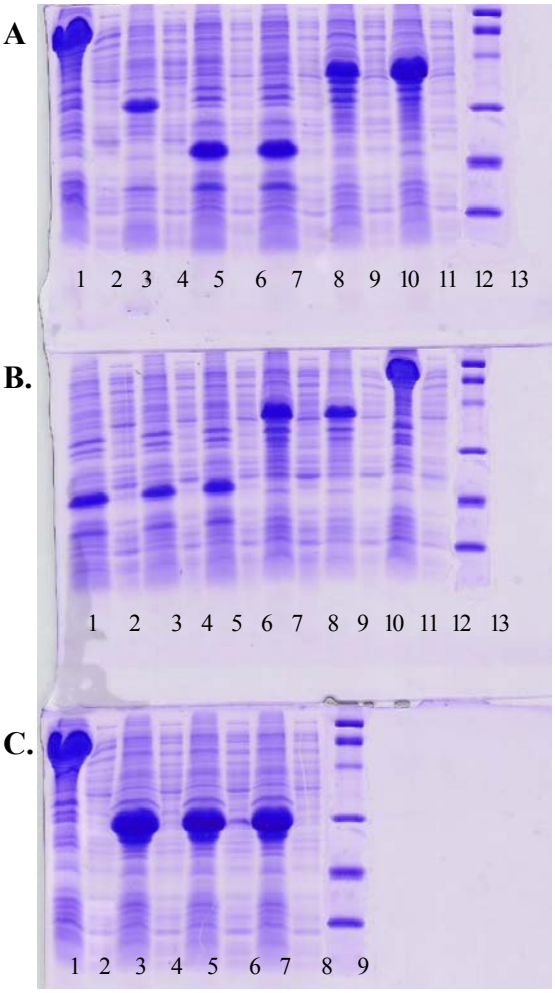
	A	B
1	C+3 (I.B.)	st
2	C+3 (S.F.)	G4 (I.B.)
3	C+4 (I.B.)	G4 (S.F.)
4	C+4 (S.F.)	G8 (I.B.)
5	C+6 (I.B.)	G8 (S.F.)
6	C+6 (S.F.)	G11(I.B.)
7	C+11 (I.B.)	G11 (S.F.)
8	C+11 (S.F.)	G13 (I.B)
9	st	G13 (S.F.)
10		P9 (I.B.)
11		P9 (S.F.)
12		P13(I.B.)
13		P13 (S.F.)
14		P14 (I.B.)
15		P14 (S.F.)

Fig. 4.20 Solubility screening of the protein P21-160. For each sample, the soluble fraction (S.F.) and the inclusion bodies (I.B.) were loaded in adjacent lines. The Protein *Molecular Weight Marker Standard* used is composed of even proteins with known molecular weight (11.6KDa, 66.2KDa, 45KDa, 35KDa, 25KDa, 18.4 KDa, 14.4KDa).

All the colonies express the protein in inclusion bodies, as showed in figure 4.20. The protein bands are coherent with the expression screening.

4.5.7 Solubility screening of a fragment of the Cyclin-dependent kinase inhibitor 1 (P21), aa 25-164, (P21-140)

The figure 4.21 shows the SDS-page concerning the solubility screening of the protein P21-140.



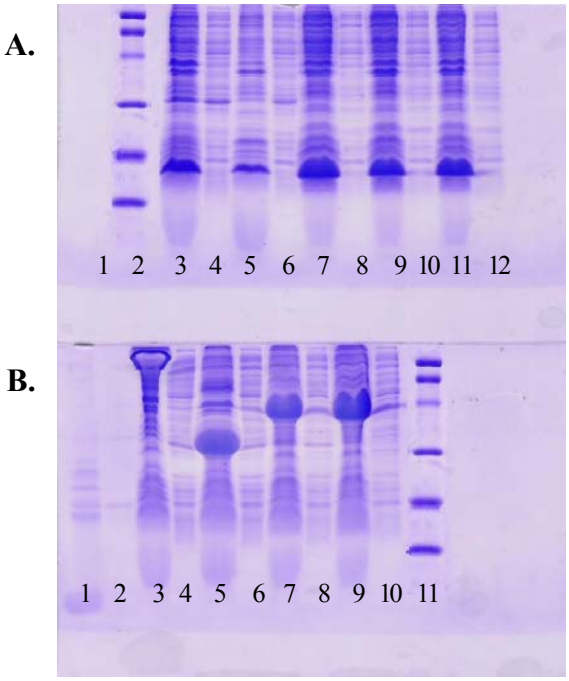
	A	B	C
1	P1 (I.B.)	C+18(I.B.)	G8 (I.B)
2	P1 (S.F.)	C+18(S.F.)	G8 (S.F.)
3	P3 (I.B.)	C+9 (I.B.)	G4 (I.B)
4	P3 (S.F.)	C+9 (S.F.)	G4 (S.F.)
5	G17 (I.B.)	C+8 (I.B.)	G11 (I.B)
6	G17 (S.F.)	C+8 (S.F.)	G11 (S.F.)
7	G9 (I.B.)	C+10 (I.B.)	G12 (I.B)
8	G9 (S.F.)	C+10 (S.F.)	G12 (S.F.)
9	G3 (I.B.)	C+19 (I.B.)	**st**
10	G3 (S.F.)	C+19 (S.F.)	
11	G7 (I.B.)	C+5 (I.B.)	
12	G7 (S.F.)	C+5 (S.F.)	
13	**st**	**st**	

Fig. 4.21 Solubility screening of the protein P21-140. For each sample, the soluble fraction (S.F.) and the inclusion bodies (I.B.) were loaded in adjacent lines. The *Low Range* standard in all gels, contains six proteins with known molecular weight (97.4KDa, 66.2KDa, 45KDa, 31KDa, 21.5KDa, 14 KDa).

All the colonies express the protein in inclusion bodies, as showed in figure 4.21. The protein bands are coherent with the expression screening.

4.5.8 Solubility screening of a fragment of the SH3 adapter protein SPIN90, aa 1-81, (SH3-81)

The figure 4.22 shows the SDS-page concerning the solubility screening of the SH3-81. All the colonies express the protein in inclusion bodies. The protein bands are coherent with the expression screening.



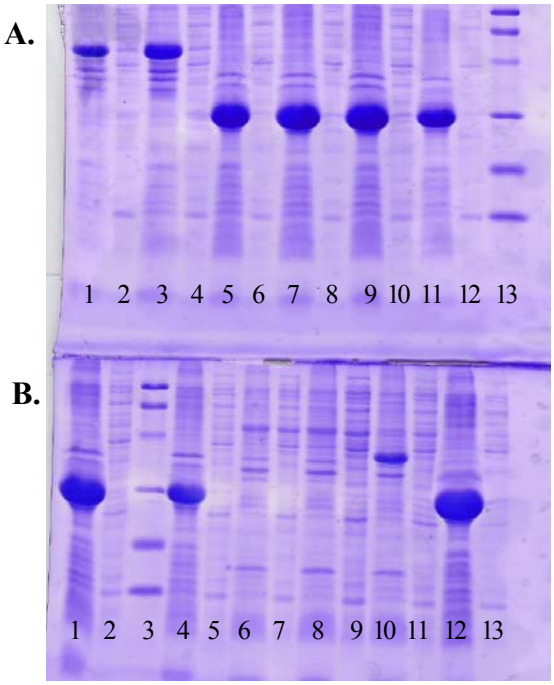
	A	B
1	/	P2 (I.B.)
2	**st**	P2 (S.F.)
3	C+10 (I.B.)	C+3 (I.B.)
4	C+10 (S.F.)	C+3 (S.F.)
5	P7 (I.B.)	C+7 (I.B.)
6	P7 (S.F.)	C+7 (S.F.)
7	G7 (I.B.)	C+12 (I.B.)
8	G7 (S.F.)	C+12 (S.F.)
9	G14 (I.B.)	G13 (I.B.)
10	G14 (S.F.)	G13 (S.F.)
11	G23 (I.B.)	**st**
12	G23 (S.F.)	

Fig. 4.22 Solubility screening of the protein SH3-81. For each sample, the soluble fraction (S.F.) and the inclusion bodies (I.B.) were loaded in adjacent lines. The *Low Range* standard in all gels, contains six proteins with known molecular weight (97.4KDa, 66.2KDa, 45KDa, 31KDa, 21.5KDa, 14 KDa).

All the colonies express the protein in inclusion bodies, as showed in figure 4.22. The protein bands are coherent with the expression screening.

4.5.9 Solubility screening of the C-terminal domain of the Tuberous sclerosis 2 protein TSC2_HUMAN, aa 1531 – 1758, (TSC2-228)

The figure 4.23 shows the SDS-page concerning the solubility screening of the protein TSC2.



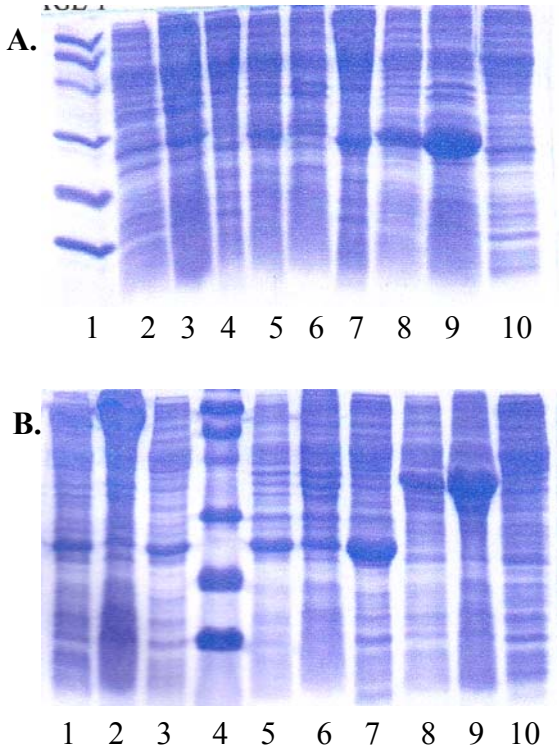
	A	B
1	G6 (I.B.)	C+4 (I.B.)
2	G6 (S.F.)	C+4 (S.F.)
3	G19 (I.B.)	st
4	G19 (S.F.)	C+18 (I.B.)
5	G15 (I.B.)	C+18 (S.F.)
6	G15 (S.F.)	P3 (I.B.)
7	G13 (I.B.)	P3 (S.F.)
8	G13 (S.F.)	P6 (I.B.)
9	G8 (I.B)	P6 (S.F.)
10	G8 (S.F.)	P1 (I.B.)
11	C+1 (I.B.)	P1 (S.F.)
12	C+1 (S.F.)	P4 (I.B.)
13	st	P4 (S.F.)

Fig. 4.23 Solubility screening of the protein TSC2. For each sample, the soluble fraction (S.F.) and the inclusion bodies (I.B.) were loaded in adjacent lines. *Low Range* standard in all gels, contains six proteins with known molecular weight (97.4KDa, 66.2KDa, 45KDa, 31KDa, 21.5KDa, 14 KDa).

All the colonies express the protein in inclusion bodies, as showed in figure 4.23. The protein bands are coherent with the expression screening.

4.5.10 Solubility screening of protein Hippocalcin

The figure 4.24 shows the SDS-page concerning the solubility screening of the protein Hippocalcin. Besides to the soluble fraction and the inclusion bodies, in figure 4.24 the unbroken cells were loaded, too.



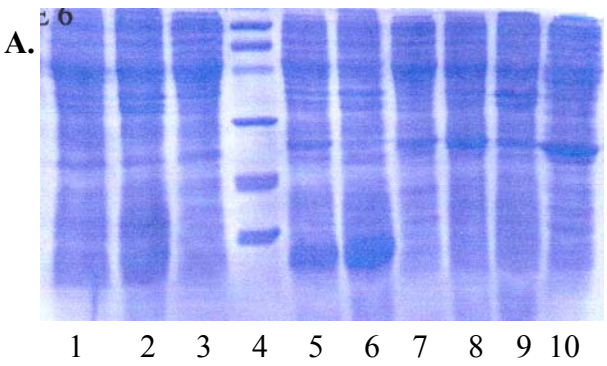
	A	B
1	st	C+ 2
2	G8	C+2(I.B.)
3	G8 (I.B.)	C+2 (S.F.)
4	G8 (S.F.)	st
5	C+4	C+3
6	C+4 (I.B.)	C+3(I.B.)
7	C+4 (S.F.)	C+3(S.F.)
8	P3	G8
9	P3 (I.B.)	G8 (I.B.)
10	P3 (S.F.)	G8 (S.F.)

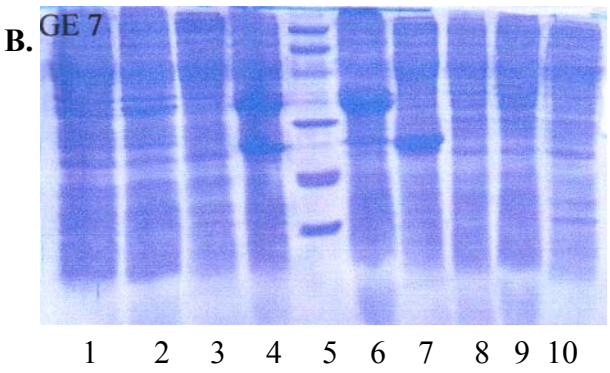
Fig. 4.24 Solubility screening of the protein Hippocalcin. For each sample, the soluble fraction (S.F.) and the inclusion bodies (I.B.) were loaded in adjacent lines. *Low Range* standard in all gels, contains six proteins with known molecular weight (97.4KDa, 66.2KDa, 45KDa, 31KDa, 21.5KDa, 14 KDa).

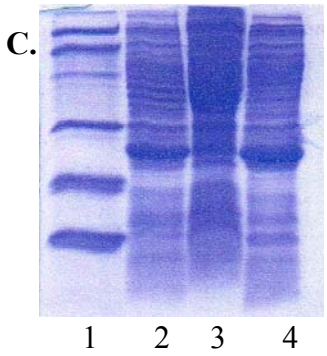
All the colonies express the protein in inclusion bodies, as showed in figure 4.24. The protein bands are coherent with the expression screening.

4.5.11 Solubility screening of a fragment of the BIR3-NAIP protein, aa 278-344

The figure 4.25 shows the SDS-page concerning the solubility screening of the BIR3 protein. Besides to the soluble fraction and the inclusion bodies, in figure 4.25 the unbroken cells were loaded too.







	A	B	C
1	G18	G13	C+14
2	G18 (I.B.)	G13 (I.B.)	C+14 (I.B.)
3	G18 (S.F.)	G13 (S.F.)	C+14 (S.F.)
4	**st**	C+ 7	
5	P15	**st**	
6	P15 (I.B.)	C+7 (I.B.)	
7	P15 (S.F.)	C+7(S.F.)	
8	C+ 20	G8	
9	C+20 (I.B.)	G8 (I.B.)	
10	C+20 (S.F.)	G8 (S.F.)	

Fig. 4.25 Solubility screening of the Bir3 protein. For each sample, the soluble fraction (S.F.) and the inclusion bodies (I.B.) were loaded in adjacent lines. *Low Range* standard in all gels, contains six proteins with known molecular weight (97.4KDa, 66.2KDa, 45KDa, 31KDa, 21.5KDa, 14 KDa).

All the colonies express the protein in inclusion bodies, as showed in figure 4.25. The protein bands are coherent with the expression screening.

4.5.12 Expression and solubility screening results

In the table 4.17 the expression and solubility screening results are summarized.

Protein category	Protein Name	Expression screening					Solubility screening	
		pDEST -17	pDEST -15	pETG-20A	pETG-30A	pETG-60A	S.F	I.B.
calmodulin targets	Endothelial differentiation-related factor 1 ISO-α			*Plys Codon Gold*	*Plys Codon Gold*	*Plys Codon Gold*	x	x
	Endothelial differentiation-related factor 1 ISO-β			*Codon Gold*	*Plys Gold*	*Plys Codon Gold*	x	x
	Cyclin-dependent Kinase inhibitor 1(P21) fragment 1-160 aa	*Plys Codon Gold*	/	*Plys Gold*	*Codon Gold*	*Plys Codon Gold*		x
	Cyclin-dependent Kinase inhibitor1 (P21) fragment 25-164 aa		*Codon Gold*	*Codon Gold*	*Codon Gold*	*Plys Codon Gold*		x
	SH3 adapter protein SPIN 90 722 aa, fragment 1-81 aa	*Plys Codon Gold*	*Codon Gold*	*Codon*	*Plys*	*Codon*		x
	Tuberous sclerosis TSC2 C-terminale domain, 1531-1758aa	*Plys Codon Gold*		*Plys Gold*	*Plys Gold*			x
EF-Hands	S100A2	*Plys Codon Gold*		*Plys Codon Gold*	*Plys*	*Plys Gold*	x	x
	S100A16	*Plys Codon Gold*		*Codon Gold*	*Gold*	*Plys*	x	x
	S100A10	*Gold*		*Plys Codon Gold*	*Codon Plys*	*Plys*	x	x
	Hippocalcin	*Plys Codon Gold*	/	*Codon Gold*	*Gold*	*Codon*		x
Hippocalcin target	BIR3-NAIP fragment 278-344 aa	*Plys Codon Gold*	/	*Codon Gold*	*Codon Gold*			x

Tab. 4.17 Expression and solubility screening results. For each sample, the *E.coli* host strains (*Plys, Codon, Gold)* which express the proteins with the different destination vectors are indicated. S.F.: soluble fraction; I.B.:inclusion bodies.

As showed in table 4.17, all the selected proteins were well expressed with almost three of the five fusion tags in all *E. coli* host strains. Moreover, from the solubility screening, it results that the calmodulin targets proteins were expressed in inclusion bodies by all the host strains with an only one exception, the protein *endothelial differentiation-related factor1* (EDF-1). On the contrary, there is a large variability in the solubility of the EF-hand proteins; in fact, in most cases, the proteins were produced in inclusion bodies and only with few tags, different

for each protein, were expressed in soluble form (S100A2 fused with NusA+$(His)_6$ tag; S100A16 fused with the NusA+$(His)_6$ fusion tag, TrxA+$(His)_6$ tag and GST+$(His)_6$ tag; S100A10 fused with NusA+$(His)_6$ tag). This result is very important for the successive protein purification steps on a larger scale. In fact, a soluble protein, fused with a specific affinity tag could be easily purified using a few chromatographic passages.

4.6 SEQUENCING

Before the large scale protein production, the selected plasmids were sequenced to confirm the correct insertion of the genes. The sequencing was performed by PRIMM laboratory in Milan. The sequencing results confirmed that the sequence of all samples was correct.

4.7 CLONING, PRODUCTION AND PURIFICATION OF *Calmodulin like to skin protein (CLSP)*

CLSP is a 146 residue (15.92 kDa) protein that shares 49% of identity with CaM. At variance with the ubiquitous CaM, CLSP is reported to be expressed only in epithelial cells, where it may be involved in the Ca^{2+}-dependent differentiation process of keratinocytes to corneocytes. Particularly, it is expressed at very late stages of differentiation just underneath the stratum corneum and is considered to play an important role in this process (Babini et al., 2005). The characteristic of the protein and his domains are reported in the table 4.18.

>Q9NZT1|CALL5_HUMAN Calmodulin-like protein 5 - Homo sapiens (Human)
MAGELTPEEEAQYKKAFSAVDTDGNGTINAQELGAALKATGKNLSEAQLRKLISEVDSDG
DGEISFQEFLTAARKARAGLEDLQVAFRAFDQDGDGHITVDELRRAMAGLGQPLPQEELD
AMIREADVDQDGRVNYEEFARMLAQE

Number of amino acids: 146
Molecular weight: 15920.5
Theoretical pI: 4.34
Amino acid composition:

Ala (A)	21	14.4%
Arg (R)	9	6.2%
Asn (N)	4	2.7%
Asp (D)	14	9.6%
Cys (C)	0	0.0%
Gln (Q)	10	6.8%
Glu (E)	17	11.6%
Gly (G)	13	8.9%
His (H)	1	0.7%
Ile (I)	5	3.4%
Leu (L)	14	9.6%
Lys (K)	6	4.1%
Met (M)	4	2.7%
Phe (F)	6	4.1%
Pro (P)	3	2.1%
Ser (S)	5	3.4%
Thr (T)	6	4.1%
Trp (W)	0	0.0%
Tyr (Y)	2	1.4%
Val (V)	6	4.1%
Asx (B)	0	0.0%
Glx (Z)	0	0.0%
Xaa (X)	0	0.0%

Total number of negatively charged residues (Asp + Glu): 31
Total number of positively charged residues (Arg + Lys): 15
Formula: $C_{682}H_{1085}N_{195}O_{236}S_4$
Total number of atoms: 2202

FT	DOMAIN	8-43	EF-hand 1.
FT	DOMAIN	44-77	EF-hand 2.
FT	DOMAIN	78-113	EF-hand 3.
FT	DOMAIN	114-146	EF-hand 4.

