Università degli Studi di Firenze

DOTTORATO DI RICERCA IN
"Scienza del Suolo e Climatologia"

CICLO XXV

COORDINATORE Prof. Luca Calamai

Interactions between proteins and soil humic substances: evidences from electrophoeresis, mass spectrometry and nuclear magnetic resonance

Settore Scientifico Disciplinare AGR/13

PhD Aspirant
Mariarita Arenella

Tutor
Dr *Giancarlo Renella*
University of Florence

Co- Tutor
Dr *Grazia Masciandaro*
ISE-CNR, Pisa

2010-2013

Table of contents

Chapter 2: Materials and Methods

Chapter 1: Introduction

1.1 The soil

Soils originate from the weathering of rocks and minerals under the effects of the physical, chemical and biological factors including climate, and macro- and microorganisms, over a period of time, well resumed by the famous pedogenesis equation of Jenny (1941):

$$Soil = f(Cl, O, R, P, T)$$

in which the soil (S) is the result of the wheathering parent materials (P) under the actions of main environmental factors such as climate (Cl), relief (R), living organisms (O), and time (T).

The typical soil composition in terms of solid, liquid and gas phases phases is shown in Figure 1.1.

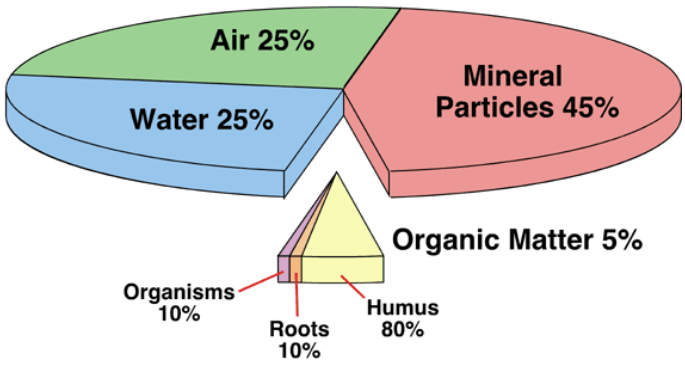


Figure 1.1. Typical soil composition

However, the relative proportions of the various soil constituents vary depending on the soil genesis, development, and result in fundamental soil physical and chemical properties such as texture, density, organic matter content, pH values, water holding capacity, soil stability, and in general soil fertility.

The soil solid phases can be inorganic or organic, with the inorganic phases consisting of stones and gravel, sand silt and clays, and the organic matter made

of biological debris, carbohydrates, proteins, lipids, nucleic acids, lignin cellulose and humic substances. Chemical composition and properties of the mineral and organic soil solid phases vary in different locations, as pedogenetic factors continuously induce mineral wethering leading to the formation of new minerals and organic moiety, conferring physico-chemical properties, specific characteristics to a given soil. Soils develop horizons of variable thickness, aspect, mineralogical composition and organic matter content, depending on the progress of pedogenesis. Soil horizons are denoted with A, B, and C horizons and are also used as diagnostic feature for soil classification. Horizon A is the surface soil layer, which is covered by plant litter of natural and forest soil, whereas in agricultural soils the A horizon is conditioned by plowing and is denoted as A_P. The A horizon is generally the richest in organic matter and nutrients, and harbour the larger soil biotic communities, although earthworms and soil fauna may move downward the soil profile to more than 1 m depth and some microbial species may be present in the deeper horizons. The B horizon is less organic and its chemical characteristics depend on the leaching processes from the A horizon and mineral depletion due leaching. The C horizon consists of weathered rock fragments mixed with materials leached by the upper horizons.

Continuous additions, movement and depletion of materials in soil, along with exchanging of energy in the form of radiation make soils thermodynamically open systems, with the main energy inputs being solar radiation, organic matter in the form of rhizodeposition, plant litter and biological remains.

Owing to the reactivity of the soil mineral particles and organic matter, and depending on the pedogenetic factors, soils generally show the formation of aggregates of variable diameters, formed by organic and inorganic constituents, and influencing the behavious of soils at the microscale to the lanscape scale. In fact, soil aggregates confer variable resistance to the sloughing action of H_2O and to erosion, and also define ecological space colonized by distinct microbial communities. The shape and size distribution of soil aggregates also induce the

formation of soil pores, which are filled by the soil solution or gases such as CO_2, CH_4, NO_3, NO_2, and N_2O, HS and HSO_4^-, produced by soil microbial activity, depending on the red-ox conditions of soil. For this reason, soils are the natural bodies that support the plant growth and share interactions with hydrosphere and atmosphere, thus playing an important role for the subsistence of life on Earth. For example, in forest soils two thirds of plant-borne organic matter is decomposed in one year (Killham 1994), with complex polymerized substances such as cellulose, hemicelluloses, and lignin, being decomposed much slower than simple sugars, nucleic acids and phospholipids. The organic matter decomposition in soil leads to the release of nutrients (e.g. N, P, S) in their mineral forms, therefore it is also called mineralization, and is carried out by the soil microbial communities. Soil microbes express specific biochemical and physiological functions such as extracellular enzymes and membrane transporters, and the mineralization rate vary depending on the pedo-climatic conditions and microbial activity, and is impacted by soil pollution and degradation.

Progresses in the bio-molecular characterization of the soil microbial communities by metagenomics have clearly shown that soils host the largest known microbial diversity among terrestrial ecosystems (Torsvik and Øvreås 2002), but to date establishing direct links between microbial diversity and specific metabolic processes is still very difficult (Nannipieri et al. 2003).

Soil mineralization activity is fundamental for the ecosystems productivity, as the mineral nutrients are the plant available fractions, although it has been reported that some carbon and nitrogen macromolecules such as glucose and aminoacids may be taken up by plants and microorganisms (Kuzyakov 2006, Paungfoo-Lonhienne 2012) However, the bioavailability of mineralized nutrients in soil is mediated by the reactivity of soil particles, especially secondary minerals and organic matter which react with nutrients through various mechanisms, as illustrated in the paragraphs 1.2 and 1.3.

Soil solid phases originate from weathering of rocks and minerals over time, caused by physical, chemical or biological factors. Physical weathering due to

temperature variations, geological forces, rainfall or ice formation, mainly causes breaking down of parent rocks without inducing mineralogical changes. Differently, chemical weathering due water dissolution or biotic activity dissolves pared materials and induces mineralogical changes with formation of new soil minerals. All three weathering factors normally act synergistically, with predominance depending on the soil location. For example, in soils located at high latitudes the predominant factors are the physical ones, mostly related to mineral stability under wide temperature changes, and related to the H_2O cycles.

1.1.1 Inorganic soil soil phases

Main inorganic soil solid phases are gravel, stones, sand, silt, clay minerals, and Fe- and Mn-(hydro)oxides, and the relative proportion of sand, silt and clays makes up the soil texture. Mineral and elemental composition of soil inorganic phases reflects that of soil parent materials, mainly formed by primary aluminosilicate minerals, in which Al, Si and O account for more than the 80% of total mass.

In soil minerals, both Si and Al are in cationic form (Al^{3+} and Si^{4+}) while O is in the anionic form O^{2-}. The three ions can combine together to form Si tetrahedrons and Al octahedron, in which they coordinate with either 4-6 O^{2-} or OH^- ions. Therefore, soil inorganic phases expose O or OH groups to the soil solution, and this is a key aspect conditioning the reactions at the surface of soil inorganic phases, for example by reacting with cations that neutralize the eventual negative charges at the mineral surfaces. Consequently, the surface chemistry in soils is dominated by the reactions that occur on oxide and hydroxide rich, highly hydrophilic surfaces, reacting with cations that neutralize the negative charges of the mineral surfaces. Formation of tetrahedra and octahedra depend on the radii of the cations and anions, and in general the central coordinated ion radius ratio determine the crystal geometry of the various minerals. The oxygen anion can be considered the main coordinated ion in soil minerals; Aluminium can have a 4^- or

6^- valence, resulting in tetrahedral and octahedral coordination status, respectively, and are present in all primary and secondary soil minerals. When in the silica tetrahedron Si is replaced by Al, the charge deficit of the mineral crystals is balanced by monovalent or divalent cations. Similarly, when in the octahedra Al is replaced by divalent cations (e.g. Mg), this generates a charge deficit which can be balanced by cations. Different arrangements of the silica tetrahedra make minerals differently prone to weathering, and determine the number covalent bonds that the minerals can form. Layer silicate classification relies on the number of tetraheral and octahedral sheets they share in their main structure (e.g. 1:1, 2:1, 2:1:1 layers), as schematically represented in in Figure 1.2.

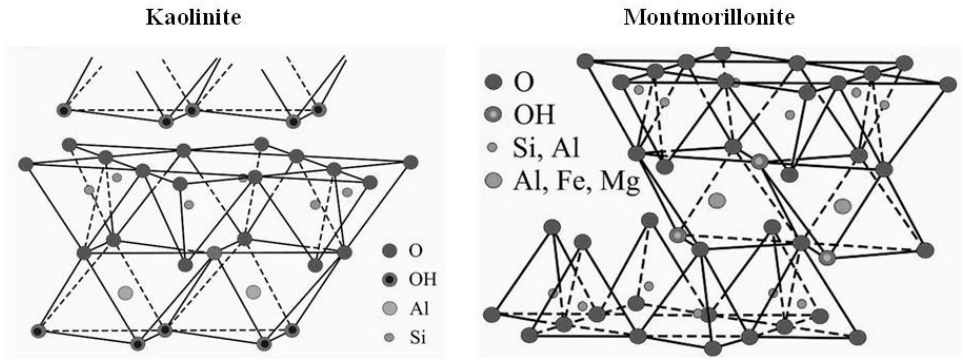


Figure 1.2. Left: structure of kaolinite (1:1 layer silicate). Right: montmorillonite (2:1 silicate)

The 1:1 layer silicates are typically constituted by one tetrahedral (silica) sheet combined with one octahedral (alumina) sheet. Kaolinite mainly represents this class of minerals, and the tetrahedral and octahedral sheets in a layer of a kaolinite crystal are held together tightly by O atoms mutually shared by the Si and Al cations in their respective sheets. These layers, in turn, are held together by H-bonding. Consequently, the structure is fixed and no expansion occurs between layers when the clay is wetted, and no H_2O nor cations can enter the structural layers of a 1:1 type mineral particle.