Tab. 4.18 Principal chemical and physical characteristics of CLSP.

The protein is highly homologus to calmodulin but the differences in the sequence are sufficient to impose a very different structure. The structure of the protein expressed in *E.coli* wasn't resolved by NMR, since the N-terminal domain is highly flexible. Consequently we have decided to produce the two domains separately. The cloning of CLSP and his N-terminal and C-terminal domains was conduced using the standard protocol as described in the following chapter.

4.7.1 Amplification and Cloning of the gene and DNA fragments coding for CLSP and for its N-terminal (N-CLSP) and C-terminal (C-CLSP) domains

The nucleotide sequence coding for human Calmodulin like to skin protein (CLSP) was amplified from cDNA by PCR using the primers oCLSPFor and oCLSPRev, whose sequence is reported in table 4. The amplified fragment of 441 bp was purified from agarose gel and blunt-end ligated into the expression vector pQE30Xa (QIAGEN) at the *StuI* site. The sequence of the gene in the plasmid pQE-CLSP so far obtained, was confirmed by single run sequencing using standard forward and reverse primers for pQE30Xa, which was performed at Primm Biotech.

Primer for amplification of the gene coding for the full length CLSP protein	
oCLSPFor [forward]	5′ATGGCCGGTGAGCTGACTCCT 3′
oCLSPRev [reverse]	5′TCACTCCTGGGCGAGCATCCT 3′
Primer for amplification of the DNA fragment coding for the N-terminal CLSP domain (M1 – A76)	
oNCLSPFor [forward]	5′ ATGGCCGGTGAGCTGACTCCT 3′
oNCLSPRev [reverse]	5′ TCAGGCCTTCCTTGCCGCCGT 3′
Primer for amplification of the of the DNA fragment coding for the C-terminal CLSP domain (A76 - E146)	
oCCLSPFor [forward]	5′ GCCAGGGCCGGCCTGGAGGAC 3′
oCCLSPRev [reverse]	5′ TCACTCCTGGGCGAGCATCCT 3′

Tab. 4.19 Sequences of the primers used to amplify the genes coding for CLSP and its N-terminal and C-terminal domains.

The DNA fragments coding for the N–terminal CLSP domain (N-CLSP, corresponding to the CLSP sequence from M1 to A76) and C–terminal CLSP domain (C-CLSP corresponding to the CLSP sequence from A76 to E146), were sub-cloned from plasmid pQE-CLSP using the primers described in table 1 for PCR amplification of pQE-CLSP. After extraction from agarose gel the fragments (of 231 bp and 216 bp respectively) were ligated to the pQE30Xa plasmid at the StuI site, and the constructs (pQE-NCLSP and pQE-CCLSP) were sequenced as described for the full length gene (Babini et al, 2006). The DNA fragments coding for CLSP and the N–terminal CLSP domain were amplified by PCR using TAQ Polymerase to verify the correct insertion of the genes in the vector. In the figure 4.26 the agarose gel of the amplified fragments is reported.

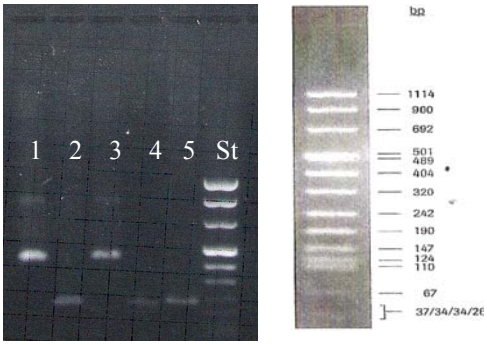


Fig. 4.26 Agarose gels concerning the amplification of the plasmidic fragments obtained using the primer reported in table 4.23. CLSP (line 1, 3). N–terminal CLSP domain (line 2, 4). C–terminal CLSP domain (line 5). Near the figure the used standard was reported.

4.7.2 Expression of recombinant CLSP (rCLSP) and of its C-terminal (rC-CLSP) and N-terminal (rN-CLSP) domains

The host strain used for expression of recombinant CLSP, C-CLSP and N-CLSP in the vector pQE30Xa was *E.coli* XL1Blue. Cells containing pQE-CLSP, pQE-CCLSP and pQE-NCLSP were grown in 3 liters of Luria Bertani medium (LB) at 37°C. Ampicillin (100 μg/ml) was added to the media as a selectable marker. After about 15 h from induction, cells were harvested by centrifugation. Protein expression was verified by SDS-PAGE as showed in figure 4.27.

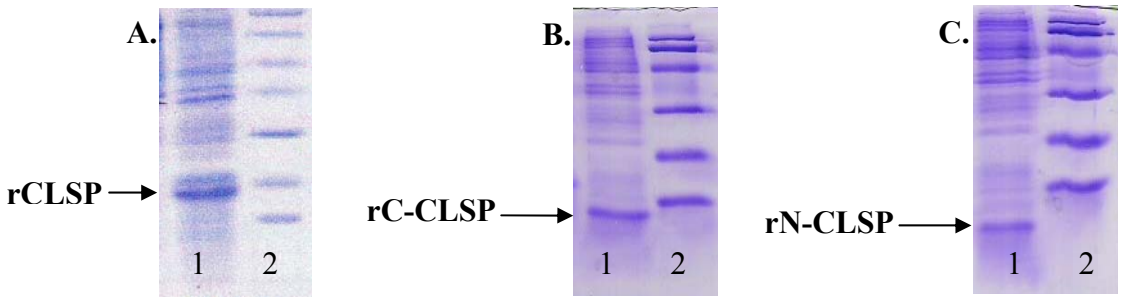


Fig. 4.27: Page A: SDS-page concerning the expression of rCLSP (Line 1). Standard Protein Molecular Weight Marker (line 2). Page B: SDS-page concerning the expression of rC-CLSP (Line)1. Standard Protein Molecular Weight Marker (line 2). Page C: SDS-page concerning the expression of rN-CLSP (Line 1). Standard Low range (line 2).

4.7.3 Purification of recombinant CLSP (rCLSP) and of its C-terminal (rC-CLSP) and N-terminal (rN-CLSP) domains

The proteins rCLSP, rC-CLSP, rN-CLSP were expressed in soluble form, so the supernatant obtained from the crude extract, and containing the soluble recombinant protein, was applied to a chromatography column.

The vector pQE30Xa allows the expression of the recombinant protein fused to an esa-hystidine chain which can be removed by the enzyme FactorXa (Amersham). So, purification of the soluble **rC-CLSP** was made using the following steps:

- ✓ metal chelate affinity chromatography to purify the protein before the cleavage. The presence of the protein in the Hitrap eluted was confirmed by SDS-PAGE, as showed in figure 4.28.

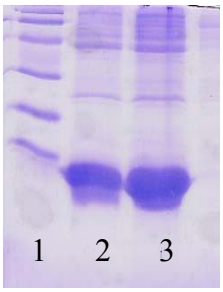


Fig. 4.28 SDS-page of rC-CLSP domain sample after the elution from Hitrap column (line 2, 3). Standard Low range (line 1).

- ✓ cleavage with Factor Xa. For removal of the histidine tag, the purified fusion protein was dialysed in FactorXa cleavage buffer [20mM TrisHCl, 1 mM $CaCl_2$, 100 mM

NaCl, pH 8.0], concentrated to 5 ml and then digested with FactorXa for 15 hours at 4°C.

- ✓ metal chelate affinity chromatography to isolate the protein from the tag detained on column and than eluited with 300mM Imidazol (line 4). In figure 4.29 the protein before the digestion (line 1), the cut (line 3) and the His_6-tag (line4) were showed.

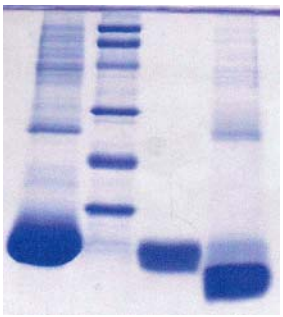


Fig. 4.29 : SDS-page of rC-CLSP domain before the digestion (line 1). In line 3 was loaded the sample after Hitrap chromatography (line 3) and the His_6-tag (line4).

- ✓ size exclusion chromatography
- ✓ anion exchange chromatography

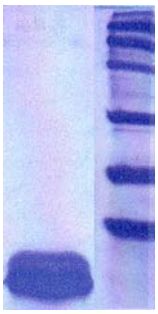


Fig. 4.30 : SDS-page of final purified rC-CLSP domain.

Purified c-CLSP concentration was determined by measuring absorbance at 276 nm, using the extinction coefficient ε_{276} = 1450 $M^{-1}cm^{-1}$, as calculated by ProtPar parameter tools (http://expasy.org/tools/protoaram.html).

Purification of **rCLSP and rN-CLSP**, both produced in a soluble form, required a denaturation refolding step on the nickel chelate column. Briefly, after loading the His_6-tagged protein sample on the Hitrap column, a denaturation solution [50mM Tris pH 8.0,

200mM NaCl, 7M Guanidinium chloride] was passed through the column, for four hours at flow 0.5 ml/min. Guanidinium was slowly removed from the column by a 16 hours continuous gradient against [50mM Tris pH 8.0, 200mM NaCl], to allow protein refolding. After elution of the protein from the Hitrap column the cleavage and the other purification steps were the same used for the C-terminal domain.

The concentrations of rCLSP and rN-CLSP were determined by measuring absorbance at 276nm, using the respective extinction coefficient $\varepsilon_{276} = 3400\ M^{-1}cm^{-1}$ for rCLSP $\varepsilon_{276} = 1450\ M^{-1}cm^{-1}$ for rN-CLSP (Babini et al., 2005).

4.7.4 Preliminary screening of folding of CLSP (rCLSP) and of its C-terminal (rC-CLSP) and N-terminal (rN-CLSP) domains

The 1H-NMR spectrum of the whole CLSP protein is reported in figure 4.31, being the protein constituted by both the N-terminal and C-terminal domains. The presence of very few signals more downfield than 9.0 ppm and more upfield than 0.5 ppm indicates that the protein was not firmly structured. In fact, the signals above 9.0 ppm concern the NH amide groups which are not exchangeable with the solvent, because are well buried within the macromolecular frame. Only a compact structure avoids the access of solvent to sites inside the globular domains. On the contrary, flexible and opened structures permit the chemical exchanges of protons between all amide groups and water, resulting in the disappearing of signals belonging to the NH groups in the protein backbone. For what it concerns the signals below 0.5 ppm, they belong to methyl groups of hydrophobic sidechains, which usually are well embedded within the protein core, thus resulting in a deshielding of the respective nuclei that resonate at very low fields. Only immobile sidechains have hydrogen nuclei falling in such a region, since methyl groups belonging to aminoacids located in flexible parts of the protein have chemical shifts at around 1 ppm, *e.g.* in the most crowded area of the NMR spectrum of a protein.

The protein unfolding implicates a major molecular flexibility that allows to the solvent to penetrate in proximity of all amidic groups and exchanges its own protons with them.

Additionally, the presence of a whole ensemble of signals, appearing as a very broad bump in the range between 6.5 and 9.0 ppm indicates a lack of univocity for the amidic signals. This phenomenon is attributable to the fact that the observed protons are part of a very mobile regions of the molecule, not stabilized by hydrogen bonds or Van der Waals interactions, as usually found in well folded structures. This fact implicates some dispersion in the position of

the atom of amidic hydrogen which causes a "broadening" of the corresponding NMR signals. In other words, the NH protons of all aminoacids should fall all at the same chemical shift, but in the protein they experience a different environment, since the macromolecule constitutes an anisotropic medium. The result is that each NH proton has its own resolved position in the NMR spectra. When the protein is very flexible, the aminoacids behave as being free and not constrained to a specific position. For this reason their chemical shifts become similar to each other and all of them overlap in a very restricted area. The same information is easily verified by inspecting the ^{1}H-^{15}N-HSQC, a 2D spectrum, showing the scalar correlations between amidic nitrogens and the protons directly bound to them. Comments on the interpretations of such a spectrum will be given below.

As mentioned before, the presence of methyl groups involved in the formation of hydrophobic patches produces 1H-NMR spectra characterized by the existence of signals below 0.5 ppm. In the spectrum reported in figure 4.47, only five signals could be recognized in such a region, having very low intensity, lower than that is expected from a so high concentration, as that for the protein sample used to record the present spectrum. The presence of these signals is compatible with a co-presence, in solution, of a small molar fraction of correctly folded protein together with a large fraction of unfolded protein.

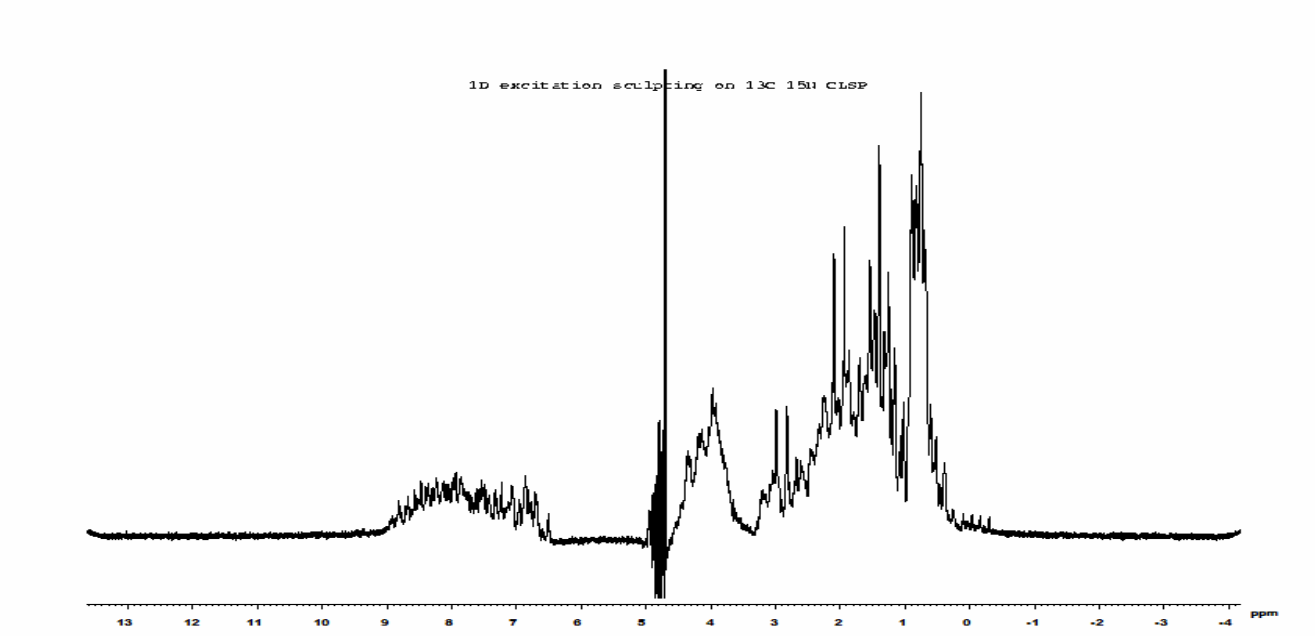


Fig. 4.31: 1D-NMR spectrum of protein CLSP

Another hypothesis is based on the fact that this molecule is constituted by a couple of domains tethered together by a linker. Thus, one domain may be characterized by a stable fold while the other one may have high flexibility. This hypothesis is based on the observation that

five methyl signals at about 0.5 ppm (each signal integrating three protons) have much smaller intensity than that of the signal falling at 6.6 ppm presumably due to a single proton.

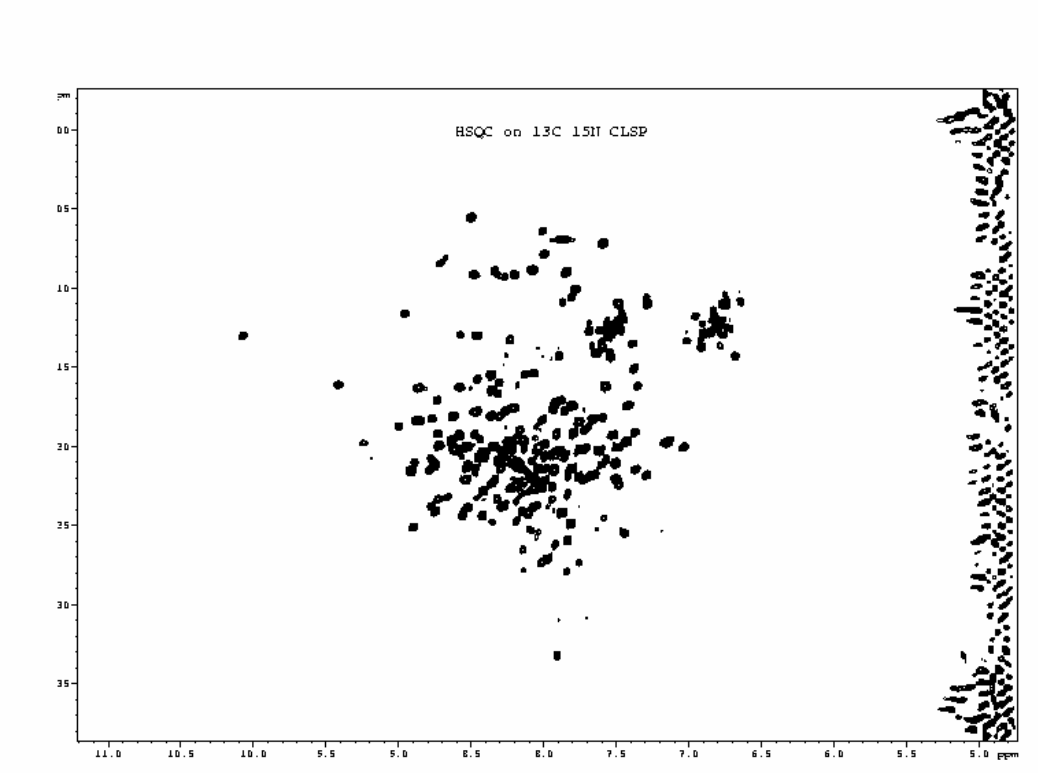


Fig. 4.32 : 2D-HSQC NMR spectrum of CLSP protein.

The spectrum shown in figure 4.32 is relative to a ^{1}H-^{15}N HSQC NMR experiment. This kind of spectrum has a higher resolving power (ability of discriminate between two signals) with respect to the 1D one, by increasing the number of dimensions on which each signals could be positioned. This 2D spectrum is characterized by a correlation peak between the frequency of the amidic proton with that one of the amidic nitrogen bonded to it. Each correlation signal corresponds to one aminoacid, so that by counting all spots in this spectrum it is possible to get the protein sequence length. The inspection of the spectrum in figure 4.32 gives the following information:

1) the sample has a high spectroscopic purity since all the signals have similar intensity. Aminoacidic impurities appear as signals with factionary intensity;

2) many signals are grouped in a very restricted region of the spectrum. This implicates a reduced or absent anisotropy of their chemical environment. The latter is obtained only by a well defined structure. In the hypothetical case of a completely linear protein, the resultant 2D spectrum would be characterized by only one broad peak collecting all signals.

The present spectrum, instead, indicates the coexistence of two types of protons: one belonging to a flexible molecular region and the other one belonging to part of the molecule characterized by a rigid structure.

The more reliable hypothesis concerning CLSP implicates a rigid structure for one domain and a flexible structure for the other one. To verify such a hypothesis, the two domains were separately expressed by applying the above described molecular biology techniques.

The spectra recorded on the isolated C-terminal domain are reported in figure 4.33 (^{1}H-NMR) and in the figure 4.34 (^{1}H-^{15}N HSQC). This domain shows spectra compatible with a rigid structure, since very intense methyl signals below 0.5 ppm, and well resolved amidic signals above 8.0 ppm are observable.

The spectra of the isolated N-terminal domain are reported in figure 4.35 (^{1}H-NMR) and in the figure 4.36 (^{1}H-^{15}N HSQC). In the ^{1}H-NMR spectrum, the methyl signals are completely absent. Also the HSQC spectrum of the N-terminal domain presents grouped signals in two restricted regions, indicative of an isotropic environment which arises from a flexible domain. The N-terminal, however, is not purified in an unfolded state, since its 1H-NOESY spectrum shows cross-peaks indicative of the presence of secondary elements such as α-helices. Anyway, despite the presence of secondary structure, the N-terminal domain of CLSP could not be "photographed" by NMR. This result is indicative of high protein dynamism which, may be speculated, is strictly connected with its function (Capozzi et al, 2006).

On the contrary, the C-terminal domain, whose structural characterization was subsequently performed, shows a spectrum with completely dispersed signals in the region of the amidic groups (Babini et al., 2006). This finding demonstrates a different structural behaviour of two domains of the same protein, sharing two identical structural motifs, both consisting of a 4-helix bundle, but the N-terminal one with high flexibility, and the C-terminal one structurally ordered.

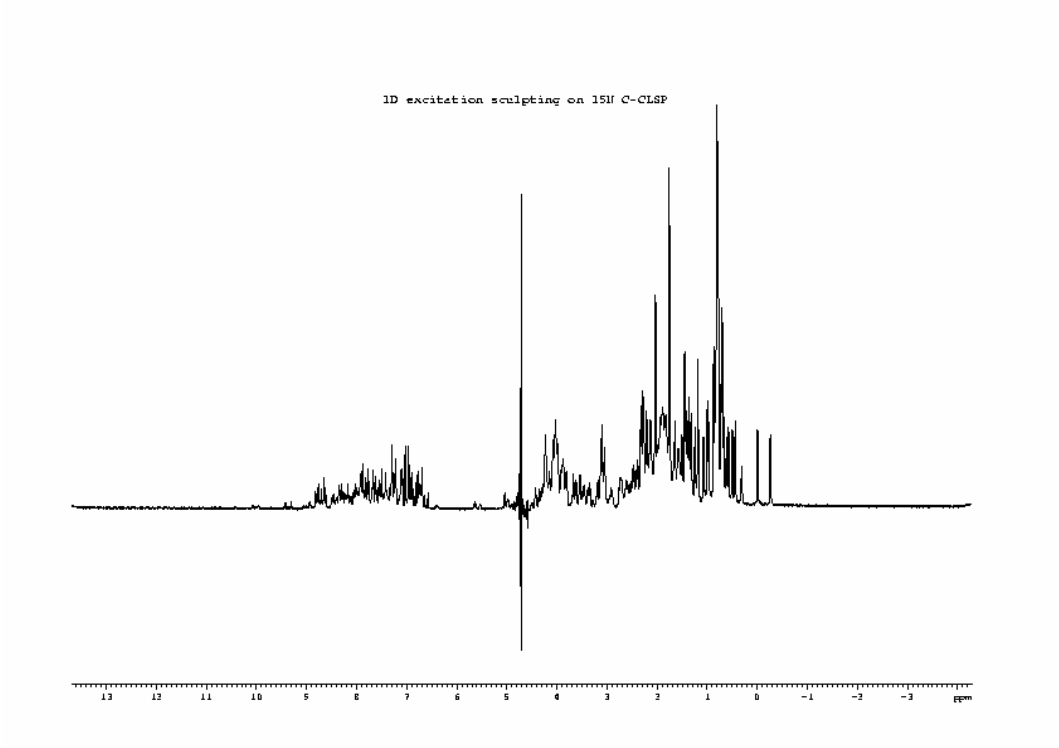


Fig. 4.33 1D-NMR spectrum of rC-CLSP domain.

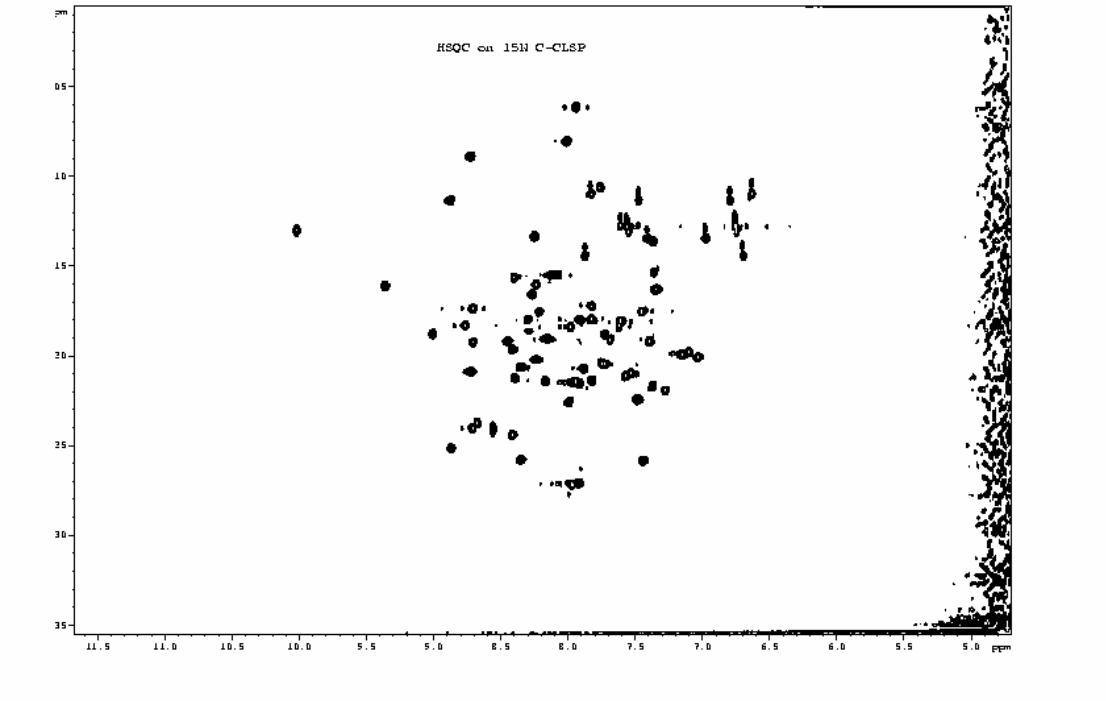


Fig. 4.34 2D-HSQC NMR spectrum of rCc-CLSP domain.

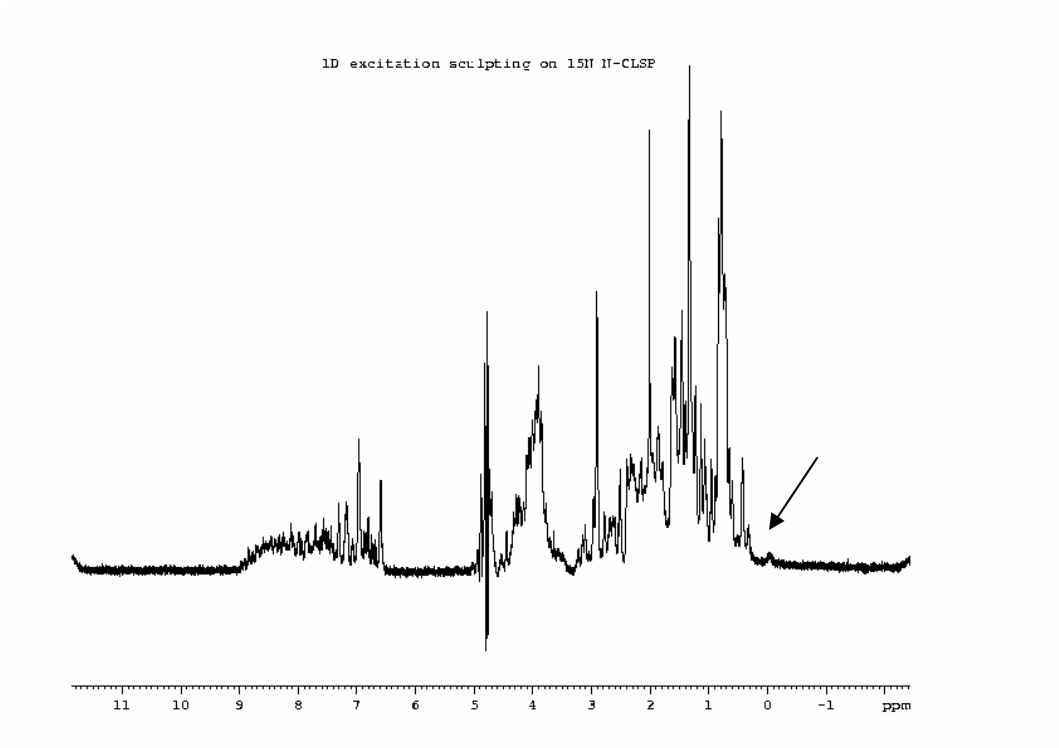


Fig. 4.35: 1D-NMR spectrum of rN-CLSP domain.

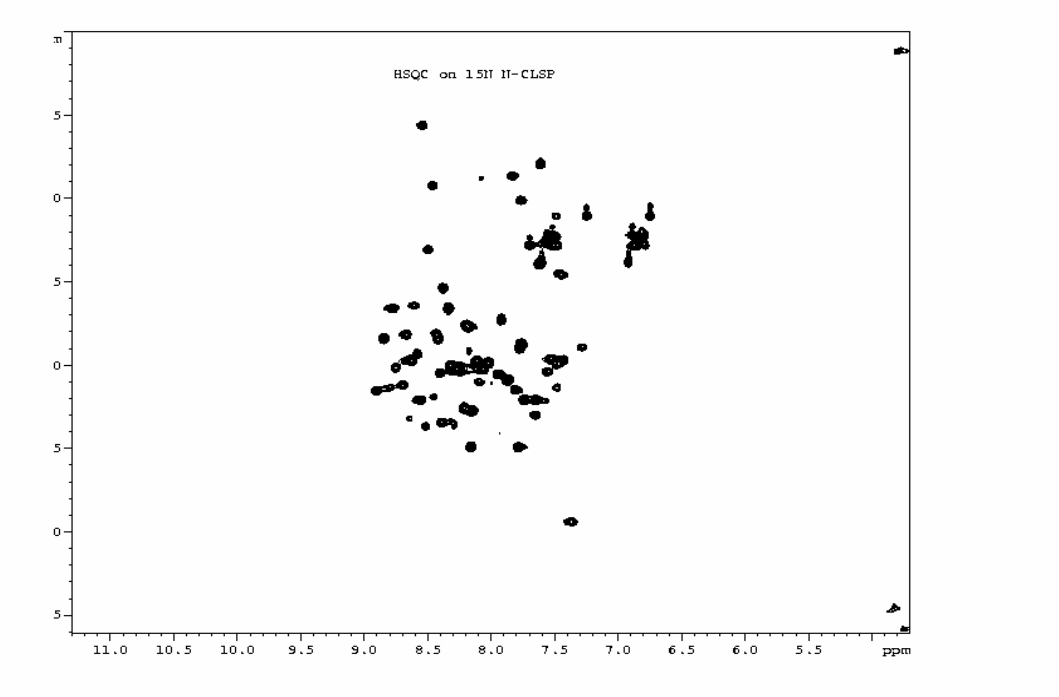


Fig. 4.36 2D-HSQC NMR spectrum of r N-CLSP domain.

In the figure 4.37 the models of CLSP (A.) and of its N-terminal domain (C.) are reported. The models have been calculated by a colleague in our research group, on the basis of the available information. In figure 4.53 B, the NMR structure of the C-terminal domain, resolved by using the samples prepared in the present research project, is showed (2B1U PDB, Babini et al., 2006). The structure of the whole protein was calculated by merging the calculated structure of the N-terminal domain and that recorded on the C-terminal domain. The figure shows that the C-terminal domain (B.) is more compact, while the N-terminal domain permits only a few interactions among the aminoacids belonging to helix 1 and helix 3 (Babini et al., 2006). For this reason, a more flexible domain was expected.

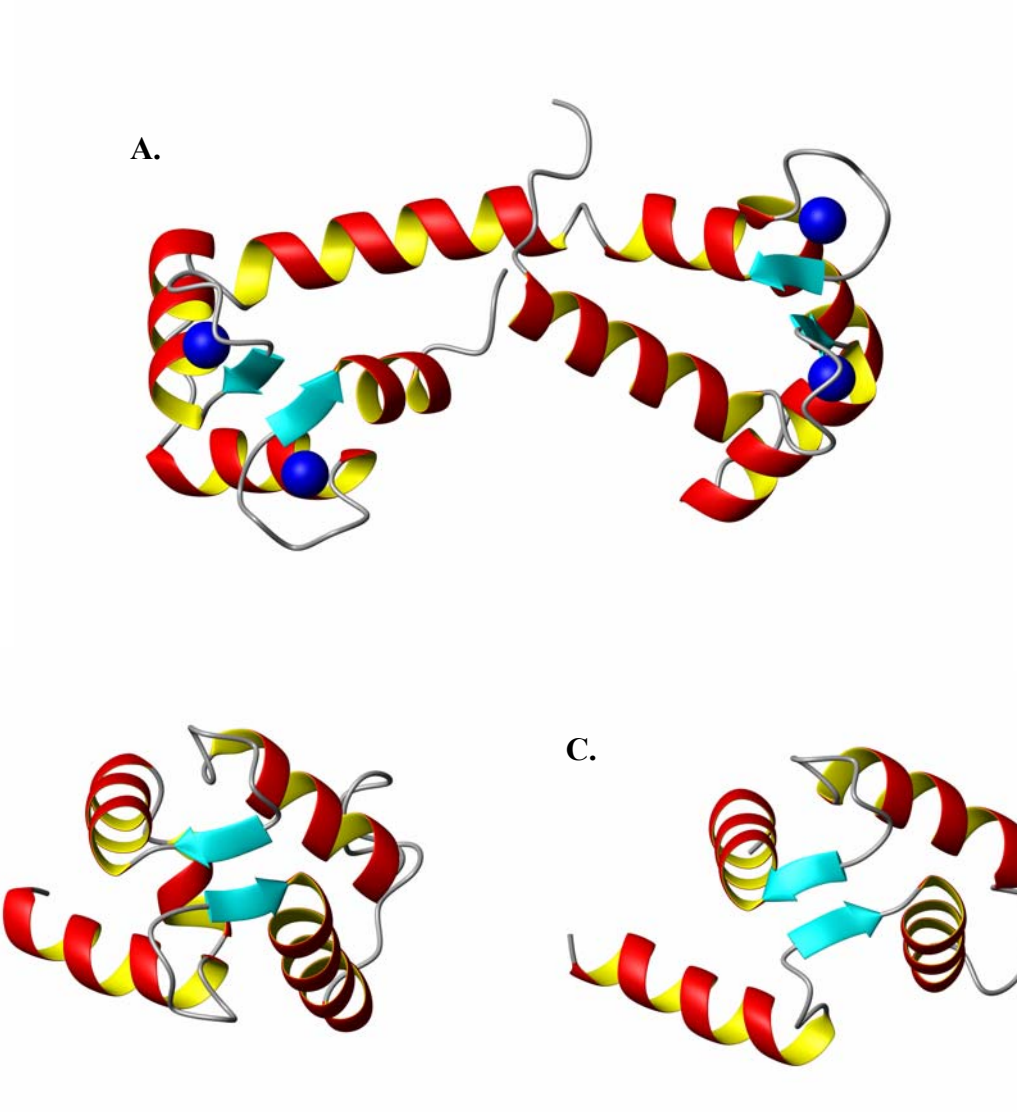


Fig. 4.37 A. Model calculated for the protein CLSP composed by four EF-hand motifs. The two globular domains, connected by a short flexible linker, are able to bind four calcium ions. **B.** NMR structure of C-terminal domain resolved by using the protein samples obtained in the present research project. **C**. A model calculated for the N-terminal domain of the same protein.

4.8 PRODUCTION AND PURIFICATION OF RECOMBINANT *S100A16* PROTEIN

S100A16 protein is a new and unique member of the EF-hand Ca^{2+}-binding proteins. S100 proteins are cell- and tissue-specific and are involved in many intra- and extracellular processes through interacting with specific target proteins. In the central nervous system S100 proteins are implicated in cell proliferation, differentiation, migration, and apoptosis as well as in recognition. S100 proteins became of major interest because of their close association with brain pathologies, for example depression or Alzheimer's disease (Cox JA et al, 2006). The amonoacidic sequence and composition is reported below.

>Q96FQ6|S10AG_HUMAN Protein S100-A16 - Homo sapiens (Human)
MSDCYTELEKAVIVLVENFYKYVSKYSLVKNKISKSSFREMLQKELNHMLSDTGNRKAAD
KLIQNLDANHDGRISFDEYWTLIGGITGPIAKLIHEQEQQSSS

Number of amino acids: 103
Molecular weight: 11801.4
Theoretical pI: 6.28

Amino acid composition:

Arg (R)	3	2.9%
Asn (N)	6	5.8%
Asp (D)	6	5.8%
Cys (C)	1	1.0%
Gln (Q)	5	4.9%
Glu (E)	8	7.8%
Gly (G)	5	4.9%
His (H)	3	2.9%
Ile (I)	8	7.8%
Leu (L)	10	9.7%
Lys (K)	10	9.7%
Met (M)	3	2.9%
Phe (F)	3	2.9%
Pro (P)	1	1.0%
Ser (S)	11	10.7%
Thr (T)	4	3.9%
Trp (W)	1	1.0%
Tyr (Y)	5	4.9%
Val (V)	5	4.9%
Asx (B)	0	0.0%
Glx (Z)	0	0.0%
Xaa (X)	0	0.0%

Total number of negatively charged residues (Asp + Glu): 14
Total number of positively charged residues (Arg + Lys): 13
Formula: $C_{522}H_{828}N_{140}O_{163}S_4$
Total number of atoms: 1657
Formula: $\mathbf{C_{522}H_{828}N_{140}O_{163}S_4}$
Atoms number: 1657

FT	DOMAIN	12-47	EF-hand 1.
FT	DOMAIN	54-89	EF-hand 2.

Tab. 4.20 Principal chemical and physical characteristics of S100a16 protein.

4.8.1 Large scale production of recombinant protein S100A16

The recombinant protein S100A16 was produced on a large scale to obtain sufficient amount for NMR analysis. The host strain used for expression of the S100A16 gene, cloned in the vector pETG-30A was *E.coli* BL21 GOLD (colony 10) expressing the protein partially in soluble form, fused with GST+(His)6-tag. The growth was performed in 3 litres of 2xYT medium and ampicillin (100μg/ml) as describe at par 3.11. The protein was expressed partially in soluble form with a yield of protein of 20mg/L. From each growing culture 1 ml of broth was taken to verify the presence of the protein by SDS-page as showed in figure 4.38.

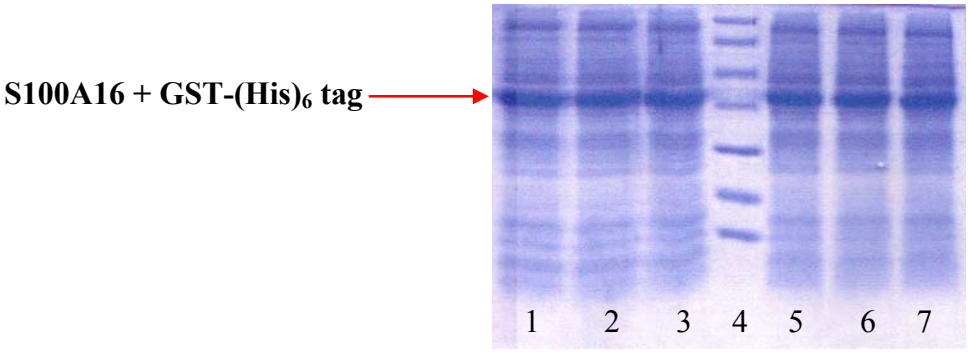


Fig. 4.38 SDS-page concerning the expression of S100A16 protein. Line 1, 2, 3, 5, 6 and 7: the over-expressed bands correspond to S100A16 protein fused with the specific GST-$(His)_6$ tag (40.8 KDa). Line 4: Protein Molecular Weight Marker Standard composed of seven proteins with known molecular weight (11.6KDa, 66.2KDa, 45KDa, 35KDa, 25KDa, 18.4 KDa, 14.4KDa)

4.8.2 Purification of recombinant protein S100A16

The soluble fraction was separated from cell debris and unbrocken cells by centrifugation. Supernatant containing soluble recombinant protein was applied to a chromatography column. Purification of the soluble S100A16 was made using a standard protocol for Histidine-tagged and GST-tagged proteins as described at par3.12.2. The steps followed were:

- ✓ metal chelate affinity chromatography and glutathione affinity chromatography (GSTrap, Amersham-Biosciences) to purify the protein before the cleavage. The presence of the protein in the GSTrap eluted was confirmed by SDS-PAGE, as showed in figure 4.39.