The 2:1 layer silicates present an octahedral sheet sandwiched between two tetrahedral sheets, characteristic of two groups: expansible (e.g. smectites and vermiculite) and non-expansible (e.g.mica) types. The interlayer expansion occurs by swelling when the minerals are wetted, the water entering the interlayer space and forcing the layers apart. Montmorillonite is the most prominent member of this mineral group in soils, in which Al^{3+} is replaced by Mg^{2+} in some of the sites of the octahedral sheet. Likewise, some Si atoms in the tetrahedral sheet may be replaced by Al. These substitutions result in a surface net negative charge, therefore these minerals have cation high exchange capacity and marked swelling and shrinkage properties. Vermiculites are also 2:1 type minerals in that an octahedral sheet occurs between two tetrahedral sheets. In most soils vermiculites, the octahedral sheet is aluminum dominated (dioctahedral), although Mg^{2+} dominated (trioctahedral) vermiculites are also common. In the tetrahedral sheet of most vermiculite, considerable substitution of Al for Si has taken place. This accounts for most of the very high net negative charge associated with these minerals. Water molecules, along with Mg^{2+} and other ions, are strongly adsorbed in the interlayer space of vermiculites. They act primarily as bridges holding the units together rather than as wedges driving them apart. The degree of swelling is, therefore considerable less for vermiculites than for smectites. For this reason, vermiculites are considered limited-expansion clay minerals, expanding more than kaolinite but much less than the smectites. The cation exchange capacity of vermiculites usually exceeds that of all other silicate clays, including montmorillonite and other smectites, because of very high negative charge in the tetrahedral sheet.

Smectite, vermiculite, and other expansible clay minerals can accommodate relatively large, inorganic cations, such as hydroxy polymers of aluminum, iron, chromium, zinc, and titanium, and cationic organic molecules, such as certain aliphatic and aromatic amines, piridine etc., between the layers. Moreover some polar organic molecules may replace adsorbed water on external surfaces and in

interlayer positions causing changes in mineral surface from hydrophilic to hydrophobic and the loss of tendency to bind water.

In 2:1:1 the crystal unit contains two silica tetrahedral sheets and two magnesium-dominated trioctahedral sheets. This silicate group is represented by chlorites, they are not present in all soil types, and become smectite or vermiculite upon weathering.

The presence of different layered silicates in soil is very important because they confer variable reactivity and chemical buffering capacity to soil, and also influence the biochemical activity by retaining the nutrients with increasing strenght.

1.1.2 Soil organic phases

The soil organic matter (SOM) is composed of molecules such as structural and functional sugars, lipids, nucleic acids, proteins and secondary metabolites, produced and released by living and dead cells, at various stages of decomposition. The SOM exists as discrete organic phases associated or not with mineral surfaces, or dissolved into the soil solution. The particulate SOM has various particle sizes, variable solubility and richness of functional groups and variable decomposition stages (Schmidt et al. 2011). The SOM content confer important physical, chemical and functional properties to soil, as it acts as cement among soil aggregates improving the soil structure, it retains water thus improving the soil water holding capacity, its dark color allows the captures of the solar radiation with buffering effects on soil temperature oscillations, and it is a reservoir of nutrients that sustain the soil biological activity.

The SOM fractions have been traditionally characterized using extractions in alkali (e.g. NaOH, Na-pyrophosphate) under inert atmosphere (e.g. N_2), followed by acidification at pH 2 with mineral acid (e.g. HCl) for separation of the humic fraction of the SOM. In the classical SOM fractionaltion scheme (Figure 1.3) the soluble phase is termed fulvic acids (FA), the precipitate phase is termed humic

acids (HA), whereas the SOM fraction which is insoluble in alkali is called humin.

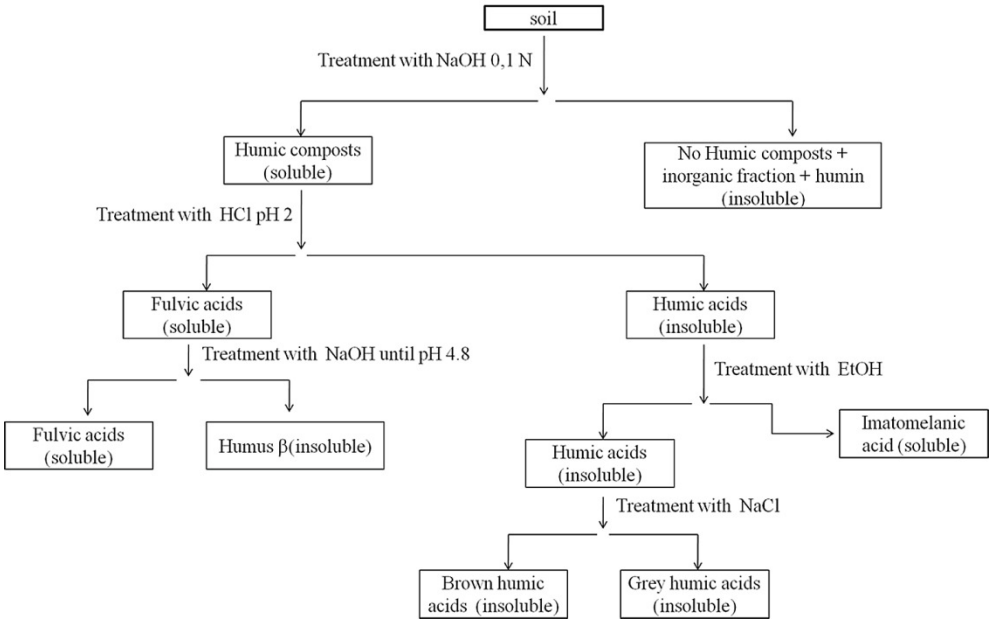


Figure 1.3. Classical SOM fractionation scheme

The SOM chemical properties and molecular composition of the humic substances HS are of major importance for the aims of my PhD project and therefore, a detailed introduction on the HS is reported in paragraph 1.2.1.

1.2 Humic substances

Humic substances are natural organic substances ubiquitous in nature and often represent the dominant component (70-80%) of the SOM, and have been traditionally described as polymeric structures with high molecular weight and of brown to black colour, formed by secondary synthesis reactions during the decomposition of plant, animal, and microbial macromolecules, rich in carboxylic and phenolic groups, and therefore present in the soil environment as negatively charged branched polyelectrolytes, with different chemical composition

depending on climatic, morphological and microbiological features under which the soils develop (Stevenson 1994).

Scientific studies on SOM date back to the 18th century as its relevance SOM was somehow recognized at the dawn of the modern Chemistry, and its positive influences on soil chemical, physical and biological fertility and plant growth were stated even before the esablishment of soil science. The most important chemists such as Wallerius in the 18th Centruy, Lemonossov and Barzelius in the mid 19th Centruty, were among those who contributed to assess the chemical composition of the SOM. In particular, Berzelius was the first who classified the HS fractions as humic acid (HA), fulvic acids (FA) and humin, on their solubility in alkali, metal complexing capacity, and chemcal inert, respectively. This HS fractionation scheme has been the most accepted for more about 150 years by soil and environmental chemists.

In this classical fractionation scheme of HS, the formation of HA was supposed to occur by the partial degradation of plant biopolymers (e.g. lignin), followed by reaction of condensation and random polymerization reactions. This hypothesis was introduced by Fischer and Schrader (1921) and supported by prominent chemists like the nobel laureate Selman Waksman. Later, the important role of microorganisms in HA synthesis was introduced, which also explained the relative richness in N of the HS, hypothesised as due to microbial protein enrichment during the microbial SOM turnover (Waksman 1938). Flaig (1966) was among the first to use the isotopes for SOM studies, who proposed that HA were forms by the oxidative polymerization of lignin-derived phenolic monomers. An alernative HA synthesis pathway was proposed by Maillard who hypothesizd that HA could be formed by abiotic reactions between reducing sugars and aminoacids or other nitrogenous componunds. Other important contributes to the classical HS view came from the work of Dragunov et al. (1948) who prposed that quinone-like molecules could be considered as the building blocks of HA, as they could be react with carboxyl groups, and Stevenson (1964) who proposed

that HA could be formed by reactions between phenolic molecules, quinones and benzoic acid, and sugars and peptides.

A partial confirmation of the classical view of the HS was brought from Schnitzer ad Khan (1972) and Schnitzer (1978) by using new chemical approaches. In particular, Schnitzer (1978) reported that HA permanganate oxidation in alkaline solutions and HA sequential oxidations with alkali or acids of increasing strenght mainly released aromatic various types of carboxylic acids, phenolic acids, and aliphatic mono or poly-carboxylic acids, thus demonstrating that the aromatic molecules limnked by aliphatic colecules formed the core structure of HA. In subsequent studies, aromatic compound such as naphthalene, phenanthrene and pyrene derivatives were identified as components of the HA structure (Schnitzer 1978). This classical view of HS was further supported by Kononova (1961), and the HS fractionation scheme in Figure 1.3, has been the main approach for studying the HS composition for decades (Tate 1992, Stevenson 1994).

The classical view of the HS is challenged by the recent advances in SOM studies, and FA and HA are no longer considered as specific compounds, but operationally defined SOM fractions of variable molecular weight and properties.

Currently the most accepted hypothesis on the HS composition is that HS are a supramolecular aggregation of self-assembling heterogeneous molecules originated from the decomposition of biological macromolecules (Wershaw 1999), stabilized by weak dispersive forces, such as hydrophobic interactions and H-bonds (Piccolo, 2001). In this view, the FA and HA fractions of the HS can be considered as formed by the supramolecular aggregation of small organic molecules, rich in functional groups, held together by weak molecular interactions (Piccolo 2002). Pioneer data on the supramolecular structure of the HS were brought by Piccolo (2001), who showed the process of supramolecular association in a model experiment in which the behavior of low-molecular weight organic compounds (LMWOCs) considered as monomeric constituents of HS, either singly or as mixture, was investigated by size exclusion chromatography (SEC).

The mixture revealed less retention peaks than in the case of single compounds, and one of the peaks indicated higher molecular weights than those of individual compounds, demonstrating that individual compounds may form supramolecular associations. Simpson et al. (2002) by two-dimensional diffusion-ordered NMR spectra (DOSY) of HA dissolved in D_2O or deuterated dimethylsulfoxide (D_2-DMSO), showed that the molecular mass of the dissolved HAs could not exceed 2000 Da, and that the observed spectra were compatible with those of fragments formed by of 3-8 sugar units, C_{16} and C_{18} fatty acid esters, lignin dimers, trimers and tetramers. Such compounds could explain 5%-50% of the analysed HA, while fragments with more than 8 units were considered highly unlikely. Later, Keller and Simpson (2006) confirmed that soil borne HS represent a very complex mixture of microbial and plant biopolymers and their degradation products but not a distinct chemical category. Zhang et al. (2011) studied the structure of HA and FA of mangrove soils by thermogravimetric analysis (TGA), Fourier Transformed Infrared Spectroscopy (FTIR), ^{13}C nuclear magnetic resonance (^{13}C NMR) and potentiometric titrations, and reported that more long-chain hydrocarbons and nitrogen compounds existed in humic acids than fulvic acids at both sites, and that HA had more aliphatic and fewer aromatic organic compounds than FA. The content of carboxylic groups accounted for most of the total acidity and contributed more to the fulvic acids than the humic acids. Sutton and Sposito (2005) in a critical review sustained that humic fractions extracted from soils contain recognizable biomolecular fragments that have been specifically excluded from traditional definitions of humic substances proposed by Stevenson (1994), Tan (2003), Hayes and Clapp (2001). The authors suggest that a definition of humic components that is relevant to experimental studies should include all molecules that are strongly associated within a humic fraction, i.e. those that cannot be removed without significant alteration of the chemical properties of the fraction.

The current uncertainty about the soil HS structure is mainly due to the lack of a repetitive sequence and to the great variety of chemical and biological reaction involved in their genesis (Ziechmann, 1994).

The HS also have biological activity and interact with plant roots, stimulating the plant growth (Blanchet 1958; Guminski 1968). Possibly, the phytostimulative activity of HS is due to the presence of hormone-like molecules or hormone precursors. An auxin activity of HS has been reported (Hamence 1946; Muscolo et al. 1993, 1996).

1.3 Proteins and peptides

Proteins are biological polymers constituted by aminoacids, synthesized by all living organisms by the transcription of genes and translation of the mRNA, and play fundamental vital roles as functional or structural cell elements. Proteins are produced by all organisms, in all habitats, from really worm to really cold environment; in extreme salinity or pressure conditions. This fact imply that the same protein, for example an enzyme catalyzing a certain reaction, in extreme different environment present different chemical characteristics for the adaptation to the conditions.

Proteins are highly heterogeneous also in their size that generally range as order of magnitude from 10-300 kDa, with a chemical behavior from hydrophilic to hydrophobic, depending on the aminoacid composition and sequence within the proteins.

Structural proteins constitute the cell scaffold and structure, whereas the functional proteins are responsible for specific cell functions as catalysts (e.g. enzymes), membrane transporters (e.g. Na/K pumps), performing regulatory functions (e.g. hormones), cell adhesion (e.g. adhesines), and receptors. Proteins can perform their functions alone or as part of a multi-protein complex, and also bound to other molecules. Enzymes represent a functionally relevant class of

proteins which catalyze chemical reactions through the reduction of activation energy, thus regulating the cell metabolism.

The genetic information carried by the mRNA is crucial because determines the aminoacid sequence (primary structure) and consequently the main protein characteristics such as the spatial arrangement of adjacent aminoacids in the linear sequence (secondary structure), its three-dimensional conformation (tertiary structure) and as a result protein surface chemical behavior. After translation, polypeptides are modified in various ways to complete their structure, designate their location or regulate their activity within the cell. After synthesis, proteins aquire a specific arrangement in space called conformation due to folding in specific protein domains, and folding also depends on the chemical composition of the surrounding environment or by post translations modifications. Folded proteins also display peculiar surface characteristics that determine their capability to interact with other molecules, which can not be predicted by the simple primary structure.

Post translational modifications are mechanisms for differentiate the protein functions through covalent modifications, such as the addition of chemical groups or the cleavages (e.g. phosphorylation, glycosylation) of the protein backbone. For example phosphorilation is a modification that activate proteins and conferres charge, and hampers the membrane crossing. Not all the aminoacids can be phosphorylated, in the mammalian phosphoproteomes serine, threonine and tyrosine are subjected to this kind of modification, while bacterial and fungal phosphoproteomes may also have histidine and aspartate phosphorilation. The introduction of the charged phosphate group induces altered conformations in local protein microenvironments (Johnson et al. 2001), and the negative charge is often paired by cationic arginine residues. Methylation and acylation are other important mechanisms in regulating gene expression. Covalent glycosylation of proteins is relatively rare in prokaryotes, more common in eukaryotes, and it cooperate in protein folding, resistance to proteolysis, and in cell-cell adhesion. Proteins can be linked also to lipids residues forming lipoproteins.

Most of the proteins of high molecular weight are macromolecules formed by the combination of several smaller subunits, either identical or different. In some multimeric proteins the subunits are held together by hydrogen bonds, by electrostatic attraction of positively and negatively charged amino acid side chains or by other non-covalent interaction. However, in several proteins the subunits are bound to each other by disulfide bonds usually supported by other non-covalent interactions.

Chemically, proteins are amphoteric polyelectrolytes (Branden and Tooze, 1991), as the aminoacids have different chemical characteristics: some proteins are apolar, while other proteins are polar and electrically charged. The hydrophobic peptides normally are folded in the inner part of globular proteins, whereas the hydrophylic pepetides are generally found outside protein surface. Owing to the aminoacid reactivity and depending on the primary structure, proteins fold immediately after synthesis, and folding confer to proteins their specific functions.

1.4 Proteins in soil

Proteins in soil originate from plants, animals and microorganisms, either through active excretion or passive release. Generally, 96-99% of soil total N is organic and after acid hydrolysis, amino acidic N accounts for 30 -50% of the total N in soil (Stevenson, 1994). It is assumed that most of the aminoacids released from acid hydrolysis derive from proteins and peptides present in soils. Modern techniques, such as Curie-point-pyrolysis gas chromatography-mass spectrometry (Cp Py-GC/MS) pyrolysis-field ionization mass spectrometry (Py-FIMS), ^{15}N nuclear magnetic resonance spectroscopy (^{15}N NMR), X-ray photoelectron spectroscopy, and X-ray absorption near-edge structures spectroscopy (XANES) have shown that proteins are the most abundant organic N compounds in soil (Nannipieri and Paul, 2009). However, origin of proteins and peptides in soil and their links with the measureable soil functions (e.g. SOM decomposition,

enzymatic activity) are largely unknown, despite the vast literature on the subject (Nannipieri and Paul, 2009).

Protein stabilization in soil is due to their interactions with surface-reactive soil solid phases such as clay minerals, Fe-, Al-, and Mn-hydr(oxides) and humic substances.

1.4.1 Interactions of proteins with soil inorganic phases

Concerning the protein interactions with the inorganic soil particles, some early evidences of clay related protein protection against proteolysis were reported by Ensminger and Gieseking (1942) and Pinck and Allison (1951). The extensive research on this topic reviewed by Stotzky (1986), has shown that protein sorption onto kaolinite and montmorillonite minerals is rapid, especially at reaction conditions below the protein isoelectric point of the protein, likely due to ionization of the amino- and carboxyl- groups. Protein sorption by 2:1 minerals (e.g. montmorillonite) is mainly related to the protein intercalation into the interlayer expanded crystal lattice, whereas protein sorption by kaolinite is more related to surface area, less to its surface charge. Clay saturation cations and the protein conformation may be not major variables influencing the protein adsorption on clay minerals, although the role of the polyvalent cations on the protein adsorption and the enzyme catalytic activity has been also demonstrated (Burns 1986; Fusi et al. 1989; Gianfreda and Bollag 1996; Quiquampoix 2000; Rupert et al. 1987; Violante and Gianfreda 1995).

Studies on the interactions between clay minerals and proteins do not always report reaction conditions and relative purity of proteins and clay minerals used for the experiments. In fact, the presence of other proteins or other organic compounds may influence the sorption of the target protein. The same situation is for the clay minerals, often referred as ‘pure’ clays, because even small amounts of impurities such as Fe-(hydro)oxides or silica, may influence the protein adsorption by clays. In fact, silica and Fe-(hydro)oxides can interact with proteins (Norde 1986; Stotzky 1986), either by electrostatic forces or by entropically

favourable structural rearrangements. The protein-clay electrostatic interactions typically fit the Freundlich isotherm model (Quiquampoix and Ratcliffe 1992). In addition to electrostatic interactions, van der Waals forces and hydrophobic interactions have been shown to be involved in protein sorption by clays (Hamzehi and Pflug 1981; Quiquampoix and Ratcliffe, 1992), particularly the small and uncharged proteins, or proteins with non polar domains. Formation of protein multiple layers has also been suggested (Violante and Gianfreda 1995). Protein adsorption by hydrophobic and ionic interactions are also accompanied by entropy gain caused by conformational changes of the protein during the adsorption (Haynes and Norde, 1995). Yu et al. (2000) studied the sorption of various proteins and peptides on pyrophyllite, and reported that the clay mineral denatured the adsorbed proteins due to dehydration and non specific interactions inducing torsion tension and destabilization of the tertiary structure of the adsorbed proteins. Resuming the vast literature on the sorption of various proteins by model soil inorganic phases, it has been shown that that sorption is mainly caused by electrostatic interactions, even more for proteins with lower conformation stability, indicating that entropy gain upon conformation rearrangements may be also an important factor in protein sorption onto soil inorganic phases. Protein sorption onto soil minerals has been resumed by Rigou et al. (2006) and Quiquampoix and Burns (2007). Other studied variables include contact mode (static or not), contact time (minutes-hours), contact pH and temperature (controlled or not), ionic strength, and protein-to-clay ratios. Due to the very broad range of protein size, conformation, electric properties of proteins, a general model of protein interactions with clay minerals is currently not available.