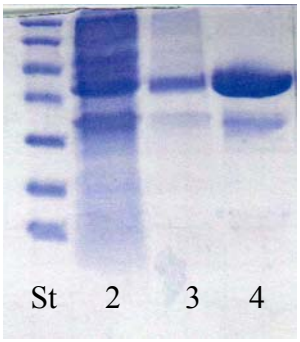


Fig. 4.39 SDS-page of protein S100A16 eluted from GSTrap column. The protein sample (line 4) was isolated from many of the other bacterial proteins visible in the flow through (line 2) and in the column washing (line 3). Line 1: Protein Molecular Weight Marker Standard described in figure 4.26.

✓ cleavage with Actev Protease to remove GST-$(His)_6$ tag.. As showed in figure 4.40 B, the cleavage was correctly performed, in fact after 15 hours in the sample are visible three bands corresponding to cut protein (S100A16, 11.8 KDa), to uncut protein (S100A16+ GST-$(His)_6$ tag, 40.8 KDa) and to GST-$(His)_6$ tag (29 KDa).

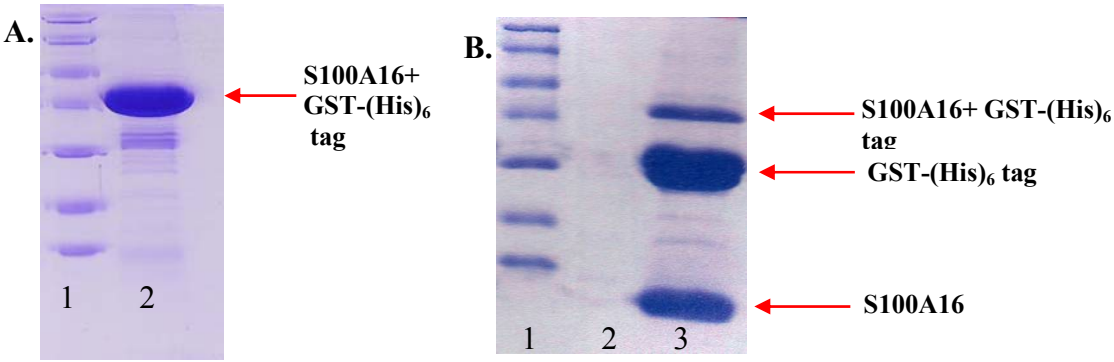


Fig. 4.40 A. SDS-page of S100A16 a after purifications into His-tag and GST-tag affinity columns before cleavage (line 2), Standard (line 1). **B.** Protein sample after digestion with Actev Protease (line 3), Standard (line 1).

After the cleavage, in the reaction mixture a fine particulate was visible. To verify the presence of the protein, the sample was centrifuged for a few minutes and the pellet was loaded on SDS-page (figure 4.41). As showed in the line 2 of the page, many bacterial protein precipitated together with some S100A16.

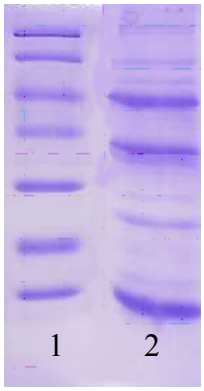


Fig. 4.41 SDS-page concerning the pellet debris formed during the digestion with Actev (line 2). Standard (line 1).

The tendency to precipitate could be due to a not correct folding state or to the fact that the experimental conditions chosen for maximal activity of the Actev protease are not the optimal for the proteins.

To find the right conditions for the protein stability, the protein was dialysed with eight different buffers, whose composition is reported in tab 4.21. After a few hours from the beginning of the dialyses, it was evident that the protein was more stable in the buffer G, selected in yellow in the table, as only in this buffer precipitation was not observed.

Buffer	Composition
A	Na Acetate pH 4.8/5; 0.5 mM $CaCl_2$
B	Tris HCl pH 8.8; 0.5 mM $CaCl_2$
C	Tris HCl pH 7.5; 0.5 mM $CaCl_2$; 100 mM NaCl
F	Tris HCl pH 7.5; 2 mM DTT; 100 mM NaCl
G	Tris HCl pH 7.5; 2 mM DTT; 5 mM EDTA; 100 mM NaCl
H	phosphate pH 7.0
I	Na Acetate pH 4.8/5; 0.5 mM $CaCl_2$; 500 mM NaCl
L	Tris HCl pH 8.8 ; 0.5 mM $CaCl_2$; 500 mM NaCl

Tab. 4.21 In the table the composition of the buffer used for the stability test of the S100A16 protein are reported.

At this purification step the protein was stable at pH 7.5, in reduced condition (DTT) and in the presence of 5 mM EDTA and 100 mM NaCl. So the successive purification steps were conduced using these experimental conditions. To increase the protein

stability, the cleavage buffer (describe at par 3.11) was added with 5mM EDTA and 2mM DTT.

- ✓ glutathione affinity chromatography to remove the GST-$(His)_6$ tag and uncut protein. In figure 4.42 is visible the cut S100A16 protein (line 1), isolated with GStrap FF column. The uncut fraction and GST (line 2), detained on column, were eluted with 10mM reduced glutathione.

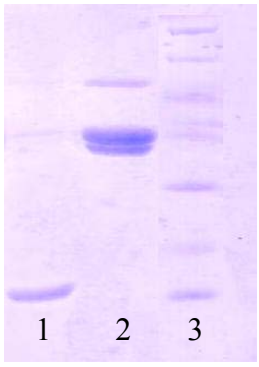


Fig. 4.42 SDS-page of S100A16 cut (line 1), isolated with GStrap FF column. The uncut fraction and GST (line 2), were detained on column. Standard (line 3).

- ✓ size exclusion chromatography to isolate the protein from the impurities. The purity of the final sample was checked by SDS-page figure 4.43.

Fig.4. 43 SDS-page of the purified S100A16 (line 2).Standard (line 3).

Purified S100A16 concentration (4mM) was determined by measuring absorbance at 276nm, using the respective extinction coefficient ε_{276} = 12950 $M^{-1}cm^{-1}$, as calculated by ProtPar parameter tools (http://expasy.org/tools/protoaram.html).

4.8.3 Preliminary screening of folding of recombinant protein S100A16

The purified samples, estimated to have concentration >1mM, were analysed by NMR spectroscopy to verify the folding state. The presence of specific signals in the 1D- and ^{1}H-NOESY spectra is an indication of the correct structural state of the protein. In figure 4.44, the 1D-NMR spectrum of S100A16 shows the presence of signals in the 8-10 ppm range, corresponding to amidic protons of polypeptide chains, explainable with the presence of secondary structure elements (α-helix).

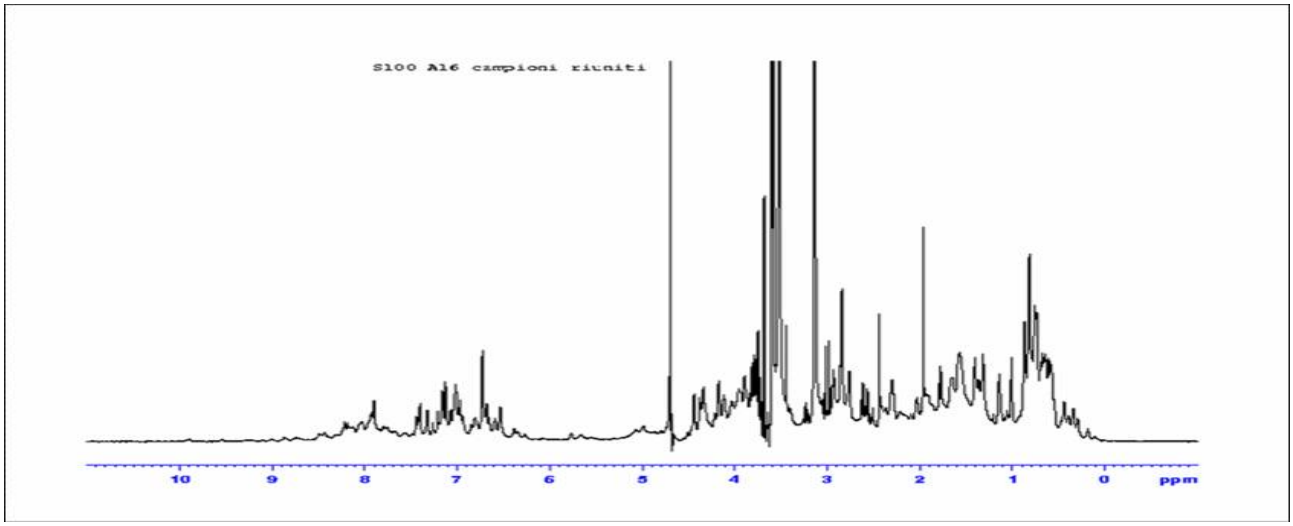


Fig.4.44 1D-NMR spectrum of S100A16.

Moreover, the presence of very sharp signals in the aromatic region (6-8 ppm), indicative of long relaxation times, shows that the protein is in a monomeric form. The 2D-1H-NOESY spectrum shown in figure 4.45, is compatible with a stable structure, because of the presence of a high number of *cross-peaks* between the aromatic (6-8 ppm) and aliphatic (0-4 ppm) regions, indicating stable interactions between aminoacids.

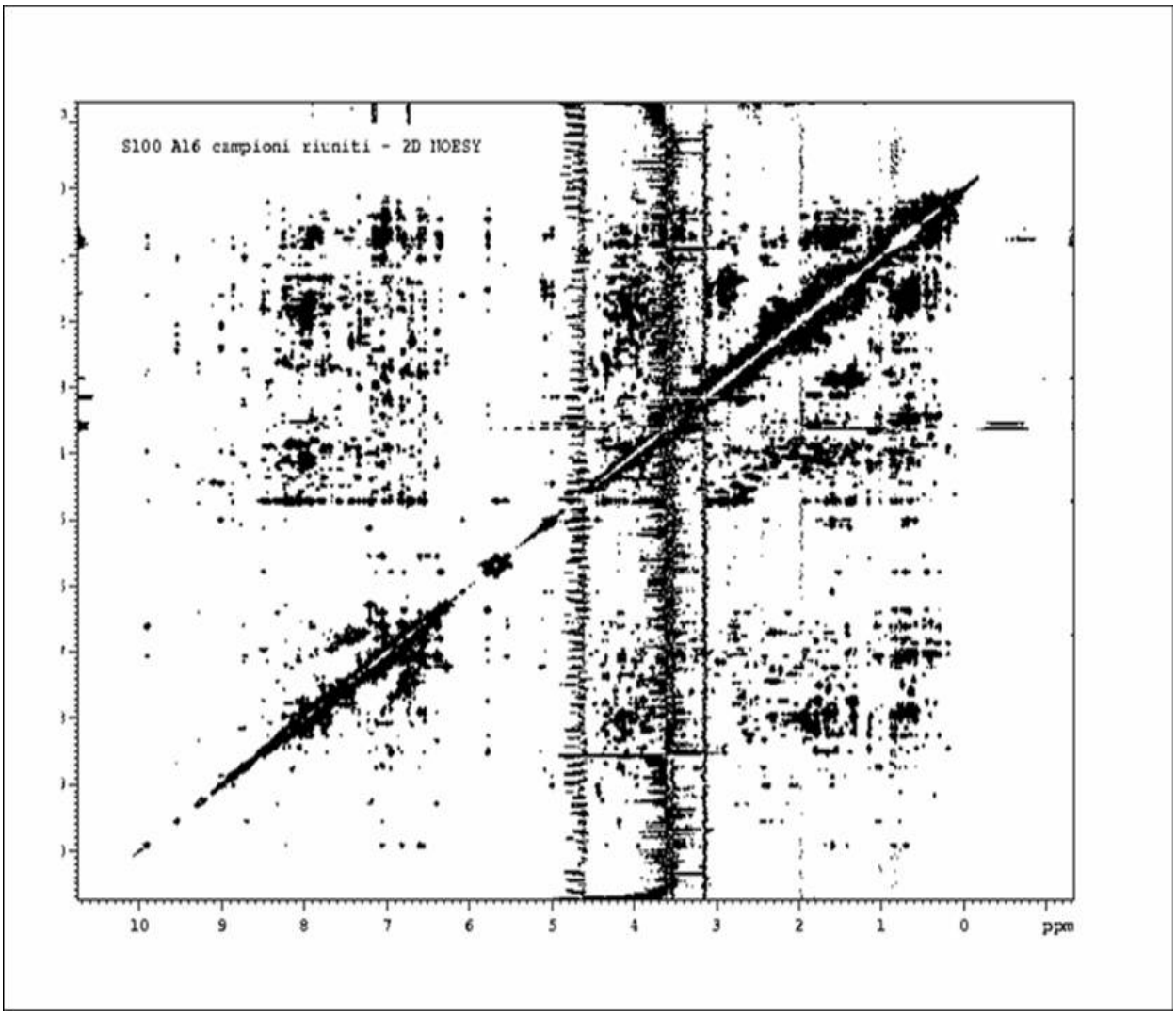


Fig. 4.45 2D-1H-NOESY NMR spectrum of S100A16 protein.

4.9 PRODUCTION AND PURIFICATION OF THE ISOFORM α OF THE *Endothelial differentiation-related factor 1* (EDF-1)

Endothelial Differentiation Factor (EDF)-1 is a calmodulin binding protein involved in the repression of endothelial cell differentiation, a crucial, late step in angiogenesis (Bolognese et al, 2006). EDF-1-CaM interaction is modulated by the concentrations of Ca^{2+} and by the phosphorylation of EDF-1 by protein kinase C.EDF-1 is the human homologue of the silkworm multiprotein bridging factor (MBF)-1, a transcriptional cofactor that mediates transactivation by stabilizing the protein-DNA interaction. Interestingly, a neuronal calmodulin-binding peptide, termed CAP 19, is the rat homologue of EDF-1. Human and rat proteins possess a putative IQ domain that could mediate their binding to CaM. Indeed, calmodulin-binding proteins share a conserved region of about 20 amino acids, designated the IQ motif, that contains a CaM-binding domain and a protein kinase C (PKC) phosphorylation site. The IQ motif was originally identified in neuromodulin that concentrates CaM at specific sites in neurons (Avon et al., 2000). The characteristic of the protein are reported below.

>O60869|EDF1_HUMAN Endothelial differentiation-related factor 1 - Homo sapiens (Human).
MAESDWDTVTVLRKKGPTAAQAKSKQAILAAQRRGEDVETSKKWAAGQNKQHSITKNTAK
LDRETEELHHDRVTLEVGKVIQQGRQSKGLTQKDLATKINEKPQVIADYESGRAIPNNQV
LGKIERAIGLKLRGKDIGKPIEKGPRAK

Number of amino acids: 148
Molecular weight: 16368.6
Theoretical pI: 9.95

Amino acid composition:

Ala (A)	15	9.9%
Arg (R)	10	6.6%
Asn (N)	5	3.3%
Asp (D)	8	5.3%
Cys (C)	0	0.0%
Gln (Q)	11	7.2%
Glu (E)	11	7.2%
Gly (G)	13	8.6%
His (H)	3	2.0%
Ile (I)	10	6.6%
Leu (L)	10	6.6%
Lys (K)	20	13.2%
Met (M)	1	0.7%
Phe (F)	1	0.7%
Pro (P)	5	3.3%
Ser (S)	7	4.6%
Thr (T)	11	7.2%
Trp (W)	2	1.3%
Tyr (Y)	1	0.7%
Val (V)	8	5.3%
Asx (B)	0	0.0%
Glx (Z)	0	0.0%
Xaa (X)	0	0.0%

Total number of negatively charged residues (Asp + Glu): 19 **Total number of positively charged residues (Arg + Lys):** 30 **Formula:** $C_{708}H_{1194}N_{222}O_{220}S_1$ **Total number of atoms:** 2345 **Charge at pH 7.5:** 1.0703084178577

Tab. 4.22 Principal chemical and physical characteristics of EDF-1.

4.9.1 Large scale production of recombinant protein EDF-1 iso α

The recombinant EDF-1 isoα protein was produced on a large scale to obtain sufficient amount for NMR analysis. The host strain used for expression of the gene cloned in the vector pETG-20A was *E.coli* BL21 Codon plus (16 colony) expressing the protein fused with Trx+(His)6-tag in inclusion bodies. The growth was performed in 3 litres of 2xYT medium and ampicillin (100μg/ml) as describe at par 3.11. At the end of the growth, after 9 hours from induction, 21 gr of cellular pellet were collected. The presence of the protein in the culture was verified by SDS-page as showed in figure 4.46.

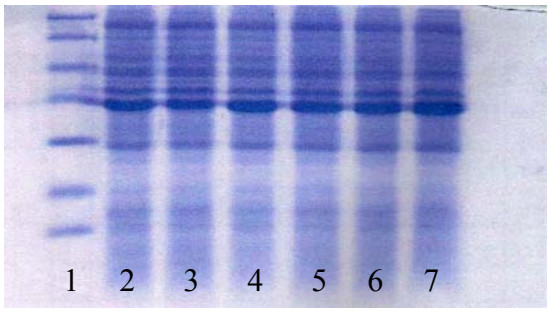


Fig. 4.46 SDS-page concerning the expression of EDF-1 iso α protein. Line 2, 3, 4, 5, 6 and 7: the over-expressed bands correspond to EDF-1 iso α protein fused with the specific Trx-$(His)_6$ tag (33 KDa). Line 1: Protein Molecular Weight Marker Standard composed of seven proteins with known molecular weight (11.6KDa, 66.2KDa, 45KDa, 35KDa, 25KDa, 18.4 KDa, 14.4KDa)

4.9.2 Purification of recombinant protein EDF-1 iso α

The EDF-1 iso α was produced in inclusion bodies. Purification of EDF-1 iso α was made using a standard protocol for Histidine-tagged with a denaturation refolding step on the nickel chelate column. The steps followed were:

- ✓ denaturation refolding step on the nickel chelate column. The inclusion bodies were isolated, solubilized and refolded as describe at paragraph 3.11 At the end of the

refolding step, the protein was eluted from Hitrap column with 300mM Imidazol. The presence of the protein was confirmed by SDS-PAGE, as showed in figure 4.47. The presence of the protein' band proves that the protein, previously solubilized with 6 M Guanidin, was retained from the column and was refolded. Moreover, being the eluted solution clear, the protein seems to have acquired a soluble and stable structure.

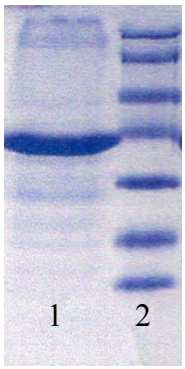


Fig. 4.47 SDS-page of EDF-1 iso α protein eluted from Hitrap column after refolding step (line 1). Line 2: Protein Molecular Weight Marker Standard described in figure 4.

✓ cleavage with Actev Protease to remove Trx-$(His)_6$ tag. As showed in figure 4.48, the cleavage was correctly occurred, in fact in the third line of the page a double band is clearly observed, one correspondent to the protein (16.7 KDa) and an other one to his tag (16.6 KDa). In the second line, the protein sample before the digestion was loaded.

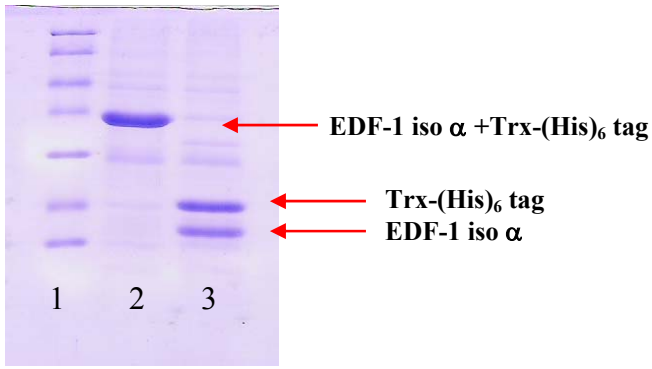


Fig. 4.48 SDS-page of EDF-1 iso α before the cleavage (line 2). Protein sample after digestion with Actev Protease (line 3), Standard (line 1).