The fate of specific proteins such as glomalin and the insecticidal *Bacillus thuringensis* endotoxin (Bt toxin) in soil, have attracted the attention of soil scientists due to their ecological and commercial relevance (De Barjac and Frachon 1990). These studies well illustrate the difficulty of extracting,

quantifying and characterizing proteins directly extracted from soils. Glomalin is an extracellular glycoprotein produced by arbuscular mycorrhizal fungi (Wright and Upadhyaya 1998), and usually quantified by either the Bradford colorimetric or immunological (ELISA). However, both techniques produce artifacts due to non specific reactions with phenolic compounds and litter and humic components. Analysis of glomalin-related proteins extracted from soil by NMR have shown the presence of humic substances (Schindel et al. 2007).

Similar observation were made on the Bt toxin, commercially available for agriculture use. Such studies have shown that the toxin is stabilized in soil by clay minerals and humic acids in its active form (Lee et al. 2003). From an *in vitro* study Helassa et al. (2009) reported that sorption of the Bt Cry1Aa (released by Bt-transformed rice) was greater onto montmorillonite than kaolinite and that desorption was less with water than with alkaline buffers in the presence of surfactants such as CHAPS, Triton-X-100 and Tween 20. The general assumption that protein adsorption can reduce its mobility and alter its biological properties due to conformational changes or orientation on the clay surface (Baron et al. 1999; Servagent-Noinville et al. 2000), is not confirmed by studies on specific proteins such as Bt toxin and glomalin. Koskella and Stotzky (1997) observed that insecticidal toxins produced by *B. thuringiensis* subspp. *kurstaki* and *tenebrionis* bound to kaolinite and montomorillonite were not mineralized by mixed microbial cultures and also retained their insecticidal activity, whereas the opposite occurred for the free toxins. In that case, protection against proteolysis could be due to protein adsorption (Kleber et al., 2007).

These model studies are of particular interest for understanding the fate of proteins in soil. In fact, most of the experimental evidences show that protein adsorption by inorganic soil phases is mostly irreversible, nearly independently on protein total concentration, and may only be partially reverted by variations in pH value and ionic strength of the soil solution. Sorption of enzymes and other proteins, actively or passively released by soil organisms, on clay minerals and humic substances can have an ecological relevance because it reduces their

availability as C and N sources for microorganisms (Calamai et al. 2000; Stotzky 1986; Violante and Gianfreda 1995). Despite protein extraction yields are important in soil proteomics, protein desorption from soil colloids is poorly known as well as the factors controlling it, such as the the effects of pH shifts induced by the different protein extraction protocols. Some information on complexity of the pH-dependent interactions is well illustrated by the variability of enzyme adsorption depending on the physico-chemical conditions. Quiquampoix et al. (1989) reported that the α-glucosidase of *Aspergillus niger* did not adsorb on the surface of montmorillonite above pH 6.0 whereas the bovine serum albumin (BSA) did not adsorb above pH 6.5 (Quiquampoix and Ratcliffe 1992). Quiquampoix and Mousain (2005) showed that acid phosphomonoesterases released by ectomycorrhizal fungi could have different behaviour towards clay minerals: sorption or total repulsion for *Suillus mediterraneensis* or *Pisolithus tinctorius*, respectively. Importance of pH-induced conformational changes on the extent of adsorption of specific proteins has been reviewed by Quiquampoix (2008).

Studies on the formation of mixed lipido-protein layers at solid phase interfaces can be also important because they can help to understand the nature of interactions in such this type of interface which likely occur in soil (Bos and Nylander 1996). Such interactions may significantly change the protei-clay interactions by affecting hydrophobicity and hydrophility properties and conformation of proteins.

1.4.2 Protein interactions with the soil organic matter (SOM)

The protein surface and conformation are highly variable and heterogeneous, exposing either positive and negative charges, groups with H-bonding capacity, whereas the non polar protein regions are typically folded in the inner structure. Owing to such complexity, proteins can interact with other proteins and several types of molecules, either of biological or non biological origin with strong

chemical bonds (e.g. covalent) or weak surface interactions (e.g. H-bonds, van der Waals interactions forces).

In the soil environment, proteins may also share chemical interactions with the soil organic matter (SOM) in several modes, including ionic interactions (both repulsive and attractive), H-bonding, hydrophobic interactions, hydration forces, acid-base interactions and van der Waals forces. Owing to complexity of proteins and HS, interactions between proteins and HS are considerably less characterized than protein-clay interactions, even if such interactions have been long postulated (Bremner 1951, Hsu and Hatcher 2005, 2006, Simonart et al. 1967, Swaby and Ladd 1964, Zang et al. 2000). Cheshire and Hayes (1990) hypothesized that interactions between proteins and polysaccharides involve mainly oxygen or hydroxyl functions of the proteins, through hydrogen bonding for neutral and negatively charged polysaccharides. Complexation of humic acids (HA) and specific domains of the human serum albumin have been reported by Ding et al. (2011). Additional information on the protein stability in soil and interaction with HS have been provided by studies on soil enzymes; enzymes are ideal for testing the proteins interactions with humic substances because effects can be also evaluated either by changes in the enzymes' conformation or kinetic properties, after sorption by humic substances, and are not resistant against proteolysis. For example, Ceccanti et al. (1989) reported that humo-enzyme complexes extracted from soils retained their catalytic activity, whereas Solaiman et al. (2007) reported enzyme degradation after contact with HS. From a model study, Tan et al (2009) reported synthetic humic acids were able to complex lysozyme, and such complexation altered the catalytic activity (Zang et al., 2000). The vast literature on this topic has been reviewed by Nannipieri et al (1996).

The electrostatic attraction between HS and differently charged aminoacid side chains has been considered the main interaction mechanism between protein and HS. However, the hypothesized supramolecular structure of HS, stabilized by weak dispersive forces such as hydrophobic interactions and H-bonds (Piccolo, 2001, 2002), requires the new hypotheses to explain the possible mechanisms

explaining the chemical interactions between proteins and HS, other than electrostatic attraction and physical entrapment.

In this light, hydrophobic and not only ionic interactions, may play a major role in protein-HS interactions. There is increasing evidences of hydrophobic interactions between proteins and humic substances (HS), and thus proteins-HS interactions also at pH values above the protein isoelectric point, when both proteins and HS are both were negatively charged (De Kruif et al. 2004). However, it can not be excluded that in real protein-HS interactions several mechanisms may act simultaneously. For example, thermodynamically favorable hydrophobic interactions following electrostatic complexation between lysozyme and synthetic humic acids were reported by Tan et al. (2008). Moreover, hydrophobic interactions may be also due to dehydration of newly formed protein-HS, as hypothesized by Tomaszewski et al. (2011). Hydrophobic interactions may also reduce the affinity of the positively charged protein for the negatively charged soil mineral phases, as reported by Tan et al. (2009) for lysozyme. However, our knowledge on the protein SOM interactions comes from oversimplified *in vitro* experiments, with highly purified HS and proteins, in which some of the soil factors such as are not involved.

Proteins can be physically entrapped by humic substances as shown by Tomaszewski et al. (2011).

Yuan and Zydney (2000) studied the role of divalent cations (e.g. Mg, Ca) in soil solution, and showed that they can form bridge bonds between negatively charged protein and SOM functional groups. The role of other important small ions (e.g. NO_3^-, NH_4^+, SO_4^{--}) present in relatively high concentration in the soil solution on the protein sorption has been also neglected so far.

Futher information on protein interactions and persistence in SOM were also provided by studies on litter decomposition (Criquet et al. 2002). Huang et al. (1998) and Miltner and Zech (1999) reported that plant proteins dominated in fresh litter whereas proteins of microbial origin predominated in aged litter. In a

model litter decomposition study, based on the proteomic analysis of *Pectobacterium carotovorum* and *Aspergillus nidulans*, either grown pure and co-culture or on beech litter, Schneider et al. (2010) showed that various proteases, pectinases and cellulases involved in litter decomposition and that decomposition was initiated and carried out by the fungus. It can not be excluded that these observations may also apply to soil, and a correlation between the SOM decomposition rate and proteins concentrations may exist, particularly in the early stages. In this sense, there is a need to monitor the protein synthesis and persistence in soil by the monitoring of intracellular proteins.

However, information on protein turnover during SOM decomposition is still scarce.

1.4.3 Peculiar proteins: Prions

The prion (PrP) is well known as a the infectious agent of the human Transmissible Spongiform Encephalopathy (TSE) , which is a rare and fatal neurodegenerative disorder characterized by tissue deposition of a misfolded isoform of the cellular prion protein (PrP^{C}), commonly referred to as PrP^{Sc} (Figure 1.4), detected in all infected species (Prusiner 2004; Wang et al. 2010). In contrast to PrP^{c}, PrP^{Sc} is aggregated, highly resistant to inactivation, with β-sheet content higher (43%) than that of PrP^{C} (3%) (Pan et al. 1993) and able to seed conversion of PrP^{c} to PrP^{Sc} 'replicate', and thus initiate prion propagation (Prusiner 2004; Wang et al. 2010). Prion proteins are also responsible for several other important animal diseases.

As proteins, prions released by infected organisms can spread out into the environment can interact with soil and, because of the close contact with soil, animals have a substantial opportunity of transmission (Saunders et al. 2012; Smith et al. 2011). Spreading of prions in environment can be caused by decomposition of infected carcasses (cattle, sheep, and deer) buried into soil, accidental dispersion from storage plants of meat and bone meal, contaminated organic fertilizers and residues used in agriculture (Andreoletti et al. 2002), and

animal infection occur through either accidental (i.e. during grazing) or intential (i.e. for complementing the mineral nutrition) soil ingestion.

It has been reported that subtilisin-like enzymes, belonging to the subfamily of serine proteases, and to a lesser extent keratinase and bacterial alkaline serine proteases can degrade various prion proteins (Atalo et al. 1993, Kristie et al. 2000, Jhonson et al. 2011). However, significant reduction of misfolding of various prion types, has been achieved using such enzyme under very high temperature and strongly alkaline pH values, which are non realistic conditions for soils, and only recently enzymatic degradation of various prions in conditions closer to those of natural soils by Prionzyme (Saunders et al. 2010) and microbial proteases produced by the soil bacterium *Streptomyces galbus* var. *achromogenes* (Tsiroulnikov et al. 2004), and from lichens (Johnson et al. 2011) have been reported, although they still need further verification. It has been reported that prions can persist in contaminated soils (Genovesi et al., 2007) and retain their pathogenicity capacity for several years (Maddison et al. 2010; Saunders et al., 2011). A recent review on the fate of prions in soil has been prepared by Smith et al. (2011).

The main mechanisms of stabilization and resistance of prions is soil are still not clear, and only few specific studies on interactions between prions and soil colloids have been carried out. Rao et al., (2006) showed an irreversible interaction of the recombinant bovine prion (RecPrP) both with catechol polymers and birnessite-catechol polymers mixtures, as model humic substances and humo-mineral complexes, suggesting that prion proteins should be strongly retained in soils rich in OM, with consequent low risks of disease spread out.

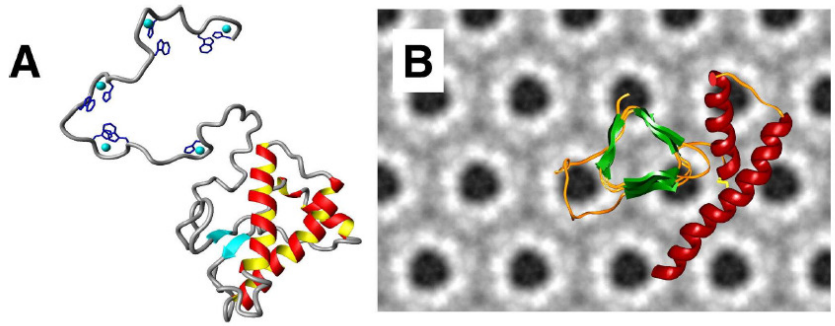


Figure 1.4. (A) NMR structure of PrPC with a notional addition of the N-terminal region (which is unstructured in NMR analyses). Copper ions (turquoise spheres) and coordinating side chains (dark blue) are shown. Note that the Cterminal globular domain has two short β strands (turquoise arrows) and three α-helices (red; image courtesy of Glenn Millhauser, University of California Santa Cruz). **(B)** A structure for PrPsc deduced from electron microscopic analyses of purified material. The background represents the two-dimensional crystals of PrP27–30 after image processing and the foreground a notional structure for PrPsc with the central region of the protein rearranged into triangular stacked β-helices (image courtesy of Cedric Govaerts and Holger Wille, University of California San Francisco)

David et al. (2011) reported that clay-enriched soils enhance prion transmissibility in selected populations of mule deers, and in general soils are key determinants in the incidence of prion-related diseases (Sanuders et al., 2011, 2012). However, as for other proteins, no specific studies on prions-HS interactions between have been carried out.

1.5 Analytical techniques for studying proteins and humic substances

Several techniques can be used to study proteins and humic substances, in particular for assessing their chemical composition, molecular structure, solubility and intra- or inter- molecular interactions.

Both proteins and humic substances must be extracted and purified from living organisms and from soils, respectively, prior to their molecular and chemical assessment.

1.5.1 Protein analysis

Protein extraction from living organisms can be performed using lysis buffers containing various chemical agents, and under different physical conditions, depending on the protein source. Among the available methods for protein extraction, freezing-thawing cycles, sonication, cell permeabilization, detergents, and organic solvents are the most used, singly or in combination, also depending on both the complexity of the biological source and the 'hardness' of the proteins to be isolated. Proteins are then precipitated by differential solubilisation using organic and inorganic solvents, including phenol, chloroform, acetone, ammonium sulphate.

Several techniques can be used to study protein structure, composition and characteristics; and, at the same time, the way they interact.

Electrophoresis is the most widely used fingerprinting technique to separate proteins according to their size and the net electric charge, under the action of an electric field. The various electrophoretic techniques are nowadays based on the characteristics of the polyacrylamide gel types, and they can be used as crude measures of molecular weight and purity, or as preparative separation prior to detailed analyses of blotting sensitivity, selective detection, isoelectric focusing (IEF). In fact, the addition of specific dyes or antibodies, or the creation of specific pH or denaturing gradients make the electrophoretical techniques suitable for a medium-high resolution protein analysis. The two dimensional gel electrophoresis (2-DE), allows an orthogonal separation according to the protein size and charges and the relatively large gel dimensions permit the detection of small changes in complex protein samples (up to 500) such as post translational modifications.

Usually, one-dimensional gel electrophoresis is normally used to separate simple protein mixtures (e.g. isoforms), whereas two-dimensional electrophoresis (2-DE) are used to separate complex protein mixtures and is preliminary to protein identification. The 2-DE technique was first proposed by Macko and Stegeman

(1969) and improved by Klose (1975) and O'Farrell (1975). Notwithstanding the limitations related to the limited number of protein spots and practical limitations in monitoring low abundant and highly hydrophobic proteins, and proteins with extreme acidic and basic isoelectric points, the 2-DE has become a standard fingerprinting technique for protein separation prior to MS analysis (Lambert et al., 2005). Important improvements of the 2-DE technique have concerned the use of fluorescent stains, in the differential gel electrophoresis (DiGE) (Unlu et al., 1997), with staining of proteins with different fluorophores so as to distinguish proteins of different biological samples.

The X-ray crystallography provides information of atomic structure within a crystal using the diffraction of X-rays by crystals, which has characteristic angles and intensities of these diffracted beam. Crystal structures of proteins and other biological macromolecules have been studied. Although since the 1950s thousands of proteins structures have been determined by X-ray crystallography this method is not routine for studying the protein interactions, due to the complex chemical procedures to obtain suitable protein crystals, and to my knowledge it has not been used for studying the protein interactions with natural colloids. A major limitation of X-ray crystallography is the protein size which should not exceed 100-1000 atoms, whereas macromolecules with thousands of atoms such as proteins are not well resolved. Some of these limitations can be overcome by using electron or neutron diffraction and synchrotron radiation techniques.

Circular dichroism (CD) is a technique based on the differential absorption of circularly polarized components of light, which can be measured by spectropolarimeters. For chiral molecules left (L) and right (R) handed polarized light components are differentially absorbed, and the resulting emitted radiation acquires elliptical polarization. In proteins, the optically active functional groups are the peptide bond (absorption below 240 nm), the aromatic amino acid side chains (absorption in the range 260 to 320 nm) and disulphide bonds (weak broad absorption bands centered around 260 nm). Other optical techniques such as,

ellipsometry, variable angle reflectometry, surface resonance plasmon, can also be used to study specific protein properties in terms of conformation and potential interactions, as the formation of specific chemical bonds produce a change in radiation propagation and diffraction.

Various spectroscopic techniques can be used for the study of the protein structures and potential interactions with other molecules. Among the techniques, the nuclear magnetic resonance (NMR) spectroscopy is one of the most used spectroscopic techniques to characterize proteins.

NMR and X-ray crystallography are the only two methods that can be applied to the study of three-dimensional molecular structures of proteins at atomic resolution. In particular, the NMR spectroscopy is the only method that allows the determination of three-dimensional structures of protein molecules in solution. In addition NMR spectroscopy is a very useful method for the study of kinetic reactions and properties of proteins at the atomic level. The NMR spectroscopy can be applied to structure determination by routine NMR techniques for proteins in the size range between 5 and 25 kDa, if phenomenon of aggregation or reduction of solubility do not occur.

Fluorescence emission spectroscopy offers a great specificity and sensitivity in protein studies, utilizing both inherent protein fluorescence activity or various natural or synthetic fluorophores. Many variants of fluorescence spectroscopy used to study proteins in solution can be applied to protein molecules adsorbed at interfaces.

In infra-red spectroscopy (IR) is a technique used for studying the protein structure, in particular the spectral region between 1100 and 1700 cm^{-1} can provide information on the protein conformation (Elliott and Ambrose, 1950). The most frequently studied IR parameter in protein studies corresponds to the stretching vibration of C=O group, as the frequency of this vibration depends on the nature of the H-bonding in which the C=O group participates, which makes it highly sensitive to the secondary structure adopted by the polypeptide chain (e.g.,

α-helices, β-sheets, turns, random coil structures), thus providing a fingerprint of protein secondary structure.

One of the first methods for sequencing amino acids in a peptide was the Edman degradation reaction. In this method, the N-terminal residue is labeled and cleaved from the peptide without disrupting the peptide bonds between the other amino acid residues. The original procedure required high amount of protein samples and with free N-terminal. Nowadays the Edman sequencing method is used coupled to electroblotting, and this combined methodology is referred as microsequencing.

Another suitable technique for studying the protein properties and reactivity is the quartz crystal microbalance, which is a device based on a piezoelectric crystal vibrating under a definite electric field. With this technique it is possible to immobilize various molecules and analytes in solution, record the initial crystal resonance, and the variations in the resonance frequency induced by chemical interactions between the partner molecules.

The characteristics and potentials of the quartz microbalance and other chemosensors to study protein interactions with soil constituents was reviewed by Puglisi et al. (2012) (see annex II of this thesis).

1.5.2 Mass spectrometry for protein analysis and identification

Mass spectrometry has nowadays established as the reference method to study proteins, as recognized by the 2002 Nobel prize in chemistry awarded to Fenn, Tanaka and Wüthrich (Aebersold and Mann 2003). The technique relies on the detection of ionized analytes according to their mass-to-charge (m/z) ratios, with the number of ions with each m/z ratio values registered by detectors and recorded as spectra of ion intensity versus the m/z ratio. Mass spectrometers employed in protein and proteomic analysis use either matrix-assisted laser desorption/ionization (MALDI) or electrospray ionization (ESI), but they vary widely in their operation modes and performances (Figures 1.5, 1.6). Both the ESI and MALDI have been the standard ionization methods coupled to gas and liquid

chromatography (LC)-MS, and also used as single or tandem modes (MS/MS). The ESI and MALDI ionization techniques, coupled with time of flight (TOF), ion trap, quadrupole orbitrap or hybrid tandem mass spectrometers to generate ion spectra have allowed the major advances in proteomic research, owing to their high resolution and sensitivity.