✓ The purification of the cut protein was made by cationic exchange chromatography. The column was conditioned with Hepes buffer pH 7.4. At this pH condition, the

protein was retained on the column for its positive charge, while Trx-(His)6-tag, negatively charged, was eluted during the loading. The protein was eluted by continuous gradient against Hepes buffer pH 7.4 and 1M NaCl, until NaCl concentration was 400mM, at a flow rate of 0.5 ml/min. In figure 4.49 the SDS-page showing the protein during the different steps is reported.

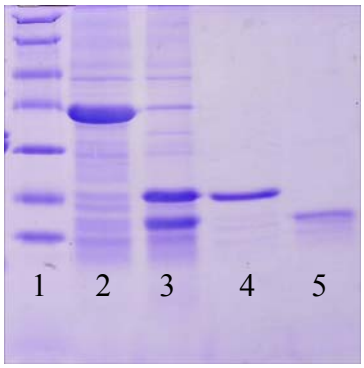


Fig. 4.49 SDS-page of EDF-1 concerning the purification steps: protein before cut (line 2), protein cut (line 3), protein detained on column and successively eluted with NaCl 400mM (line 4), fusion tag [Trx-(His)6] eluted during loading (line 5).

Purified EDF-1 concentration, 1mM, was determined measuring the absorbance at 276 nm, using the extinction coefficient ε_{276} = 12450 M-1cm-1, as calculated by ProtPar parameters tools (http://us.expasy.org/tools/protparam.html).

4.9.3 Preliminary screening of folding of recombinant protein EDF-1 iso α

The purified samples were analysed by NMR spectroscopy to verify the folding state. The presence of specific signals in the monodimensional and HSQC spectra is an indication of the correct structural status of the protein in the sample. The monodimensional spectrum of EDF-1 is reported in figure 4.50. The absence of broad signals indicates that the protein does not form aggregates in its quaternary structure, but is a stable monomeric form. Moreover, the presence of signals in the 7.5-11 ppm range where the amidic protons resonates, indicates that the backbone of the protein chain is protected, well inside the protein globule, against the exchange with solvent.

Signals in the methyl spectral region (0.0-1.0 ppm) are not as well resolved as typical of rigid structures. In order to have a proof of the correct folding of the protein, the sample was investigated by analysis of its HSQC spectrum. The figure 4.51, shows the HSQC spectrum

characterized by the presence of signals spread in both the 1H and 15N, dimensions suggesting a stable tertiary structure typical of a correct folding.

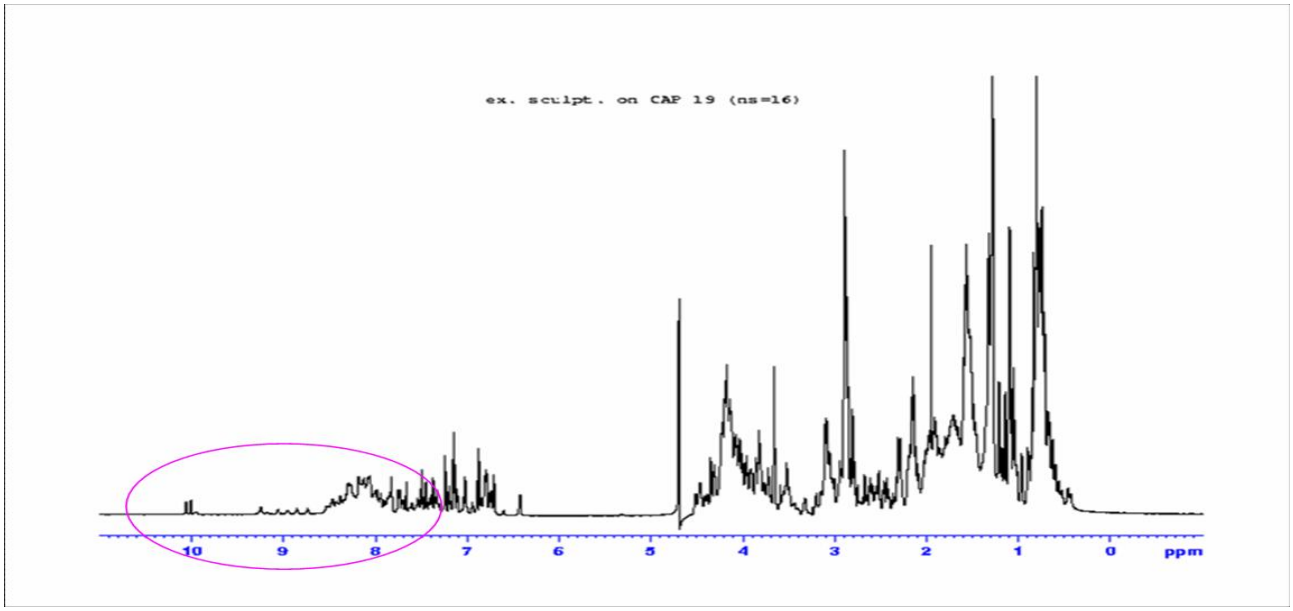


Fig.4.50 1D-NMR spectrum of EDF-1 iso α.

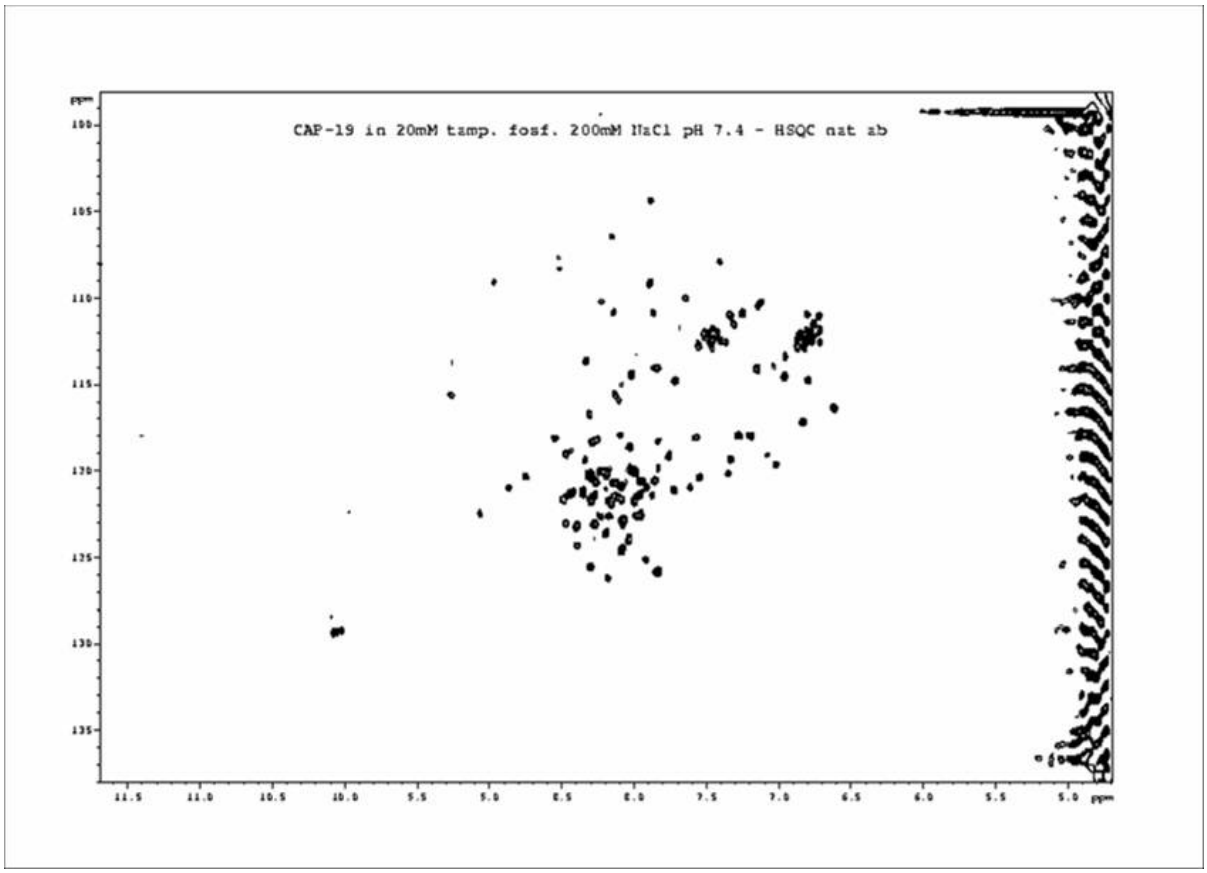


Fig. 4.51 2D-HSQC NMR spectrum of EDF-1 iso α.

4.9.4 Stability test of recombinant protein EDF-1 iso α

The protein structure wasn't determined since a proteasic activity was observed, confirmed by electrophoresis analysis (appearance of bands corresponding to fragments of the protein). The structural NMR analysis requires many days for spectra acquisition during which the sample should be stable. Some test to find better experimental conditions and block the protein degradation were conduced.

1. The protein was dialysed in different conditions reported in tab . The degradative state was checked after 2 hours , 20 hours and 4 days by SDS-page in figure 4.52

Buffer 1	Buffer 2	Buffer 2
Tris 50 mM NaCl 200 mM PMSF 1mM pH 6.5	Phosphate 20mM Inhibitor Protease cocktail pH 7.2	Tris 50 mM NaCl 200 mM pH 8.0

Tab. 4.23. In the table the composition of the tested buffer are reported.

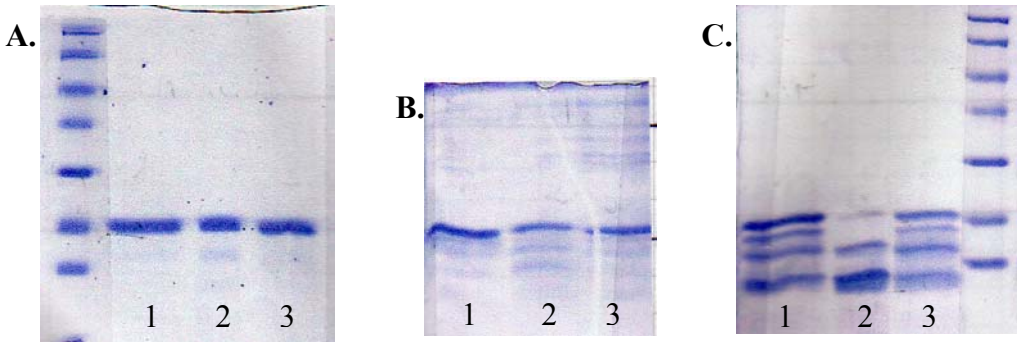


Fig. 4.52 SDS-pages concerning the degradative state of the EDF-1 iso α sample after 2 hours (Page A), 20 hours (Page B) and after 4 days (page C), in the three different conditions (lines 1, 2, 3) reported in table 4.23.

Observing the gels it's possible to note that the buffer 2 was the worst condition, although also in the other conditions the degradation wasn't stopped.

2. During the cationic exchange chromatography, the buffers reported in table 4.24 were experimented. The inhibitor protease cocktail, PMSF 1mm and EDTA 5mM were added after the elution of the cut protein from the column. After six days the samples were checked by SDS-page in figure 4.53.

Buffer 4	Buffer 5	Buffer 6	Buffer 7
MOPS pH 7.4 PMSF 1mM cocktail	MOPS pH 7.4 PMSF 1mM EDTA 5mM cocktail	HEPES pH 7.4 PMSF 1mM cocktail	HEPES pH 7.4 PMSF 1mM EDTA 5mM cocktail

Tab. 4.24 In the table the composition of the tested buffer are reported.

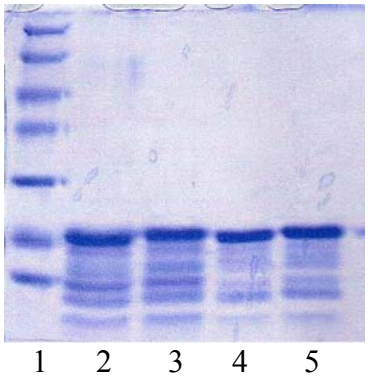


Fig. 4.53 SDS-pages concerning the degradative state of the EDF-1 iso α sample in different conditions. Line 2: buffer4; line3: buffer5; line 3: buffer6; line 5: buffer7

Also in this case the sample degradation wasn't blocked, even if the buffer 6 and 7 seemed to be better than 4 and 5 one.

5 CONCLUSIONS

In the last decade, the number of available allergen sequences has rapidly increased due to the introduction of molecular biology techniques for allergen characterization. Sequence analysis of allergen-encoding cDNAs revealed the presence of typical calcium-binding motifs, termed EF-hands, within allergens from various sources. Parvalbumin represented the first calcium-binding allergen described. It represents the major allergen in fish and belongs to a subfamily of closely related calcium-binding proteins that contain two functional and one silent EF-hand motif.

This thesis work is part of a high-throughput project faced to the production of recombinant proteins belonging to one of the principal classes of the calcium-binding proteins causing food allergies, and their targets, selected for their relevance in cellular metabolism of all organisms. It has been demonstrated that calcium binding to the loop induces conformational changes that causes the recognition of the IgE-allergen (Dissertori O. et al., 2007). These results indicate that the interaction of these proteins with the physiological calcium plays an important role in the biologic and allergenic activity. The thesis' aim was to obtain pure recombinant soluble protein for structural characterization by NMR. This information is essential to understand the mechanism of protein functions and for the general purpose of identifying new drug targets.

The cloning and the high-throughput expression were performed by applying the *Gateway* system. The Gateway technology is a universal cloning method based on a site-specific recombination which provides a rapid and highly efficient way to move DNA sequences into multiple vector system. One of the advantages of this method is the opportunity to use different types of DNA sequences (PCR products, cDNA clones, restriction fragments) and multiple destination vectors. In fact, the DNA sequences initially are inserted in a plasmid (pENTR) and then transferred in multiple destination vectors (pDEST) with an only one recombination reaction. The destination vectors are characterized by specific fusion tag (protein or peptide), which allows to simplify the subsequent purification steps by specific affinity chromatography columns. Additionally, certain affinity tags have a beneficial effect on protein solubility especially in bacterial protein expression. This is an important problem in the recombinant proteins production. It often happens, in fact, that the recombinant protein produced in a host organism is not correctly structured. Each protein, in this sense, has a different "behaviour" and the more applied strategy to increase the successful probabilities in protein expression consists in the experimentation of more vector-host combinations. On this

subject, in the protein expression of complex multicellular organisms, big advantages could be derived by the employment of organisms different from *Escherichia coli* (the simplest and by far the most widely used organism for protein expression) as the yeast (*Saccharomices cerevisiae*) or insect (*Baculovirus*). In these cells some of the post-translational modifications typical of the eukaryotic cells, which could be essential for the acquisition of the correct three-dimensional protein structure, can happen. In definitive, once a pENTR containing the gene of interest is generated, with single recombination reaction it is possible to transfer simultaneously the gene to different pDEST, and to obtain the protein expressed with different peptide or fusion proteins in different host organism.

The genes coding for the selected recombinant proteins were successfully cloned into pENTR plasmids by recombination reaction. This result was obtained for 32 of the 39 genes initially selected. Subsequently the genes were successfully transferred from each pENTR into five different expression vectors (pETG-20A, pETG-30A, pETG-60A, pDEST17, pDEST15) carrying specific protein fusion tags. For protein expression three different *E.coli* host strains were used (BL21 *Gold, pLys, Codon plus*) and tested for the expression screening which showed that the selected proteins were well expressed with almost three of the five fusion tags in all *E. coli* host strains. Moreover, from the solubility screening, it results that the Calmodulin targets proteins were expressed in inclusion bodies by all the host strains with an only one exception, the protein Endothelial differentiation related factor1 (EDF-1). There is instead a big variability in the solubility of the EF-hand proteins; in fact, in most cases, the proteins were produced in inclusion bodies and only with few tags, different for each protein, were expressed in soluble form. This result is very important for the successive protein purification steps on a larger scale. In fact, a soluble protein, fused with a specific affinity tag could be easily purified using a few chromatographic passages. This is the case of the of the recombinant protein S100A16, and of the C-terminal domain of the protein CLSP, which were extracted from 3 litres of bacterial culture and purified with only up to three steps, until a sample was obtained with high concentration suitable for the preliminary folding analysis by NMR.

Purification of recombinant CLSP and of its N-terminal domain, both produced in a soluble form, required a denaturation refolding step on the nickel chelate column. The same protocol was applied for the calmodulin target Endothelial differentiation-related factor1 (EDF-1), isoform α, produced in inclusion bodies. The method allowed the production of pure protein samples, perfectly soluble and with a sufficient concentration for the NMR analysis.

In the results chapter, the spectra obtained at the European Centre of Magnetic Resonances (CERM) of the University of Florence are reported. From the spectra analysis, it results that the C-terminal domain of the protein CLSP was correctly folded and the protein structure was subsequently determined (Babini et al., 2005).

The S100A16 NMR spectra showed that the soluble protein sample was correctly folded. Such a result opens the way to the subsequent step of structure characterization. Actually, the protein structure is unpredictable by bioinformatics methods (homology modelling), because of the low sequence homology with proteins with known structure.

The NMR spectra of EDF-1 and of N- terminal CLSP reveal that the proteins are very flexible, such as that it is impossible to obtain the mean structure by NMR. This result is not indicative of an unfolded protein; on the contrary, the protein dynamics may be strictly related, and point, to its physiological functions (Capozzi et al, 2006).

In the light of the results gained during this thesis work, only a high throughput strategy permits one to characterized at least one method useful to obtain a protein sample suitable for structural studies. Moreover, we found out that a number of soluble proteins, expressed and purified with protocols as much as possible optimized, even after the application of many alternative refolding procedures, show high flexibility. The flexibility is implicitly related to the physiological mechanism by which these proteins developed their function and, probably, their allergenicity was derived from this peculiarity. These proteins are also able to assume more exchangeable conformations, so that almost one of them may be recognized as an allergen. This result was clearly stressed in a paper, recently published by our group {REF, which reviews and extrapolates a general trait on the dynamics of this class of proteins as emerging by studies published by most of the research groups involved in this field. In fact, as it is concluded in this work, the structure of EF-hand proteins is highly variable and couldn't be always "photographed" by NMR; this is the reason by which we have considered the necessity to adopt a high-throughput cloning strategies to define, by chances, the conditions to obtain at least one quasi-stable structure, that permits to build calculated models, where it is possible to find the one responsible of the IgE recognition.

By this way, it is possible to foresee a structural conformation of the protein which is superimposable with the structure of a certain allergen, in order to hypothesize a potential cross-reactivity.

Last, but not least, the results obtained during the present PhD thesis's work permits, once more, to point out that the simple aminoacidic sequence is not sufficient to define the protein

properties. Rather, it is required that the structural characterization, in a natural state, must be performed to evaluate the potential activities of proteins.

Annex 1

Nucleotidic sequences of the proteins selected for the research project with the specific primers designed to amplify the genes from cDNA.