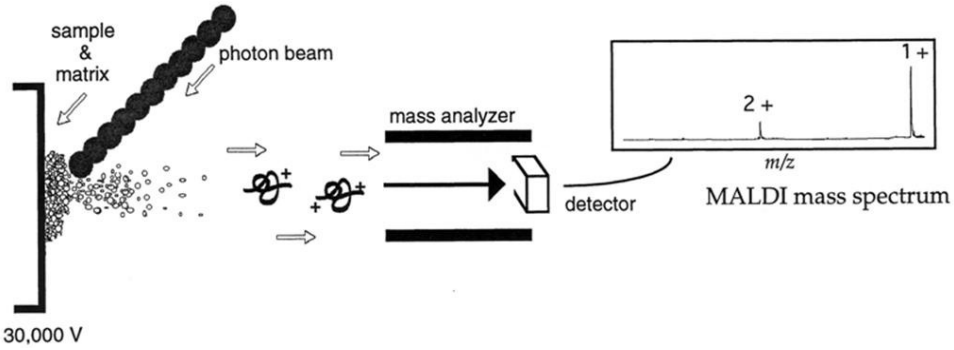


Figure 1.5. Illustration of matrix-assisted laser desorption ionization (MALDI). Gas phases ions generated by laser vaporization of a solid matrix/analyte mixture in which matrix strongly absorb the low energy laser radiation and acts as receptacle for energy deposition, that results in the vaporization and ionization of both matrix and analyte ions

Other analytical techniques such as desorption electrospray ionization (DESI), surface-assisted laser desorption/ionization (SALDI) along with new detectors for peptide mass fingerprinting such as MALDI-Qq TOF which also allows the aminoacid sequencing and the Fourier Transform ion cyclotron resonance (FT-ICR), also hold promising improvements of either sensitivity and accuracy of protein mass spectrometry.

Protein identification is based on the analysis of peptides generated by proteolyic digestion, prior to chromatographic separation. The most used proteolytic enzyme is trypsin, which hydrolyzes the protein on the C-terminal side of lysine and arginine, unless the subsequent amino acid in the sequence is a proline. This is advantageous as every peptide other than the protein C terminus has at least two

sites for efficient protonation, the N-terminal amino group and the C-terminal basic residue, so peptides are readily ionized and detected as positive ions.

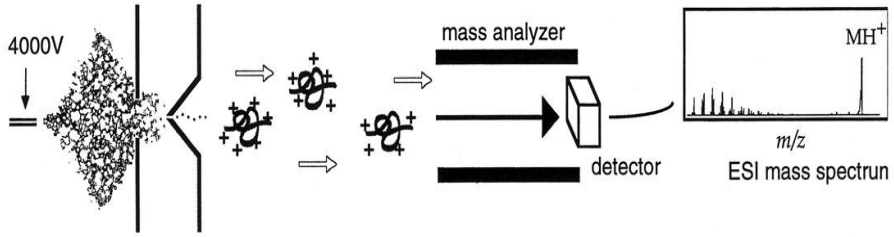


Figure 1.6. ESI generates ions directly from solution by creating a fine spray of highly charged droplets in the presence of a strong electric field. Subsequent vaporization of these charged droplets results in the production of gaseous ions

Therefore, the tryptic proteolysis produces unique peptides for a protein having a definite molecular mass, which can be separated by chromatography prior to MS/MS analysis. In this analytical line, the MS/MS spectra from individual digested peptides can be measured, and the experimental mass values are then compared with calculated peptide mass or fragment ion mass values, obtained by applying cleavage rules to the entries in databases containing the protein primary sequences. In this way, the MS/MS spectra generated by protein analysis are assigned by their precursor peptides by sequence searching. By using this method, a sequence search engine matches the detected spectra with theoretical fragment ion spectra generated from all known protein sequences. The usual approach for sequence search engines is to iterate over each observed MS/MS spectrum and, for each spectrum, scan through a FASTA file of protein sequences, selecting only the peptides with the same precursor m/z ratio, within a given tolerance interval. For this subset of candidate peptides, theoretical MS/MS spectra are generated and compared to the observed spectrum and a *score* is calculated to quantify the matching of each theoretical spectrum with the observed spectrum. The sequence coverage is the percentage of protein covered by the matching peptides. This value changes depending on the analyzed protein: in small proteins,

relatively few peptides have to be matched and typically high coverage is observed, whereas large and complex proteins produce high numbers of peptides to be matched and low coverage percentages is generally observed. The MS data are then filtered according to quality criteria (described below), the data are searched against proteomic databases. All search engines allow for the setting of a number of variables such as protein molecular mass range, pI range, mass tolerance for peaks, mass tolerance for collision ion detection (CID) precursor ions, mass tolerance for fragment ions, number of charges, number of peptides required for a match, and possible modifications to certain residues such as alkylation of cysteine or oxidation of methionine. There is no universal standards for scoring the database output.

Mascot uses the probability-based Mowse scoring algorithm, which yields a score based on the probability that the top hit is a random event. Given an absolute probability that the top match is random, and knowing the size of the sequence database being searched, the engine calculates an objective measure of the significance of the result. The X!Tandem algorithm similarly calculates an expectation value that the top match for a spectrum is a random event. The SEQUEST algorithm identifies aminoacid sequences in the database that match the measured mass of the Peptide and comparing fragment ions against the MS/MS spectrum it generates a preliminary score for each amino acid sequence. A cross correlation analysis is then performed on the top 500 preliminary scoring peptides by correlating theoretical, reconstructed spectra against the experimental spectrum and assigns a cross correlation score to any peptide match (Xcorr). The PetideProphet model the distribution of scores as a mixture of two populations, correct and incorrect assignments. This requires that the search scores allow significant separation of the two distributions. Then, based on the mixture model, probabilities of correctness to all identifications can be calculated, together with global false discovery rate (FDR) for a given probability threshold.

Other types of protein search engines called spectral search engines such as BiblioSpec, Bonanza, SpectraST , X! Hunter, match new observed spectra with a library of consensus spectra derived from previous identifications. In some cases the spectral search engine is more efficient than sequence search engines, but databases are limited to only previously observed peptides.

De novo sequencing is another approach for protein identification based on search engine capable of producing protein sequences directly from the spacing of the peaks. The analytical approach provides short sequences (tags) of a few consecutive amino acids that can be determined *de novo*, and then use these tags to drastically limit the search space for a conventional sequence search. In this way, only peptides that fall within the correct mass range and contain one of the possible tags determined from the de novo part will be considered. However, this is a labour intensive approach requiring high quality and high resolution MS/MS spectra, of particular interest when a specific protein with no available sequences is searched.

Software assisted protein mass spectrometry

Protein mass spectrometry analysis is a quite sophisticated technique, involving critical aspects in protein tryptic hydrolysis, peptide gel or chromatographic separation, peptide ionization and ion separation and detection. This methodological approach is valid if formation of artifacts due to sample low quality, low separation efficiency and bad mass spectrometer performance can be excluded. For this reason, parallel to evolution of protein MS, several software have been developed for the analysis of the workflow analysis of MS data by signals produced by the detectors, including the file conversion, statistical analysis and protein identification. An example of such software is the Trans Proteomic Pipeline which encompasses most of the steps in protein data analysis in a single integrated software, from mass spectrometer output file conversion to protein-level statistical validation (Deutsch et al. 2010). Among the TPP routines, the Pep 3D tool allows the visualizing of the LC/MS data in a two dimensional

density plot (retention time vs m/z), in which each spot represents a peptide. This tool allows to visualize the precursor ions selected for collision-induced dissociation (CID), and successfully identified peptides. In mass spectrometers using a reversed-phase microcapillary LC system is coupled online with an electrospray ionization (ESI) tandem mass spectrometer, the instrument automatically selects specific precursor ion(s) for fragmentation in a collision cell, thus generating collision-induced dissociation (CID) or tandem mass (MS/MS) spectra. These CID spectra are then searched against sequence databases using software tools (e.g. MASCOT) which allows the identification of peptide sequence.

1.5.3 Analytical techniques for soil extracted proteins and soil proteomics

Techniques used for the analysis of protein extracted from soil and for soil proteomics have seen a strong technological and methodological evolution in the last decade. Methods for direct protein extraction from soils are based on the use of various lysis buffers containing protease inhibitor cocktails, NaOH, surfactants (e.g. SDS), phenol and chelating agents (e.g. EDTA), with or without sonication, autoclaving and freeze-thawing cycles (Tabs 1.1, 1.2). A review of methods for direct extraction of whole soil proteins has been published by Keller and Hettich (2009).

Early soil proteomic studies have been carried out using mainly monodimensional electrophoresis or 2-DE approaches, based on the detection of the a differential protein expression by soil microbial communities. However, although the early proteomic studies have provided pioneer information on the presence of specific proteins in soil, very little information was gained about the complexity of the entire soil protein complement, as high throughput protein analytical technologies and identification tools were not available. This type of information was obtained thanks to the use of protein MS analysis. Schulze et al. (2005) presented the first comprehensive environmental proteomic approach by the extracting proteins from

the soil dissolved organic matter and from clay minerals of a forest soil. After protein purification by gel filtration to remove humic acids phenolic compounds and protein separation by SDS-PAGE, proteins were digested with trypsin, and the trypic peptides separated by nanoflow liquid chromatography and analysed by mass spectrometry. Proteins involved in ribosomal transcription, membrane proteins and enzymes of plants, animals and soil microorganism were detected. To date the study by Schulze et al. (2005) is the only soil proteomic study where enzymes involved in SOM mineralization such as cellulases and laccases were identified as associated to soil solid phases.

Table.1.1 Soil and environmental (meta)proteomic studies in chronological order, with main extraction conditions and outcome

Reference	Aim	Analytical conditions	Outcome	Comment
Ogunseitan (1993)	Extraction and characterization of whole soil proteins.	Two extraction methods: 1) samples in boiling lysis buffer (50 mM Tris-HCl pH 6.8, 100 mM dithiothreitol, 2% SDS, 10% glycerol, and 0.2% bromophenol blue) 2) incubation at at 0°C in a lysis buffer (50 mM Tris-HCl pH 7.6, 1 mM EDTA, 10% sucrose, 1 mM dithiothreitol, 300 µg/ml lysozyme, 0.1% polyoxyethylene 20 acetyl ether), followed by four freeze-thaw cycles. Proteins characterizationby SDS-PAGE.	The freeze-thaw method yielded more proteins (20 -50 µg $^{g-1}$). Protein size obtained by both methods ranged from less than 14 kDa to greater than 97 kDa.	The first paper on soil and sediment metaproteomic. Both the tested extraction methods were suggested as suitable for soil and environmental metaproteomics.Interferences by phenolic compounds on soil-extracted proteins that make impossible a quantitative comparison between extraction methods were demonstrated in later papers (Roberts and Jones, 2008).
Wright and Upadhyaya (1996, 1998)	Arbuscular mycorrhizal glycoprotein (glomalin) extraction from twelve soils	Autoclaving + extraction by citrate buffer or malate buffer (pH 7.0 or pH 8.0). Protein characterization by ELISA.	Protein characterization by ELISA confirmed that the extracted protein was glomalin.	In a later paper on Schindler et al. (2007) showed the possible formation of artifacts

Ogunseitan (1997)	Study of extracellular enzymes produced by microbial communities of freshwater and wastewater sludge and from the respective culturable microbial communities and by a *P. putida* strain.	Proteins from environmental matrices, culturable microorganisms and *P. putida* were extracted by pulsed sonication on ice from 3 min. Soluble proteins recovery by centrifugation (25 min 25000 *g*). Protein immobilization onto nitrocellulose filters (0.2 µm) pre-soaked by the appropriate buffer and detection of enzyme activity by chromogen substrates and native-PAGE.	The protein extraction protocol permitted the detection of catechol oxidase, nitrate reductase, peroxidase, and xanthine dehydrogenase activities and the differentiation of extracellular activities between sludge and freshwater samples.	A similar conservative approach has not been sufficiently tested in soil proteomics. The comparison with control experiments based on the measurement of enzyme activity of a *P. putida* strain, shows the detection limits in soil proteomics. The xanthine dehydrogenase activity was undetectable at population density below 10^3 CFU, and below 3.3 ng of total protein. The importance of sample amount for environmental proteomics was discussed later by Thompson et al (2008).
Craig and Collins (2000)	Immuno-detection of BSA adsorbed to siliceous minerals (DACIA) using 3 different buffer (Urea, PBS and HF)	The lysis buffer contained 10 mM PBS (pH 7.4), 0.87% NaCl, 8M urea (pH 8.5) and 4M HF. Samples were incubated at 4°C under shaking for 24h, followed by a washing step with 0.1 M carbonate-bicarbonate (pH 9.6).	Using DACIA immunological test, HF resulted a suitable extractant for mineral-sorbed proteins	The use of BSA as test protein is questionable due to its inherent instability. Irreversible protein denaturation due to acidic pH values of the extractant should be considered for soil proteomic studies.
Craig and Collins (2002)	Characterization of BSA in the presence of quartz and illite and using 3 different buffers	The lysis buffer contained: 2% SDS (pH 6.4), 6 M guanidine-HCl (pH 5.6), 8 M urea, 10 mM PBS, 0.149 M NaCl pH (7.4), 5% ammonia (pH 11.5), 10% K_2-EDTA (pH 10.0).Protein characterization by DACIA and	Proteins were tightly bound to mineral surfaces via short-range bonds. The extraction efficiency was not satisfactory.	Use of immunological test (DACIA) for identifying mineral bound protein, after digestion of the mineral phase with HF was proposed.

		spectrofometric analysis		
Singleton et al. (2003)	Metaproteomic study of Cd-polluted soils.	Bead-beating or freezing-thawing cycles, Tris-HCl buffer + protease inhibitors, sucrose, dithiothreitol and EDTA. Protein characterization by SDS-PAGE.	Soil Cd pollution decreased the content of extracted proteins and induced the synthesis of some low molecular weight proteins.	Low number of protein bands on SDS-PAGE in spite of the theoretical soil metaproteome and the used harsh method for disgregating soil matrix. No specific proteins could be identified.
Murase et al. (2003)	Extraction of extracellular proteins from a greenhouse soil	The lysis buffer contained 67 mM phosphate buffer at pH values 6.0, 7.0 and 7.7, and pH 7.7+NaN_3. Filtration by filter paper and 0.2μm cellulose acetate filters. TCA and ethanol were used for protein precipitation and characterization by SDS-PAGE.	The best results were obtained using buffer at pH 6.0 but a low amount of protein was detected on gel. Impurities were present in protein SDS-PAGE.	The amount of protein extracted was very low in relation with the theoretical soil protein content, likely due to the weakness of the used extractant. Information on other proteins were not provided.
Wilmes and Bond, 2004	Indirect proteomic analysis of microbial communities extracted from activated sludge	Microbial communities extraction by centrifugation. Protein extraction by 7 M urea, 2 M thiourea, 4% (w/v) CHAPS, 10 mM Tris-1 mM EDTA, 50 mM DTT, 25 mM PefablocSC and 2 mM Pefabloc protector. Protein precipitation by TCA. Protein fingerprinting by 2-DE, MALDI-TOF	Specific functional proteins were identified	The used de novo sequencing approach helped to identify the functional proteins of the extracted microbial communities
Schulze et al. (2005)	Off gel proteomic fingerprint of dissolved organic matter (DOM)	Flotation of DOM with Na polytungstate + washing with H_2O, 10% HF + gel filtration using Sepharose 4B. Protein analysis by SDS-PAGE and LC/MS/MS	One hundred forty-eight proteins isolated by SDS-PAGE. Graphics showed the different percentage of identified proteins belonging to organisms from different kingdoms.	No gel image was reported. No identified protein list with related identification parameters was shown.
Benndorf et al. (2007)	Enrichment of soil with *C. necator* JMP134,	Protein extraction by 0.1 M NaOH+phenol, protein precipitation by 0.1M CH_3COONH_4 in	Four proteins of *C. necator* JMP134 were identified by	Low protein recovery particularly from soil and low number

	Rhodoferax sp. P230 and *S. herbicidovorans* B488.	CH_3OH. Protein extraction and characterization by SDS-PAGE,2D-E and MS/MS.	SDS-PAGE. Proteomic analysis of groundwater by SDS-PAGE and 2-DE showed 20 bands and ca. 100 spots, respectively. Twenty-nine proteins of 19 SDS-PAGE bands and 26 proteins of 50 2-DE spots were identified by MS.	of identified proteins. SDS-PAGE gel showed impurities on gel.
Maron et al. (2008)	Protein fingerprint of soil microbial communities of 3 soils coupled to soil microbial communities characterization.	Extraction of soil microrganisms by centrifugation density gradient, resuspension in NaCl and further separation by Nycodenz density gradient. Protein extraction by ultrasonic treatment and lysis buffer. Protein characterization by SDS-PAGE and microbial characterization by ARISA,	This was the first attempt to combine between genetic and proteomic approaches for the study of soil functionality.	Large number of protein bands detected on SDS-PAGE but no proteins were identified.
Benndorf et al. (2009)	Proteomic study of river sediments and lava granules	Sonication for 10 min, extraction with 20 mM Tris-HCl pH 7.5. 2-DE analysis.		No protein identification.
Chen et al. (2009)	Comparison between different direct soil protein extraction methods.	Sequential extraction with citrate + SDS buffers, followed by phenol extraction. SDS-PAGE and 2-DE protein characterization.	The protocol with SDS buffer was more efficient.	No protein identification.
Williams et al. (2010).	Metaproteomic and metagenomics study of soil microbial communities in toluene-	Protein extraction of soil microorganisms, separated from soil by density gradient centrifugation(Nycodenz), using the lysis buffer(0.5 M TRIS-HCl (pH 8.7), 0.9 M sucrose,	Fourty-seven proteins were identified; several proteins were common between toluene-amended soil communities and toluene impacted bacteria,	One of the few integrated approaches for the study of soil microbial diversity and functional responses to specific stressors. The used proteomic

	amended soil, and toluene-amended soil inoculated with microbial cultures.	0.05 M EDTA, 0.1M KCl, and 2% 2-mercaptoethanol). Phenol treatment (0.5 mL phenol for several phase inversion separation). Protein precipitation by 0.1 M CH3COONH4/ 1% β-mercaptoethanol in methanol and 80% acetone. Proteins analysis by SDS-PAGE - MALDI-TOF/TOF MS, using NCBInr in MASCOT.	mostly stress related proteins, analyzed by fatty acid methyl ester analysis profiles. The 16S rRNA gene analysis indicated high dominance with 80% of the OTUs related to the *Bacillus* genus.	approach highlighted the type of stress imposed by toluene to soil microbial communities.
Chourey et al. (2010).	Off gel soil proteomic analysis by direct and indirect extraction *P. putida* was added to soil as internal standard strain.	Three extraction protocols: 1) SDS-TCA lysis buffer (5% SDS, 50 mM Tris-HCl, pH 8.5, 0.15 M NaCl, 0.1 mM EDTA, 1 mM MgCl2, 50 mM DTT). Protein precipitation by 100% TCA and acetone. Protein pellet in guanidine buffer (6 M guanidine-HCl, 10 mM DTT in Tris-$CaCl_2$ buffer at 60°C for 1 h. Proteins trypsin digestion and 2) Guanidine-HCl lysis buffer. Protein digestion by double treatment with trypsin, protein reduction using 20 mM DTT. 3) Isolation of soil microorganisms by centrifugation with cold PBS 10.000 *g* for 10 min at 4 °C. Protein extraction by SDS-TCA or guanidine lysis buffers. Protein characterization by 2D-LC-MS/MS. Protein identification using SEQUEST.	One thousand forty-three identified protein of *P. putida* in liquid culture, 925 identified protein of *P. putida* extracted from soil. The approach showed to be suitable for Gram-negative and Gram-positive bacteria, as direct soil protein resulted in identification of more than 500 unique proteins with no apparent bias in terms of protein size, localization, functions model strains grown in pure cultures.	First soil metaproteomic study using 2D-LC-MS/MS Relatively low protein identification rates from soil and reduced protein identification of the inoculated *P. putida* strain. Large amount of unknown proteins from using the direct extraction approach.
Bastida et al. (2010).	Microbial proteomic responses	Soil microbial communities were extracted by phosphate	Nineteen proteins belonging to *B. cereus, B.*	Relatively low number of proteins were idetified, and