Proteins	Nucleotidic sequence		
1a) + NEUROMODULIN - P17677 - 238 aa	caccATGCTGTGCTGTATGAGAAGAACCAAACAGGTTGAAAAAAATGATGACGACCAAAAGATTGAACAAGATG GTATCAAACCAGAAGATAAAGCTCATAAGGCCGCAACCAAAATTCAGGCTAGCTTCCGTGGACACATAACAAGG AAAAAGCTCAAAGGAGAGAAGAAGGATGATGTCCAAGCTGCTGAGGCTGAAGCTAATAAGAAGGATGAAGCCCC TGTTGCCGATGGGGTGGAGAAGAAGGGAGAAGGCACCACTACTGCCGAAGCAGCCCCAGCCACTGGCTCCAAGC CTGATGAGCCCGGCAAAGCAGGAGAAACTCCTTCCGAGGAGAAGAAGGGGGAGGGTGATGCTGCCACAGAGCAG GCAGCCCCCCAGGCTCCTGCATCCTCAGAGGAGAAGGCCGGCTCAGCTGAGACAGAAAGTGCCACTAAAGCTTC CACTGATAACTCGCCGTCCTCCAAGGCTGAAGATGCCCCAGCCAAGGAGGAGCCTAAACAAGCCGATGTGCCTG CTGCTGTCACTGCTGCTGCTGCCACCACCCCTGCCGCAGAGGATGCTGCTGCCAAGGCAACAGCCCAGCCTCCA ACGGAGACTGGGGAGAGCAGCCAAGCTGAAGAGAACATAGAAGCTGTAGATGAAACCAAACCTAAGGAAAGTGC CCGGCAGGACGAGGGTAAAGAAGAGGAACCTGAGGCTGACCAAGAACATGCCtga		
	Primers	Nucleotide number	Primer sequence 5'-3'
	Forward:ONEUaF3	22	CACCATGCTGTGCTGTATGAGA
	Reverse:ONEUaR4	20	TCAGGCATGTTCTTGGTCAG
1b) + NEUROMODULIN - P17677 - GAP-43$_{5-226}$ 222 aa	CaccATGAGAAGAACCAAACAGGTTGAAAAAAATGATGACGACCAAAAGATTGAACAAGATGGTATCAAACCAG AAGATAAAGCTCATAAGGCCGCAACCAAAATTCAGGCTAGCTTCCGTGGACACATAACAAGGAAAAAGCTCAAA GGAGAGAAGAAGGATGATGTCCAAGCTGCTGAGGCTGAAGCTAATAAGAAGGATGAAGCCCCTGTTGCCGATGG GGTGGAGAAGAAGGGAGAAGGCACCACTACTGCCGAAGCAGCCCCAGCCACTGGCTCCAAGCCTGATGAGCCCG GCAAAGCAGGAGAAACTCCTTCCGAGGAGAAGAAGGGGGAGGGTGATGCTGCCACAGAGCAGGCAGCCCCCCAG GCTCCTGCATCCTCAGAGGAGAAGGCCGGCTCAGCTGAGACAGAAAGTGCCACTAAAGCTTCCACTGATAACTC GCCGTCCTCCAAGGCTGAAGATGCCCCAGCCAAGGAGGAGCCTAAACAAGCCGATGTGCCTGCTGCTGTCACTG CTGCTGCTGCCACCACCCCTGCCGCAGAGGATGCTGCTGCCAAGGCAACAGCCCAGCCTCCAACGGAGACTGGG GAGAGCAGCCAAGCTGAAGAGAACATAGAAGCTGTAGATGAAACCAAACCTAAGGAAAGTGCCCGGCAGGACGA GGGTtga		
	Primers	Nucleotide number	Primer sequence 5'-3'
	Forward:ONEUbF3	21	CaccATGAGAAGAACCAAAC A
	Reverse:ONEUbR4	17	tcaACCCTCGTCCTGCC
1c) + NEUROMODULIN - P17677 - GAP-43$_{41-226}$ 186 aa	caccAGCTTCCGTGGACACATAACAAGGAAAAAGCTCAAAGGAGAGAAGAAGGATGATGTCCAAGCTGCTGAGG CTGAAGCTAATAAGAAGGATGAAGCCCCTGTTGCCGATGGGGTGGAGAAGAAGGGAGAAGGCACCACTACTGCC GAAGCAGCCCCAGCCACTGGCTCCAAGCCTGATGAGCCCGGCAAAGCAGGAGAAACTCCTTCCGAGGAGAAGAA GGGGGAGGGTGATGCTGCCACAGAGCAGGCAGCCCCCCAGGCTCCTGCATCCTCAGAGGAGAAGGCCGGCTCAG CTGAGACAGAAAGTGCCACTAAAGCTTCCACTGATAACTCGCCGTCCTCCAAGGCTGAAGATGCCCCAGCCAAG GAGGAGCCTAAACAAGCCGATGTGCCTGCTGCTGTCACTGCTGCTGCTGCCACCACCCCTGCCGCAGAGGATGC TGCTGCCAAGGCAACAGCCCAGCCTCCAACGGAGACTGGGGAGAGCAGCCAAGCTGAAGAGAACATAGAAGCTG TAGATGAAACCAAACCTAAGGAAAGTGCCCGGCAGGACGAGGGTtga		
	Primers	Nucleotide number	Primer sequence 5'-3'
	Forward:ONEUcF3	18	caccAGCTTCCGTGGACA
	Reverse:ONEUbR4	17	tcaACCCTCGTCCTGCC
2a) Endothelial differentiation-related factor 1 (CAP-19) ISO-α	caccATGGCCGAGAGCGACTGGGACACGGTGACGGTGCTGCGCAAGAAGGGCCCTACGGCCGCCCAGGCCAAAT CCAAGCAGGCTATCTTAGCGGCACAGAGACGAGGAGAAGATGTGGAGACTTCCAAGAAATGGGCTGCTGGCCAG AACAAACAACATTCTATTACCAAGAACACGGCCAAGCTGGACCGGGAGACAGAGGAGCTGCACCATGACAGGGT GACCCTGGAGGTGGGCAAGGTGATCCAGCAAGGTCGGCAGAGCAAGGGGCTTACGCAGAAGGACCTGGCCACGA AAATCAATGAGAAGCCACAGGTGATCGCGGACTATGAGAGCGGACGGGCCATACCCAATAACCAGGTGCTTGGC AAAATCGAGCGGGCCATTGGCCTCAAGCTCCGGGGAAAGGACATTGGAAAGCCCATCGAGAAGGGGCCTAGGGC GAAAtga		
	Primers	Nucleotide number	Primer sequence 5'-3'
	Forward:OCAP19aF3	16	caccATGGCCGAGAGC
	Reverse:OCAP19aR4	21	tcaTTTCGCCCTAGGCCCCT T
2b) Endothelial differentiation-related factor 1 (CAP-19) ISO-β	caccATGGCCGAGAGCGACTGGGACACGGTGACGGTGCTGCGCAAGAAGGGCCCTACGGCCGCCCAGGCCAAAT CCAAGCAGGCTATCTTAGCGGCACAGAGACGAGGAGAAGATGTGGAGACTTCCAAGAAATGGGCTGCTGGCCAG AACAAACAACATTCTATTACCAAGAACACGGCCAAGCTGGACCGGGAGACAGAGGAGCTGCACCATGACAGGGT GACCCTGGAGGTGGGCAAGGTGATCCAGCAAGGTCGGCAGAGCAAGGGGCTTACGCAGAAGGACCTGGCCACGA AAATCAATGAGAAGCCACAGGTGATCGCGGACTATGAGAGCGGACGGGCCATACCCAATAACCAGGTGCTTGGC AAAATCGAGCGGGCCATTGGTGAGTGTCCCTCCACCCTTCGCCGGGTCCGCtga		
	Primers	Nucleotide number	Primer sequence 5'-3'
	Forward:OCAP19aF3	16	caccATGGCCGAGAGC

	Reverse:OCAP19bR4	14	tcaGCGGACCCGGC
3) PEP-19 P48539 62 AA	caccATGAGTGAGCGACAAGGTGCTGGGCCAACCAATGGAAAAGACAAGACATCTGGTGAAAATGATGGACAGA AGAAAGTTCAAGAAGAATTTGACATTGACATGGATGCACCAGAGACAGAACGTGCAGCGGTGGCCATTCAGTCT CAGTTCAGAAAATTCCAGAAGAAGAAGGCTGGGTCTCAGTCCtag		
	Primers	Nucleotide number	Primer sequence 5'-3'
	Forward:OPEP19F3	20	caccATGAGTGAGCGACAAG
	Reverse:OPEP19R4	24	ctaGGACTGAGACCCAGCCT TCTT
4a) P21 P38936 164aa	caccATGTCAGAACCGGCTGGGGATGTCCGTCAGAACCCATGCGGCAGCAAGGCCTGCCGCCGCCTCTTCGGCC CAGTGGACAGCGAGCAGCTGAGCCGCGACTGTGATGCGCTAATGGCGGGCTGCATCCAGGAGGCCCGTGAGCGA TGGAACTTCGACTTTGTCACCGAGACACCACTGGAGGGTGACTTCGCCTGGGAGCGTGTGCGGGGCCTTGGCCT GCCCAAGCTCTACCTTCCCACGGGGCCCCGGCGAGGCCGGGATGAGTTGGGAGGAGGCAGGCGGCCTGGCACCT CACCTGCTCTGCTGCAGGGGACAGCAGAGGAAGACCATGTGGACCTGTCACTGTCTTGTACCCTTGTGCCTCGC TCAGGGGAGCAGGCTGAAGGGTCCCCAGGTGGACCTGGAGACTCTCAGGGTCGAAAACGGCGGCAGACCAGCAT GACAGATTTCTACCACTCCAAACGCCGGCTGATCTTCTCCAAGAGGAAGCCCtaa		
	Primers	Nucleotide number	Primer sequence 5'-3'
	Forward:OP21aF3	18	caccATGTCAGAACCGGC
	Reverse:OP21aR4	17	ttaGGGCTTCCTCTTGG
4b) P21 P38936 (1-160) 160 aa	caccATGTCAGAACCGGCTGGGGATGTCCGTCAGAACCCATGCGGCAGCAAGGCCTGCCGCCGCCTCTTCGGCC CAGTGGACAGCGAGCAGCTGAGCCGCGACTGTGATGCGCTAATGGCGGGCTGCATCCAGGAGGCCCGTGAGCGA TGGAACTTCGACTTTGTCACCGAGACACCACTGGAGGGTGACTTCGCCTGGGAGCGTGTGCGGGGCCTTGGCCT GCCCAAGCTCTACCTTCCCACGGGGCCCCGGCGAGGCCGGGATGAGTTGGGAGGAGGCAGGCGGCCTGGCACCT CACCTGCTCTGCTGCAGGGGACAGCAGAGGAAGACCATGTGGACCTGTCACTGTCTTGTACCCTTGTGCCTCGC TCAGGGGAGCAGGCTGAAGGGTCCCCAGGTGGACCTGGAGACTCTCAGGGTCGAAAACGGCGGCAGACCAGCAT GACAGATTTCTACCACTCCAAACGCCGGCTGATCTTCTCCtga		
	Primers	Nucleotide number	Primer sequence 5'-3'
	Forward:OP21aF3	18	caccATGTCAGAACCGGC
	Reverse:OP21bR4	18	tcaGGAGAAGATCAGCCG
4c) P21 P38936 140 aa	caccGTGGACAGCGAGCAGCTGAGCCGCGACTGTGATGCGCTAATGGCGGGCTGCATCCAGGAGGCCCGTGAGC GATGGAACTTCGACTTTGTCACCGAGACACCACTGGAGGGTGACTTCGCCTGGGAGCGTGTGCGGGGCCTTGGC CTGCCCAAGCTCTACCTTCCCACGGGGCCCCGGCGAGGCCGGGATGAGTTGGGAGGAGGCAGGCGGCCTGGCAC CTCACCTGCTCTGCTGCAGGGGACAGCAGAGGAAGACCATGTGGACCTGTCACTGTCTTGTACCCTTGTGCCTC GCTCAGGGGAGCAGGCTGAAGGGTCCCCAGGTGGACCTGGAGACTCTCAGGGTCGAAAACGGCGGCAGACCAGC ATGACAGATTTCTACCACTCCAAACGCCGGCTGATCTTCTCCAAGAGGAAGCCCtaa		
	Primers	Nucleotide number	Primer sequence 5'-3'
	Forward:OP21cF3	17	caccGTGGACAGCGAGC
	Reverse:OP21bR4	18	tcaGGAGAAGATCAGCCG
5a) RNA-binding protein EWS Q01844 656 aa	**cacc**ATGGCGTCCACGGATTACAGTACCTATAGCCAAGCTGCAGCGCAGCAGGGCTACAGTGCTTACACCGCCC AGCCCACTCAAGGATATGCACAGACCACCCAGGCATATGGGCAACAAAGCTATGGAACCTATGGACAGCCCACT GATGTCAGCTATACCCAGGCTCAGACCACTGCAACCTATGGGCAGACCGCCTATGCAACTTCTTATGGACAGCC TCCCACTGGTTATACTACTCCAACTGCCCCCCAGGCATACAGCCAGCCTGTCCAGGGGTATGGCACTGGTGCTT ATGATACCACCACTGCTACAGTCACCACCACCCAGGCCTCCTATGCAGCTCAGTCTGCATATGGCACTCAGCCT GCTTATCCAGCCTATGGGCAGCAGCCAGCAGCCACTGCACCTACAAGACCGCAGGATGGAAACAAGCCCACTGA GACTAGTCAACCTCAATCTAGCACAGGGGGTTACAACCAGCCCAGCCTAGGATATGGACAGAGTAACTACAGTT ATCCCCAGGTACCTGGGAGCTACCCCATGCAGCCAGTCACTGCACCTCCATCCTACCCTCCTACCAGCTATTCC TCTACACAGCCGACTAGTTATGATCAGAGCAGTTACTCTCAGCAGAACACCTATGGGCAACCGAGCAGCTATGG ACAGCAGAGTAGCTATGGTCAACAAAGCAGCTATGGGCAGCAGCCTCCCACTAGTTACCCACCCCAAACTGGAT CCTACAGCCAAGCTCCAAGTCAATATAGCCAACAGAGCAGCAGCTACGGGCAGCAGAGTTCATTCCGACAGGAC CACCCCAGTAGCATGGGTGTTTATGGGCAGGAGTCTGGAGGATTTTCCGGACCAGGAGAGAACCGGAGCATGAG TGGCCCTGATAACCGGGGCAGGGGAAGAGGGGGATTTGATCGTGGAGGCATGAGCAGAGGTGGGCGGGGAGGAG GACGCGGTGGAATGGGCAGCGCTGGAGAGCGAGGTGGCTTCAATAAGCCTGGTGGACCCATGGATGAAGGACCA GATCTTGATCTAGGCCCTCCTGTAGATCCAGATGAAGACTCTGACAACAGTGCAATTTATGTACAAGGATTAAA TGACAGTGTGACTCTAGATGATCTGGCAGACTTCTTTAAGCAGTGTGGGGTTGTTAAGATGAACAAGAGAACTG GGCAACCCATGATCCACATCTACCTGGACAAGGAAACAGGAAAGCCCAAAGGCGATGCCACAGTGTCCTATGAA GACCCACCCACTGCCAAGGCTGCCGTGGAATGGTTTGATGGGAAAGATTTTCAAGGGAGCAAACTTAAAGTCTC CCTTGCTCGGAAGAAGCCTCCAATGAACAGTATGCGGGGTGGTCTGCCACCCCGTGAGGGCAGAGGCATGCCAC CACCACTCCGTGGAGGTCCAGGAGGCCCAGGAGGTCCTGGGGGACCCATGGGTCGCATGGGAGGCCGTGGAGGA GATAGAGGAGGCTTCCCTCCAAGAGGACCCCGGGGTTCCCGAGGGAACCCCTCTGGAGGAGGAAACGTCCAGCA CCGAGCTGGAGACTGGCAGTGTCCCAATCCGGGTTGTGGAAACCAGAACTTCGCCTGGAGAACAGAGTGCAACC AGTGTAAGGCCCCAAAGCCTGAAGGCTTCCTCCCGCCACCCTTTCCGCCCCCGGGTGGTGATCGTGGCAGAGGT GGCCCTGGTGGCATGCGGGGAGGAAGAGGTGGCCTCATGGATCGTGGTGGTCCCGGTGGAATGTTCAGAGGTGG CCGTGGTGGAGACAGAGGTGGCTTCCGTGGTGGCCGGGGCATGGACCGAGGTGGCTTTGGTGGAGGAAGACGAG GTGGCCCTGGGGGGCCCCCTGGACCTTTGATGGAACAGATGGGAGGAAGAAGAGGAGGACGTGGAGGACCTGGA AAAATGGATAAAGGCGAGCACCGTCAGGAGCGCAGAGATCGGCCCTACtag		
	Primers	Nucleotide number	Primer sequence 5'-3'
	Forward:OEWSaF3	22	caccATGGCGTCCACGGATT AC
	Reverse:OEWSaR4	18	ctaGTAGGGCCGATCTCT
5b) RNA-binding protein EWS Q01844 256-447	CaccCAACAGAGCAGCAGCTACGGGCAGCAGAGTTCATTCCGACAGGACCACCCCAGTAGCATGGGTGTTTATG GGCAGGAGTCTGGAGGATTTTCCGGACCAGGAGAGAACCGGAGCATGAGTGGCCCTGATAACCGGGGCAGGGGA AGAGGGGGATTTGATCGTGGAGGCATGAGCAGAGGTGGGCGGGGAGGAGGACGCGGTGGAATGGGCAGCGCTGG		

191 aa	AGAGCGAGGTGGCTTCAATAAGCC TGGTGGACCCATGGATGAAGGACCAGATCTTGATCTAGGCCCTCCTGTAGATCCAGATGAAGACTCTGACAACA GTGCAATTTATGTACAAGGATTAAATGACAGTGTGACTCTAGATGATCTGGCAGACTTCTTTAAGCAGTGTGGG GTTGTTAAGATGAACAAGAGAACTGGGCAACCCATGATCCACATCTACCTGGACAAGGAAACAGGAAAGCCCAA AGGCGATGCCACAGTGTCCTATGAAGACCCACCCACTGCCAAGGCTGCCGTGGAATGGTTTGATGGGAAAGATT TTCAAGGGAGCAAACTTAAAGTCTCCCTTGCTCGGAAGtga		
	Primers	Nucleotide number	Primer sequence 5'-3'
	Forward:OEWSbF3	17	CaccCAACAGAGCAGCA
	Reverse:OEWSbR4	21	tcaCTTCCGAGCAAGGGAGA C
6a) similar to CAVP-target protein	caccATGGGTCTGCGGGAGGACAGAGCCCCGGCAGCTCCCAGAGCTTCCTTCCAGGACGCCTACCTGATGAAGC TGCGCCGCCAGACGCTCACCTTCACCACCAACCGGCTCAAGGAGTTCCTGGGCGAGCAGCGGCGGCGCCGGGCT GAGGCTGCCACCCGCCACAAGATGTCTAAAGCAGCTCCTGCCAAAAAGCCAGTGGCTGTGGCCCCAGCTCCTGG ATGTACCCTGGACATCAATGACCCACAGGTCCAGAGTGCGGCCATTCGTATCCAGGCCTCTTACCGGGGCCACA GGTCCCGGAAGGAGCTGCGCGAGAAGGGGCCGCCGCGGGTGCTGGAGCCGCTGAAGGACGTGGTGCTGATCGAA GGCAGCGCGGCCAAGCTCACTTGCCGCATTTCGGCTTTCCCGGACCCATTCATCCGCTGGAGTAAGGACGGCAA GGAGCTACGTGACGGTCCCAAGTACCGCTACGTCTTCGAGGACCCTGACGTGGTGGCACTGGTGGTGCGCGACG GCGAGCTGGCAGACCTGGGCCAGTACAGCATCAACGTCACCAACCCCTTCGGCCAGTGCTCCGACTCGGCGCGC ATCCTCGTGGAAGCCGGTGAGGTGATGCCCTGCCTCATCTCCGGGTCCTTCACGGGGTTTCGCTCCCCTTTCCT TCCGCTAGTCCCGACGAAGATTCAAAAGGGACCCGACAACACTAAGGCGCGCAAAGGCACCACCGTGACGCTGA CTGCGGAGATCCTGGGAGAGCCTGCGCCCGACGTAGGCTGGACCAAGGACGGGGAGGACATCGAGGAGGATGAC AGGGTGTTCTTCGAGATCGGCAGCACCACCACGACGCTGACCATTCGCCGGGCCACGCCTCAGGACAGCGGCAA GTACGAGGTGTACGTGGAGAACAGCCTGGGCATGGACCAGAGCTTCGCTCGCGTCGACGTGGCCtga		
	Primers	Nucleotide number	Primer sequence 5'-3'
	Forward:OCAVPF3	17	caccATGGGTCTGCGGG
	Reverse:OCAVPR4	17	tcaGGCCACGTCGACGC
7) SH3 adapter protein SPIN90 Q9NZQ3 FRAGMENT (1-81) 81 aa	caccATGTACCGCGCGCTGTACGCGTTCCGCTCGGCGGAGCCCAACGCGCTGGCGTTCGCCGCGGGCGAGACCTT CCTGGTGCTAGAGCGAAGCAGCGCGCACTGGTGGCTGGCCGCGCGGGCGCGCAGTGGTGAGACGGGCTACGTGC CGCCAGCCTACCTGCGCCGCCTGCAGGGCCTGGAGCAGGATGTCCTCCAGGCCATTGACCGGGCCATCGAGGCT GTACACAACACAGCCATGCGGGATtga		
	Primers	Nucleotide number	Primer sequence 5'-3'
	Forward:OSP90F3	15	caccATGTACCGCGC
	Reverse:OSP90R4	18	tcaATCCCGCATGGCTGT
8) Tuberous sclerosis 2 protein Tuberin TSC2_HUMAN P49815 C-terminal domain Rap-GAP (1531 – 1758) 228 aa	CaccGTGCAGCTCCTCGACCAGATCCCATCATACGACACCCACAAGATCGCCGTCCTGTATGTTGGAGAAGGCC AGAGCAACAGCGAGCTCGCCATCCTGTCCAATGAGCATGGCTCCTACAGGTACACGGAGTTCCTGACGGGCCTG GGCCGGCTCATCGAGCTGAAGGACTGCCAGCCGGACAAGGTGTACCTGGGAGGCCTGGACGTGTGTGGTGAGGA CGGCCAGTTCACCTACTGCTGGCACGATGACATCATGCAAGCCGTCTTCCACATCGCCACCCTGATGCCCACCA AGGACGTGGACAAGCACCGCTGCGACAAGAAGCGCCACCTGGGCAACGACTTTGTGTCCATTGTCTACAATGAC TCCGGTGAGGACTTCAAGCTTGGCACCATCAAGGGCCAGTTCAACTTTGTCCACGTGATCGTCACCCCGCTGGA CTACGAGTGCAACCTGGTGTCCCTGCAGTGCAGGAAAGACATGGAGGGCCTTGTGGACACCAGCGTGGCCAAGA TCGTGTCTGACCGCAACCTGCCCTTCGTGGCCCGCCAGATGGCCCTGCACGCAAATATGGCCTCACAGGTGCAT CATAGCCGCTCCAACCCCACCGATATCTACCCCTCCAAGTGGATTGCCCGGCTCCGCCACATCAAGCGGCTCCG CCAGCGGATCTGCGAGGAAGCCtga		
	Primers	Nucleotide number	Primer sequence 5'-3'
	Forward:OTSCF3	19	CaccGTGCAGCTCCTCGAC
	Reverse:OTSCR4	18	tcaGGCTTCCTCGCAGAT
9) + Calcineurin B Homologous Protein (CHP1) 195 aa Q99653	caccATGGGTTCTCGGGCCTCCACGTTACTGCGGGACGAAGAGCTCGAGGAGATCAAGAAGGAGACCGGCTTTT CCCACAGTCAAATCACTCGCCTCTACAGCCGGTTCACCAGCCTGGACAAAGGAGAGAATGGGACTCTCAGCCGG GAAGATTTCCAGAGGATTCCAGAACTTGCCATCAACCCACTGGGGGACCGGATCATCAATGCCTTCTTTCCAGA GGGAGAGGACCAGGTAAACTTCCGTGGATTCATGCGAACTTTGGCTCATTTCCGCCCCATTGAGGATAATGAAA AGAGCAAAGATGTGAATGGACCCGAACCACTCAACAGCCGAAGCAACAAACTGCACTTTGCTTTTCGACTATAT GATTTGGATAAAGATGAAAAGATCTCCCGTGATGAGCTGTTACAGGTGCTACGCATGATGGTCGGAGTAAATAT CTCAGATGAGCAGCTGGGCAGCATCGCAGACAGGACCATTCAGGAGGCTGATCAGGATGGGGACAGTGCCATAT CTTTCACAGAATTTGTTAAGGTTTTGGAGAAGGTGGATGTAGAACAGAAAATGAGCATCCGATTTCTTCACtaa		
	Primers	Nucleotide number	Primer sequence 5'-3'
	Forward:OCHP1F3	17	caccATGGGTTCTCGGG
	Reverse:OCHP1R4	20	ttaGTGAAGAAATCGGATGC
10) Calcineurin B Homologous Protein 2 (CHP2) O43545 196 aa	caccATGGGGTCGCGCAGCTCCCACGCCGCGGTCATTCCCGACGGGGACAGTATTCGGCGAGAGACCGGCTTCT CCCAAGCCAGCCTGCTCCGCCTGCACCACCGGTTCCGGGCACTGGACAGGAATAAGAAGGGCTACCTGAGCCGC ATGGATCTCCAGCAGATAGGGGCGCTCGCCGTGAACCCCCTGGGAGACCGAATTATAGAAAGCTTCTTCCCCGA TGGGAGCCAGCGAGTGGATTTCCCAGGCTTTGTCAGGGTCTTGGCTCATTTTCGCCCTGTAGAAGATGAGGACA CAGAAACCCAAGACCCCAAGAAACCTGAACCTCTCAACAGCAGAAGGAACAAACTTCACTATGCATTTCAGCTC TATGACCTGGATCGCGATGGGAAGATCTCCAGGCATGAGATGCTGCAGGTTCTCCGTCTGATGGTTGGGGTACA GGTGACAGAAGAGCAGCTGGAGAACATCGCTGACCGCACGGTGCAGGAGGCTGATGAAGATGGGGATGGGGCTG TGTCCTTCGTGGAGTTCACCAAGTCCTTAGAGAAGATGGACGTTGAGCAAAAAATGAGCATCCGGATCCTGAAG tga		
	Primers	Nucleotide number	Primer sequence 5'-3'
	Forward:OCHP2F4	14	caccATGGGGTCGC
	Reverse:OCHP2R4	16	tcaCTTCAGGATCCGG
11) TESCALCIN	caccATGGGCGCTGCCCACTCCGCGTCTGAGGAGGTGCGGGAGCTCGAGGGCAAGACCGGCTTCTCATCGGATC		

Q96BS2 214 aa	AGATCGAGCAGCTCCATCGGAGATTTAAGCAGCTGAGTGGAGATCAGCCTACCATTCGCAAGGAGAACTTCAAC AATGTCCCGGACCTGGAGCTCAACCCCATCCGATCCAAAATTGTTCGTGCCTTCTTCGACAACAGGAACCTGCG CAAGGGACCCAGTGGCCTGGCTGATGAGATCAATTTCGAGGACTTCCTGACCATCATGTCCTACTTCCGGCCCA TCGACACCACCATGGACGAGGAACAGGTGGAGCTGTCCCGGAAGGAGAAGCTGAGATTTCTGTTCCACATGTAC GACTCGGACAGCGACGGCCGCATCACTCTGGAAGAATATCGAAATGTGGTCGAGGAGCTGCTGTCGGGAAACCC TCACATCGAGAAGGAGTCCGCTCGCTCCATCGCCGACGGGGCCATGATGGAGGCGGCCAGCGTGTGCATGGGGC AGATGGAGCCTGATCAGGTGTACGAGGGGATCACCTTCGAGGACTTCCTGAAGATCTGGCAGGGGATCGACATT GAGACCAAGATGCACGTCCGCTTCCTTAACATGGAAACCATGGCCCTCTGCCACtga		
	Primers	Nucleotide number	Primer sequence 5'-3'
	Forward:OTESCF3	14	caccATGGGCGCTG
	Reverse:OTESCR4	15	tcaGTGGCAGAGGGC
12) visinin-like protein 1 191aa p28677	caccATGGGGAAGCAGAATAGCAAACTGGCCCCTGAAGTGATGGAGGACCTGGTGAAGAGCACAGAGTTTAATG AGCATGAACTCAAGCAGTGGTACAAAGGATTTCTCAAGGACTGTCCAAGTGGGAGGCTAAATCTCGAGGAATTT CAGCAGCTCTATGTGAAGTTCTTTCCTTATGGAGACGCCTCCAAGTTTGCCCAGCATGCCTTCCGAACCTTCGA CAAGAATGGGGACGGCACCATTGACTTCCGAGAGTTCATCTGCGCTCTGTCCATCACCTCCAGGGGCAGCTTTG AGCAGAAGCTGAACTGGGCCTTCAATATGTATGACCTGGATGGTGATGGCAAGATCACCCGAGTGGAGATGCTG GAGATCATCGAGGCTATCTACAAAATGGTAGGCACTGTGATCATGATGAAAATGAATGAGGATGGCCTGACGCC TGAGCAGCGAGTAGACAAGATTTTCAGCAAGATGGATAAGAACAAAGATGACCAGATTACACTGGATGAATTCA AAGAAGCTGCAAAGAGCGACCCTTCCATTGTATTACTTCTGCAGTGCGACATCCAGAAAtga		
	Primers	Nucleotide number	Primer sequence 5'-3'
	Forward:OVIS1F3	21	caccATGGGGAAGCAGAATAG
	Reverse:OVIS1R4	20	tcaTTTCTGGATGTCGCACT
13a) + visinin-like protein 3 (HIPPOCALCIN-LIKE PROTEIN 1)P37235 193aa	caccATGGGCAAACAGAACAGCAAGCTGCGGCCCGAGGTGCTGCAGGACCTGCGGGAGAACACGGAGTTCACCG ACCACGAGCTGCAGGAGTGGTACAAGGGCTTCCTCAAGGACTGCCCCACCGGCCACCTGACCGTGGACGAGTTC AAGAAGATCTACGCCAACTTCTTCCCCTACGGCGACGCTTCCAAGTTCGCCGAGCACGTCTTCCGCACCTTCGA CACCAACGGCGACGGCACCATCGACTTCCGGGAGTTCATCATTGCGCTGAGCGTGACCTCGCGGGGCAAGCTGG AGCAGAAGCTCAAGTGGGCCTTCAGCATGTACGACCTGGACGGCAACGGCTACATCAGCCGCAGCGAGATGCTG GAGATCGTGCAGGCCATCTACAAGATGGTGTCGTCTGTGATGAAGATGCCGGAGGATGAGTCCACCCCGGAGAA GCGCACAGACAAGATCTTCAGGCAGATGGACACCAACAATGACGGCAAACTGTCCTTGGAAGAATTCATCAGAG GTGCCAAGAGCGACCCCTCCATCGTCCGCCTGCTGCAGTGCGACCCCAGCAGTGCCAGTCAGTTCtga		
	Primers	Nucleotide number	Primer sequence 5'-3'
	Forward:OVIS3F3	20	caccATGGGCAAACAGAACA
	Reverse:OVIS3R4	19	tcaGAACTGACTGGCACTG
14) CALCYPHOSINE 189 AA Q13938	caccATGGACGCCGTGGATGCCACCATGGAGAAACTCCGGGCACAGTGCCTGTCCCGCGGGGCCTCGGGCATCC AGGGCCTGGCCAGGTTTTTCCGCCAACTAGACCGGGACGGGAGCAGATCCCTGGACGCTGATGAGTTCCGGCAG GGTCTGGCCAAACTCGGGCTGGTGCTGGACCAGGCGGAGGCAGAGGGTGTGTGCAGGAAGTGGGACCGCAATGG CAGCGGGACGCTGGATCTGGAGGAGTTCCTTCGGGCGCTGCGGCCCCCCATGTCCCAGGCCCGGGAGGCTGTCA TCGCAGCTGCATTTGCCAAGCTGGACCGCAGTGGGGACGGCGTCGTGACGGTGGACGACCTCCGCGGGGTGTAC AGTGGCCGTGCCCACCCCAAGGTGCGCAGTGGGGAGTGGACCGAGGACGAGGTGCTGCGCCGCTTCCTGGACAA CTTCGACTCCTCTGAGAAGGACGGGCAGGTCACACTGGCGGAATTCCAGGACTACTACAGCGGCGTGAGTGCCT CCATGAACACGGATGAGGAGTTCGTGGCCATGATGACCAGTGCCTGGCAGCTGtga		
	Primers	Nucleotide number	Primer sequence 5'-3'
	Forward:OCAPSF3	16	caccATGGACGCCGTG
	Reverse:OCAPSR4	17	tcaCAGCTGCCAGGCAC
16a) CALNEURON Q9BXU9 219 AA	caccATGCCGTTCCACCATGTGACCGCCGGCTTGTTGTACAAGGGGAATTACCTCAACCGATCGCTCTCTGCTG GCAGTGACAGCGAACAGCTGGCTAATATCTCCGTGGAGGAGCTCGATGAAATCCGAGAGGCCTTTCGGGTTCTG GACCGGGATGGGAACGGCTTCATCTCCAAGCAGGAGCTGGGCATGGCCATGCGCTCTTTGGGGTACATGCCAAG CGAGGTGGAGCTGGCCATCATCATGCAGCGCTTGGACATGGACGGGGATGGCCAGGTGGATTTTGATGAATTCA TGACCATTCTTGGCCCCAAACTGGTGTCTTCAGAAGGTCGCGATGGTTTTCTTGGGAACACGATAGACAGCATA TTCTGGCAGTTTGACATGCAAAGGATAACTCTGGAAGAGTTGAAGCACATTCTCTATCATGCCTTCCGAGACCA CCTAACGATGAAGGACATTGAGAACATCATTATCAATGAGGAAGAGAGCCTGAATGAGACCTCGGGGAACTGCC AAACAGAGTTTGAAGGAGTGCATTCCCAGAAGCAGAACAGACAGACCTGCGTCCGGAAGAGCCTCATATGCGCC TTTGCTATGGCCTTCATCATCAGTGTCATGCTGATTGCAGCCAACCAGATACTCCGGAGCGGCATGGAGtag		
	Primers	Nucleotide number	Primer sequence 5'-3'
	Forward:OCALNF3	18	caccATGCCGTTCCACCA
	Reverse:OCALNR4	18	ctaCTCCATGCCGCTCCG
16b) CALNEURON Q9BXU9 DOMINIO N-TERMINALE 33-104 72 aa	caccATCTCCGTGGAGGAGCTCGATGAAATCCGAGAGGCCTTTCGGGTTCTGGACCGGGATGGGAACGGCTTCA TCTCCAAGCAGGAGCTGGGCATGGCCATGCGCTCTTTGGGGTACATGCCAAGCGAGGTGGAGCTGGCCATCATC ATGCAGCGCTTGGACATGGACGGGGATGGCCAGGTGGATTTTGATGAATTCATGACCATTCTTGGCCCCAAAtg a		
	Primers	Nucleotide number	Primer sequence 5'-3'
	Forward:OCALNF5	18	caccATCTCCGTGGAGGA
	Reverse:OCALNR6	17	tcaTTTGGGGCCAAGAA
17) + GUANYLYL CYCLASE ACTIVATING PROTEIN 1 P43080 201 AA	caccATGGGCAACGTGATGGAGGGAAAGTCAGTGGAGGAGCTGAGCAGCACCGAGTGCCACCAGTGGTACAAGA AGTTCATGACTGAGTGCCCCTCTGGCCAACTCACCCTCTATGAGTTCCGCCAGTTCTTCGGCCTCAAGAACCTG AGCCCGTCGGCCAGCCAGTACGTGGAACAGATGTTTGAGACTTTTGACTTCAACAAGGACGGCTACATTGATTT CATGGAGTACGTGGCAGCGCTCAGCTTGGTCCTCAAGGGGAAGGTGGAACAGAAGCTCCGCTGGTACTTCAAGC TCTATGATGTAGATGGCAACGGCTGCATTGACCGCGATGAGCTGCTCACCATCATCCAGGCCATTCGCGCCATT AACCCCTGCAGCGATACCACCATGACTGCAGAGGAGTTCACCGATACAGTGTTCTCCAAGATTGACGTCAACGG		

	GGATGGGGAACTCTCCCTGGAAGAGTTTATAGAGGGCGTCCAGAAGGACCAGATGCTCCTGGACACACTGACAC GAAGCCTGGACCTTACCCGCATCGTGCGCAGGCTCCAGAATGGCGAGCAAGACGAGGAGGGGGCTGACGAGGCC GCTGAGGCAGCCGGCtga		
	Primers	Nucleotide number	Primer sequence 5'-3'
	Forward:OGCA1F3	18	caccATGGGCAACGTGAT
	Reverse:OGCA1R4	15	tcaGCCGGCTGCCTC
18) + GUANYLYL CYCLASE ACTIVATING PROTEIN 2 Q9UMX6 200 AA	CACCATGGGGCAGGAGTTTAGCTGGGAGGAGGCGGAGGCAGCTGGCGAGATAGATGTGGCGGAGCTCCAGGAGT GGTACAAGAAGTTTGTGATGGAGTGCCCCAGCGGCACACTCTTTATGCATGAGTTTAAGCGCTTCTTCAAGGTC ACAGACGATGAGGAGGCCTCCCAGTATGTAGAGGGCATGTTCCGAGCCTTCGACAAGAATGGGGACAACACCAT CGACTTCCTGGAGTACGTGGCAGCTCTGAATCTCGTGCTGAGGGGCACCCTGGAGCACAAGCTGAAGTGGACAT TCAAGATCTATGATAAGGATGGCAATGGCTGCATCGACCGCCTGGAGCTACTCAACATTGTGGAGGGAATTTAC CAGCTGAAGAAAGCCTGCCGGCGAGAGCTACAGACTGAGCAAGACCAGCTGCTCACACCCGAGGAGGTCGTGGA CAGGATCTTCCTCCTGGTGGATGAGAATGGAGATGGCCAGCTGTCTCTGAACGAGTTTGTTGAAGGTGCCCGTC GGGACAAGTGGGTGATGAAGATGCTGCAGATGGACATGAATCCCAGCAGCTGGCTCGCTCAGCAGAGACGGAAA AGTGCCATGTTCTGA		
	Primers	Nucleotide number	Primer sequence 5'-3'
	Forward:OGCAP2F1	21	CACCATGGGGCAGGAGTTTA G
	Reverse:OGCAP2R2	20	TCAGAACATGGCACTTTTCC
19a) + Guanylyl cyclase activating protein 3 O95843 iso1 209	caccATGGGGAATGGCAAATCTATAGCTGGTGATCAGAAAGCAGTTCCTACACAAGAGACCCATGTGTGGTACA GAACATTTATGATGGAATATCCATCCGGCCTGCAAACACTACATGAATTTAAGACACTTTTGGGTCTGCAAGGT CTGAATCAGAAGGCCAATAAACATATTGATCAAGTTTATAATACCTTTGACACGAACAAGGATGGATTTATTGA CTTTTTGGAGTTTATTGCTGCTGTAAATCTAATCATGCAAGAAAAAATGGAGCAAAAATTAAAATGGTATTTTA AGCTGTATGATGCTGATGGAAATGGTTCTATTGACAAAAATGAACTACTGGACATGTTCATGGCGGTACAAGCC CTCAATGGCCAGCAAACTCTGAGTCCTGAAGAATTCATCAACTTGGTGTTCCATAAGATCGATATAAACAATGA TGGGGAATTGACTTTAGAAGAATTTATCAATGGCATGGCAAAAGATCAGGATCTCCTGGAGATTGTTTACAAGA GCTTCGACTTCTCCAATGTGCTGAGAGTAATCTGTAATGGGAAGCAGCCAGACATGGAGACAGACTCCTCCAAA TCTCCTGACAAGGCTGGTCTAGGGAAGGTGAAAATGAAGtag		
	Primers	Nucleotide number	Primer sequence 5'-3'
	Forward:OGCAP3F3	18	caccATGGGGAATGGCAA
	Reverse:OGCAP3R4	27	ctaCTTCATTTTCACCTTCC CTAGACC
20) S102 P29034 97 aa	caccATGTGCAGTTCTCTGGAGCAGGCGCTGGCTGTGCTGGTCACTACCTTCCACAAGTACTCCTGCCAAGAGG GCGACAAGTTCAAGCTGAGTAAGGGGGAAATGAAGGAACTTCTGCACAAGGAGCTGCCCAGCTTTGTGGGGGAG AAAGTGGATGAGGAGGGGCTGAAGAAGCTGATGGGCAGCCTGGATGAGAACAGTGACCAGCAGGTGGACTTCCA GGAGTATGCTGTTTTCCTGGCACTCATCACTGTCATGTGCAATGACTTCTTCCAGGGCTGCCCAGACCGACCCt ga		
	Primers	Nucleotide number	Primer sequence 5'-3'
	Forward:OS102F3	21	caccATGTGCAGTTCTCTGG A
	Reverse:OS102R4	16	tcaGGGTCGGTCTGGG
21) S116 Q96FQ6 103AA	caccATGTCAGACTGCTACACGGAGCTGGAGAAGGCAGTCATTGTCCTGGTGGAAAACTTCTACAAATATGTGTC TAAGTACAGCCTGGTCAAGAACAAGATCAGCAAGAGCAGCTTCCGCGAGATGCTCCAGAAAGAGCTGAA CCACATGCTGTCGGACACAGGGAACCGGAAGGCTGCGGATAAGCTCATCCAGAACCTGGATGCCAATCAT GATGGGCGCATCAGCTTCGATGAGTACTGGACCTTGATAGGCGGCATCACCGGCCCCATCGCCAAACTCA TCCATGAGCAGGAGCAGCAGAGCAGCAGCtag		
	Primers	Nucleotide number	Primer sequence 5'-3'
	Forward:OS116F3	22	caccATGTCAGACTGCTACA CG
	Reverse:OS116R4	18	ctaGCTGCTGCTCTGCTG
22a) PEFLIN Q9UBV8 284 AA INTERA	caccATGGCCAGCTATCCTTACCGGCAGGGCTGCCCAGGAGCTGCAGGACAAGCACCAGGAGCCCCTCCGGGTA GCTACTACCCTGGACCCCCCAATAGTGGAGGGCAGTATGGTAGTGGGCTACCCCCTGGTGGTGGTTATGGGGGT CCTGCCCCTGGAGGGCCTTATGGACCACCAGCTGGTGGAGGGCCCTATGGACACCCCAATCCTGGGATGTTCCC CTCTGGAACTCCAGGAGGACCATATGGCGGTGCAGCTCCCGGGGGCCCCTATGGTCAGCCACCTCCAAGTTCCT ACGGTGCCCAGCAGCCTGGGCTTTATGGACAGGGTGGCGCCCCTCCCAATGTGGATCCTGAGGCCTACTCCTGG TTCCAGTCGGTGGACTCAGATCACAGTGGCTATATCTCCATGAAGGAGCTAAAGCAGGCCCTGGTCAACTGCAA TTGGTCTTCATTCAATGATGAGACCTGCCTCATGATGATAAACATGTTTGACAAGACCAAGTCAGGCCGCATCG ATGTCTACGGCTTCTCAGCCCTGTGGAAATTCATCCAGCAGTGGAAGAACCTCTTCCAGCAGTATGACCGGGAC CGCTCGGGCTCCATTAGCTACACAGAGCTGCAGCAAGCTCTGTCCCAAATGGGCTACAACCTGAGCCCCCAGTT CACCCAGCTTCTGGTCTCCCGCTACTGCCCACGCTCTGCCAATCCTGCCATGCAGCTTGACCGCTTCATCCAGG TGTGCACCCAGCTGCAGGTGCTGACAGAGGCCTTCCGGGAGAAGGACACAGCTGTACAAGGCAACATCCGGCTC AGCTTCGAGGACTTCGTCACCATGACAGCTTCTCGGATGCTAtga		
	Primers	Nucleotide number	Primer sequence 5'-3'
	Forward:OPEFF3	22	caccATGGCCAGCTATCCTT AC
	Reverse:OPEFR4	24	tcaTAGCATCCGAGAAGCTG TCAT
22b) PEFLIN Q9UBV8 ESP1 (EFH1 2) (116-	CaccGATCCTGAGGCCTACTCCTGGTTCCAGTCGGTGGACTCAGATCACAGTGGCTATATCTCCATGAAGGAGC TAAAGCAGGCCCTGGTCAACTGCAATTGGTCTTCATTCAATGATGAGACCTGCCTCATGATGATAAACATGTTT		

182) 67 aa	GACAAGACCAAGTCAGGCCGCATCGATGTCTACGGCTTCTCAGCCCTGTGGAAATTCtga		
	Primers	Nucleotide number	Primer sequence 5'-3'
	Forward:OPEFEspF3	20	CaccGATCCTGAGGCCTACT
	Reverse:OPEFEspR4	19	tcaGAATTTCCACAGGGCT
22c) >PEFLIN ESP2 (EFH3_4) (183 – 250) 68 aa	CaccATCCAGCAGTGGAAGAACCTCTTCCAGCAGTATGACCGGGACCGCTCGGGCTCCATTAGCTACACAGAGC TGCAGCAAGCTCTGTCCCAAATGGGCTACAACCTGAGCCCCCAGTTCACCCAGCTTCTGGTCTCCCGCTACTGC CCACGCTCTG CCAATCCTGCCATGCAGCTTGACCGCTTCATCCAGGTGTGCACCCAGCTGtga		
	Primers	Nucleotide number	Primer sequence 5'-3'
	Forward:OPEFEspF5	21	CaccATCCAGCAGTGGAAGAA
	Reverse:OPEFEspR6	17	tcaCAGCTGGGTGCACA
23a) ALG-2 ESP1 (EFH1_2) 67 aa	caccCAGAGCTTCCTGTGGAACGTTTTCCAGAGGGTCGATAAAGACAGGAGTGGAGTGATATCAGACACCGAGC TTCAGCAAGCTCTCTCCAACGGCACGTGGACTCCCTTTAATCCAGTGACTGTCAGGTCGATCATATCCATGTTT GACCGTGAGAACAAGGCCGGCGTGAACTTCAGCGAGTTCACGGGTGTGTGGAAGTACtga		
	Primers	Nucleotide number	Primer sequence 5'-3'
	Forward:OALG2Esp1F1	18	caccCAGAGCTTCCTGTG
	Reverse:OALG2Esp1R2	21	tcaGTACTTCCACACACCCGT
23b) ALG-2 ESP2 (EFH3_4) 66 aa	caccATCACGGACTGGCAGAACGTCTTCCGCACGTACGACCGGGACAACTCCGGGATGATCGATAAGAACGAGC TGAAGCAGGCCCTCTCAGGTTTCGGCTACCGGCTCTCTGACCAGTTCCACGACATCCTCATTCGAAAGTTTGAC AGGCAGGGACGGGGGCAGATTGCCTTCGACGACTTCATCCAGGGCTGCATCGTCtga		
	Primers	Nucleotide number	Primer sequence 5'-3'
	Forward:OALG2Esp2F1	17	caccATCACGGACTGGC
	Reverse:OALG2Esp2R2	17	tcaGACGATGCAGCCCT
24) EP15_HUMAN P42566 896AA ESP1 (EH1) (15-85) 71 aa	caccGGGAATCCTGTATATGAAAAATACTATAGACAGGTTGATACAGGCAATACTGGAAGGGTGTTGGCTTCTG ATGCTGCTGCTTTCCTGAAAAAATCAGGGCTTCCAGACTTGATACTTGGAAAGATTTGGGATTTAGCCGACACA GATGGCAAAGGTATCCTGAACAAACAAGAATTCTTTGTTGCTTTGCGTCTTGTGGCATGTGCCCAGAATtga		
	Primers	Nucleotide number	Primer sequence 5'-3'
	Forward:OEP15Esp1F3	22	caccGGGAATCCTGTATATGAA
	O Reverse:EP15Esp1R4	18	tcaATTCTGGGCACATGC
25) BIR3-NAIP (238-362) Q13075 125 aa	CaccCAGTATATTCAAAGCTACAAGGGATTTGTTGACATAACGGGAGAACATTTTGTGAATTC CTGGGTCCAGAGAGAATTACCTATGGCATCAGCTTATTGCAATGACAGCATCTTTGCTTACGAAGAACTA CGGCTGGACTCTTTTAAGGACTGGCCCCGGGAATCAGCTGTGGGAGTTGCAGCACTGGCCAAAGCAGGTCTTTT CTACACAGGTATAAAGGACATCGTCCAGTGCTTTTCCTGTGGAGGGTGTTTAGAGAAATGGCAGGAAGGTGATG ACCCATTAGACGATCACACCAGATGTTTTCCCAATTGTCCATTTCTCCAAAATATGAAGTCC TCTGCGGAAGTGACTCCAGACCTTCAGAGCCGTGGTtga		
	Primers	Nucleotide number	Primer sequence 5'-3'
	Forward:OBIR3IIF3	25	CaccCAGTATATTCAAAGCTACAAG
	Reverse:OBIR3IIR4	19	tcaACCACGGCTCTGAAGG
26) S100A10 97 aa	CaccATGCCATCTCAAATGGAACACGCCATGGAAACCATGATGTTTACATTTCACAAATTCGCTGGGGATAAAG GCTACTTAACAAAGGAGGACCTGAGAGTACTCATGGAAAAGGAGTTCCCTGGATTTTTGGAAAATCAAAAAGAC CCTCTGGCTGTGGACAAAATAATGAAGGACCTGGACCAGTGTAGAGATGGCAAAGTGGGCTTCCAGAGCTTCTT TTCCCTAATTGCGGGCCTCACCATTGCATGCAATGACTATTTTGTAGTACACATGAAGCAGAAGGGAAAGAAGt ag		
	Primers	Nucleotide number	Primer sequence 5'-3'
	Forward:OS100A10F3	19	CaccATGCCATCTCAAATG
	Reverse:OS100A10R4	24	ctaCTTCTTTCCCTTCTGCTTCAT
27)CLSP 146 aa	caccATGGCCGGTGAGCTGACTCCTGAGGAGGAGGCCCAGTACAAAAAGGCTTTCTCCGCGGTTGACACGGATG GAAACGGCACCATCAATGCCCAGGAGCTGGGCGCGGCGCTGAAGGCCACGGGCAAGAACCTCTCGGAGGCCCAG CTAAGGAAACTCATCTCCGAGGTTGACAGCGACGGCGACGGCGAAATCAGCTTCCAGGAGTTCCTGACGGCGGC AAGGAAGGCCAGGGCCGGCCTGGAGGACCTGCAGGTCGCCTTCCGCGCCTTCGACCAGGATGGCGACGGCCACA TCACCGTGGACGAGCTCAGGCGGGCCATGGCGGGGCTGGGGCAGCCGCTGCCGCAGGAGGAGCTGGACGCCATG ATCCGCGAGGCCGACGTGGACCAGGACGGGCGGGTGAACTACGAGGAGTTCGCGAGGATGCTCGCCCAGGAGtg		

<table>
<tr><td></td><td colspan="3">a</td></tr>
<tr><td></td><td>Primers</td><td>Nucleotide number</td><td>Primer sequence 5'-3'</td></tr>
<tr><td></td><td>Forward:OCLSPF1</td><td>16</td><td>caccATGGCCGGTGAG</td></tr>
<tr><td></td><td>Reverse:OCLSPR2</td><td>16</td><td>tcaCTCCTGGGCGAGC</td></tr>
<tr><td>28) Hippocalcin 193 aa</td><td colspan="3">caccATGGGCAAGCAGAACAGCAAGCTGCGGCCCGAGATGTTGCAGGACCTGCGAGAGAACACAGAGTTCTCAG
AGCTGGAGCTGCAGGAGTGGTACAAGGGCTTCCTCAAGGACTGCCCCACAGGAATCCTCAATGTGGATGAGTTC
AAGAAGATCTACGCCAACTTCTTTCCCTATGGTGACGCCTCCAAGTTTGCCGAGCACGTCTTCCGCACCTTTGA
CACCAACAGCGATGGCACCATAGACTTTCGGGAGTTCATCATTGCGCTGAGCGTGACCTCGCGCGGCCGCCTGG
AGCAGAAGCTCATGTGGGCCTTCAGCATGTATGACCTGGACGGCAACGGCTACATCAGCCGGGAGGAGATGCTG
GAGATCGTGCAGGCCATTTACAAGATGGTTTCGTCCGTGATGAAGATGCCGGAGGACGAGTCGACCCCGGAAAA
GAGGACTGAGAAAATCTTCCGCCAAATGGACACAAACAACGACGGCAAGCTGTCCTTGGAGGAGTTCATCCGCG
GGGCCAAAAGCGACCCGTCCATCGTGCGTCTGCTGCAGTGCGACCCCAGCAGCGCCTCCCAGTTCtga</td></tr>
<tr><td></td><td>Primers</td><td>Nucleotide number</td><td>Primer sequence 5'-3'</td></tr>
<tr><td></td><td>Forward:OHPCAF1</td><td>19</td><td>caccATGGGCAAGCAGAAC</td></tr>
<tr><td></td><td>Reverse:OHPCAR2</td><td>18</td><td>tcaGAACTGGGAGGCGCT</td></tr>
<tr><td>29>BIR3-NAIP frammento piccolo 68 aa mw 7610.6 (278-344)</td><td colspan="3">caccGAAGAACTACGGCTGGACTCTTTTAAGGACTGGCCCCGGGAATCAGCTGTGGGAGTTGCAGCACTGGCCA
AAGCAGGTCTTTTCTACACAGGTATAAAGGACATCGTCCAGTGCTTTTCCTGTGGAGGGTGTTTAGAGAAATGG
CAGGAAGGTGATGACCCATTAGACGATCACACCAGATGTTTTCCCAATTGTCCATTTCTCTgga</td></tr>
<tr><td></td><td>Primers</td><td>Nucleotide number</td><td>Primer sequence 5'-3'</td></tr>
<tr><td></td><td>Forward:OBIR3IIF5</td><td>21</td><td>caccGAAGAACTACGGCTGG A</td></tr>
<tr><td></td><td>Reverse:OBIR3IIR6</td><td>22</td><td>tcaGAGAAATGGACAATTGG GA</td></tr>
</table>

REFERENCES

Arnau J, Lauritzen C, Pedersen J, Petersen GE. Current strategies for the use of affnity tags and tag removal for the purification of recombinant proteins. Protein Expression and Purification 2006;48:1–13.

Aalberse RC, Ree RV. Crossreactive carbohydrate determinants. Clin Rev Allergy Immunol 1997;15:375-87.

Aalberse RC. Molecular mechanisms in allergy and clinical immunology. J Allergy Clin Immunol 2000;106:228-38.

Affinity Chromatography Principles and Methods (Amersham Pharmacia Biotech AB 2001).

Anfinsen CB. Principles that govern the folding of protein chains. Science 1973;181:223-230.

Avon E, De Benedictis L, Maier JAM, Mariotti M. Interaction between Endothelial Differentiation-related Factor-1 and Calmodulin *in Vitro* and *in Vivo.* The Journal of Biological Chemistry 2000;275:(31)24047–24051.

Babini E, Bertini I, Capozzi F, Chirivino E, Luchinat C. A structural and dynamic characterization of the EF-hand protein CLSP. Structure. 2006 Jun;14(6):1029-38.

Babini E, Bertini I, Capozzi F, Luchinat C, Quattrone A, Turano M. Principal component analysis of the conformational freedom within the EF-hand superfamily. J Proteome Res. 2005 Nov-Dec;4(6):1961-71.

Bentley GA, Bhat TN, Boulot G, Fischmann TO, Mariuzza RA, Phillips SE, et al. Crystallographic refinement of the three-dimensional structure of the FabD1.3-lysozyme complex at 2.5-A resolution. J Biol Chem 1991;266:12915-20

Bernstein IL, Chapman JA, Lee RE, Oppenheimer J. Food allergy: a practice parameter. Annals of allergy, asthma & immunology 2006;96:S1-S68.

Bettelheim P, Duchene M, Pettenburger K, Sillaber C, Valent P, Valenta R, et al. Identification of profilin as a novel pollen allergen; IgE autoreactivity in sensitized individuals. Science 1991;253:557-60.

Bolognese F, Pitarque-Martì M, Lo Cicero V, Mantovani R,.Maier JA.Characterization of the human EDF-1 minimal promoter: Involvement of NFY and Sp1 in the regulation of basal transcription. Gene 2006; 374:87–95.

Branden C, Tooze J. Introduxtion to protein structure. 2nd ed. New York:Garland Publishing; 1999.

Breiteneder H, Mills ENC. Molecular properties of food allergens. J Allergy Clin Immunol 2005;115:14-23.

Breiteneder H, Radauer C. A classification of plant food allergens. J Allergy Clin Immunol 2004;113:821– 30.

Bruijnzeel-Koomen C, Ortolani C, Aas K, Bindslev-Jensen C, Bjorksten B, Moneret-Vautrin D, et al. Adverse reactions to food European academy of allergology and clinical immunology subcommittee. Allergy 1995;50:623– 35.

Bugajska-Schretter A, Grote M, Vangelista L, Valent P, Sperr WR, Rumpold H, et al. Purification, biochemical, and immunological characterisation of a major food allergen: different immunoglobulin E recognition of the apo- and calcium-bound forms of carp parvalbumin. Gut 2000;46:661– 9.

Bush RK, Hefle SL. Food allergens. Crit Rev Food Sci Nutr 1996;36(Suppl):S119– 63.

Capozzi F, Casadei F, Luchinat C. EF-hand protein dynamics and evolution of calcium signal transduction: an NMR view.J Biol Inorg Chem. 2006 Nov;11(8):949-62. Review.

Chakraborty C, Nandi S, Jana S. Prion disease: a deadly disease for protein misfolding. Curr. Pharm. Biotechnol 2005; 6:167-177.

Chazin WJ, Nelson MR. Structures of EF-hand Ca2+-binding proteins: Diversity in the organization, packing and response to Ca2+ Binding. BioMetals 1998;11:297-318.

Concepcion GP, David MP, Padlan EA. Why don't humans get scrapie from eating sheep? A possible explanation based on secondary structure predictions. Med. Hypotheses 2005;64: 919-924.

Cox JA, Durussel I, Heizmann CW, Sturchler E, Weibel M. S100A16, a Novel Calcium-binding Protein of the EF-hand Superfamily. The Journal of Biological Chemistry 2006;281(50):38905–38917.

Dissertori O, Ferreira F, Lackner P, Wopfner N. Calcium-binding proteins and their role in allergic diseases. Immunol Allergy Clin North Am. 2007 Feb;27(1):29-44.

EuroPREVALL - THE PREVALENCE, COST AND BASIS OF FOOD ALLERGY ACROSS EUROPE. Integrated project. Annex1-"Description of work". March 2005.

Gateway Technology, version E 2003. Catalog. nos.12535-019 and 12535-027 Invitrogen.

Gel filtration Principles and Methods (Amersham Pharmacia Biotech AB 1998).

Hebenstreit D., Ferreira F. Structural changes in calcium-binding allergens: use of circular dichroism to study binding characteristics. Allergy 2005;60:1208–1211.

Heizmann CW. The importance of calcium-binding proteins in childhood diseases. The Journal of Pediatrics 2005;147:731-8.

Helm RM, Burks AW. Mechanisms of food allergy. Curr Opin Immunol 2000;12:647–53.

Ion Exchange Chromatography Principles and Methods(Amersham Pharmacia Biotech AB 2001).

Lehrer SB, Ayuso R, Reese G. Seafood allergy and allergens: a review. Mar Biotechnol 2003;5:339 – 48.

Lehninger LA, Nelson LD, Cox MM. Principi di biochimica. Zanichelli Second Edition 1994;170-172.

Mills ENC, Jenkins JA, Alcocer MJC, Shewry PR. Structural and biological relationships of plant food allergens. Crit Rev Nutr Food Sci 2004;44:379– 407.

Mills ENC, Valovirta E, Madsen C, Taylor SL, Vieths S, Anklam E, et al. Information provision for allergic consumers: where are we going with food allergen labelling? Allergy 2004;59:1262– 8.

Padlan EA. X-ray crystallography of antibodies. Adv Protein Chem 1996;49:57-133.

Pascal B, LaBaer J. High throughput protein production for functional proteomics. Trends in Biotechnology 2003;21(9):383-8.

Pizzin G, Bentley S, Maggi E. Allergia alimentare per consumo di prodotti di origine animale: stato dell'arte (2003) Ann. Fac. Medic. Vet. di Parma, Vol. XXIII: 261-269.

Poulsen LK, Hansen TK, Norgaard A, Vestergaard H, Stahl Skov P, Bindslev-Jensen C. Allergens from fish and egg. Allergy 2001;56(Suppl 67):39- 42.

Roux KH, Sathe SK, Teuber SS. Effects of food processing on the stability of food allergens. Biotechnology Advances 2005;23 423-429.

Schirmer RH, Schulz GE. Principles of Protein Structure. In 'Springer Advanced Texts in Chemistry' 1990 (Ed. C. R. Cantor).

Svensson M, Hakansson A, Mossberg AK, Linse S, Svanborg C. Conversion of alpha-lactalbumin to a protein inducing apoptosis. Proc. Natl. Acad. Sci. U S A 2000;97:4221-4226.

The Recombinant Protein Handbook, Protein Amplification and Simple Purification (Amersham pharmacia biotech AB 2000).

Valenta R, Twardosz A, Swoboda I, Hakey B, Spitzauer S, Kraft D. Calcium-binding Proteins in Type I Allergy - Elicitors and Vaccines from Calcium - The molecular basis of calcium, action in biology and medicine. Ed Roland Pochet 2000;365-377.

Wal JM. Cow's milk proteins/allergens. Ann Allergy Asthma Immunol 2002;89(Suppl 1):3-10.

Wetlaufer DB. Nucleation, rapid folding, and globular intrachain regions in proteins. Proc. Natl. Acad. Sci 1973;70:697-701.