	to soil pollution with hydrocarbons.	buffer, inoculated and cultured onto LB medium; then they were lysed with 20 mM Tris-HCl (pH 7.5) plus 0.2 g L^{-1} SDS and also using the method proposed by Benndorf et al. (2007). Protein analysis by SDS-PAGE and LC/MS analysis.	*thuringiensis, B. Anthracis* were identified.	SDS-PAGE gel image were not shown. The main conclusion was that *Bacillus sp.* dominated the soil microbial communities of hydrocarbon polluted soil could be biased by the intermediated microbial cultural step
Wang et al. (2011).	Metaproteomic analysis of crop soil Soil microbial composition analysis by TRLFP.	Two protein extraction protocols: 1) SDS 1.25% w:v, 1M Tris-HCl, pH 6.8, 20 mM DTT+ buffered (pH 8) phenol 2) 0.25M citrate buffer (pH 8) + buffered (pH 8) phenol. Proteins were precipitated by CH_3COONH_4 in methanol. Protein analysis by SDS-PAGE, 2DE and MS/MS.	Two-hundred eighty protein randomly selected were analyzed by MS: 122 protein were identified. Correspondence between protein MS/MS results and TRLFP results.	First report matching soil proteomic and genomic data. The extraction protocol with SDS buffer was more efficient than that with citrate buffer. Several proteins were still not identified
Keiblinger et al. (2012)	Comparison of four different soil protein extraction method	Pretreatment with PVPP. Four protein extraction protocols were compared: 1) 50 mM Tris, 1% SDS pH 7.5 and 10% TCA. 2) NaOH, phenol buffer and 0.1 M ammonium acetate 3) 50 mM Tris, 1% SDS pH 7.5 and phenol pH 8.0. 4) pretreatment with TCA and Methanol. Then 50 mM Tris, 1% SDS pH 7.5 and phenol (pH 8.0).	The method with SDS and phenol were the two most efficient for the studied soils	The protein extraction efficiency from soil samples is severely hampered by the complex matrix. The SDS-phenol was showed to be the best extractant.

Table. 1.2 Sample of soil and environmental model proteomic studies in chronological order, with main extraction conditions and outcome

Reference	**Aim**	**Analytical conditions**	**Outcome**	**Comment**
Luo et al. (2007)	Proteomic study of *B. cereus* grown in soil organic matter extracts.	The *B.cereus* proteome was analysed after growth in soil organic matter and in LB medium until mid exponential phase were analyzed. Cells were resuspended in urea CHAPS buffer and lysed by freeze-thaw cycles and pulsed sonication. Protein characterization by 2-DE and MALDI-TOF MS (*B. cereus* sequence database).	The 2-DE analysis revealed 234 upregulated proteins and 201 down regulated proteins of *B. cereus* grown in soil organic matter extracts. Thirty-five proteins were overexpressed and eight proteins were underexpressed after *B. cereus* growth in soil organic matter, all involved in the fundamental cellular metabolism.	The first model approach for microbial proteomics related to a specific soil component. The study revealed that the studied organic matter was relatively poor in nutrients.
Taylor and Williams (2009)	Comparison of direct and indirect protein extraction method from soil after inoculation with *S. maltophilia* and BSA to soil as protein internal standard.	Direct method: soil proteins extracted directly from soil. Indirect method: microorganisms extracted from soil by density gradient centrifugation. Protein extraction for both methods: 0.5 M Tris–HCl (pH 8.7), 0.9 M sucrose, 0.05 M EDTA, 0.1 M KCl, and 2% 2-mercaptoethanol, phenol. Protein precipitation: 0.1M ammonium acetate and 1% β-mercaptoethanol in methanol and 80% acetone. Protein characterization by SDS-PAGE.	SDS-PAGE of proteins using indirect method was better than direct method. BSA as control was evident.	Direct extraction of protein from soil was affected by organic matter and clay particles. Low amount of protein and unclear SDS-PAGE patterns from direct protein extraction from soil were obtained and no protein identification was provided.

Schneider et al. (2010).	Proteomic analysis of bacterial and fungal species involved in litter decomposition. *P. carotovorum* and *A. nidulans* were used as bacterial and fungal model organisms.	Enzymes in the secretome were extracted by ultrafiltration (10 kDa cutoff membrane) and analyzed by SDS-PAGE. Proteins were subjected to in-gel tryptic digestion and analysed by MS/MS. Protein identification was done by using the MASCOT algorithm.	The *P. carotovorum* grew better in co-culture with the fungus. The bacterium showed a limited litter decomposition capability but profited by the fungal degradation products.	Proteomic analysis revealed the differences in the regulation and in the production levels of degradative enzymes in the secretome of the studied bacterial and fungal strains and their ecological role in litter decomposition.
Giagnoni et al. (2011)	Proteomic analysis of *C. metallidurans* in artificial soils containing quartz sand, kaolinite, montomorillonite, goethite and humic acids, singly or mixed for the evaluation of the effects of selected soil solid phases on the bacterial proteome.	The bacterial proteome was extracted by PBS-SDS lysis buffer containing nuclease and protease inhibitor cocktail. Proteome analysis was done by 2-DE and MALDI-TOF (MASCOT algorithm in NCBInr).	The presence of highly reactive clays affected the analysis of the bacterial proteome as less proteins were recovered in the artificial soils as compared to the sand and kaolinites substrates. No proteins could be recovered by microcosms containing sole montmorillonite.	Microbial proteome analysis in soils is influenced by the reactivity of the soil solid phases, by reducing the protein extraction efficiency.
Giagnoni et al. (2012).	Proteomic analysis of *C. metallidurans* CH34 incubated in an artificial soil at different contact times.	The bacterial proteome was extracted by PBS-SDS lysis buffer containing nuclease and protease inhibitor cocktail. Proteome analysis was done by 2-DE and MS (MASCOT algorithm in NCBInr).	The number of protein spots in 2-DE was reduced during the incubation time. Apparently, a large fraction of the bacterial proteome was stabilized by soil solid phases after the bacterial death.	The paper underlined the importance of protein extracellular stabilization for soil proteomics. Relatively low rates of protein identification was reported, maybe due to interactions between proteins and humicsubstances.

Benndorf et al. (2007, 2009) charaterized the regulation of a 1,2-dioxygenase and other proteins involved in microbial degradation of dichlorophenoxy acetic acid and benzene in contaminated soils, sediments and aquifers. Williams et al. (2010) performed a metaproteomic study in toluene-amended soil, and toluene-amended soil inoculated with microbial cultures using an SDS-PAGE - MALDI-TOF/TOF MS approach, and identified fourty-seven proteins, with several of them being in common between toluene-amended soil communities and toluene impacted bacteria. Chourey et al. (2010) conducted the first off-gel soil metaproteomic study using 2D-LC-MS/MS, including a comparison with proteins extracted from a *P. putida* liquid culture. The used approach led to the identification of more than 500 unique proteins with no apparent bias in terms of protein size, localization, functions as compared to proteins extracted from the *P. putida* strain grown in pure culture.

In this PhD project, MS and NMR techniques have been used to study the interactions between specific proteins and soil-borne HS.

1.5.4 Analysis of SOM and HS

The SOM chemical composition and molecular structure has been traditionally studied by physical and chemical fractionation methods. Physical fractionation methods aim at assessing spatial arrangement of inorganic phases and SOM, to understand the SOM mobility and bioaccessibility and derive information on potential SOM decomposition. Physical fractionation involves the application of various degrees of disaggregating treatments (dry and wet sieving, slaking), dispersion (ultrasonic vibration in water), density separation and sedimentation; it aims to separate SOM aggregates.

Chemical fractionation procedures are based on the extraction of SOM in H_2O, alkali or saline solutions, in organic solvents, or methods based on acid solution to evaluate the SOM resistance to hydrolysis and oxidation. Chemical extractions

are advantageous because provide purified SOM and SOM fractions, free of mineral components.

Solubility in water is considered an important SOM property because it indicates the bioavailability (Marschner and Kalbitz, 2003), and the HS are operationally defined and based on their solubility properties. The most common extraction method for HS is based on the use of NaOH and $Na_4P_2O_7$ solutions capable of extracting large quantities of HS from soils (Stevenson, 1994), with variable extraction efficiency depending on the soil type (Olk, 2006). Humic acids (HA) have usually little solubility in water, and solubilisation is possible in alkaline solutions. The 0.1 M $Na_4P_2O_7$ is the typical extractant for HS. The NaOH is a stronger extractant but may induce SOM auto-oxidation during the extraction. Moreover, during NaOH extraction, H^+-bridges within SOM are replaced by Na^+, causing SOM solubilization and also a rearrangement of organic associations (Piccolo 2002). SOM desorption is further promoted through competition of the OH^- anions with the anionic groups of SOM for adsorption sites at mineral surfaces and through pH-induced changes in functional group dissociation. Polyvalent cation bridges between SOM and soil minerals are not affected by NaOH. They are only disrupted by extraction with $Na_4P_2O_7$ (Schnitzer 1978).

Extraction with $Na_4P_2O_7$ at pH 7 was proposed to obtain the fraction of SOM bound to clay minerals by metal bridges and SOM complexed by polyvalent cations, without removing Fe and Al from soil, forming parent materials (Alexandrowa 1960). In comparative studies, it was reported that NaOH extracted up to 80% of SOM, while extraction with $Na_4P_2O_7$ was usually less efficient, with up to 30% of SOM (Kononova 1966; Stevenson 1994). Extraction with 0.5M NaOH removed more C and N from coarser fractions than from fine fractions. Therefore, Schnitzer and Schuppli (1989) proposed 0.5M NaOH as a method to remove more 'free' SOM. Cameron et al. (1972) and Piccolo and Mirabella (1987) suggested that NaOH extracts more high-molecular weight SOM than does $Na_4P_2O_7$ solution at pH 7, while $Na_4P_2O_7$ extraction removed more SOM from

finer soil particles, indicating that $Na_4P_2O_7$ extracts 'complexed OM' (Schnitzer and Schuppli 1989). For extraction of humic substances, most often a mixture of e.g. 0.1M NaOH + 0.1M $Na_4P_2O_7$ was used.

Moreover, the SOM extracted with NaOH is not a homogeneous soil fraction, because the extraction procedure simultaneously affects organo–mineral and organo–organo interactions. Na^+ ions also interfere with the flocculation of clays, causing disaggregation. Compared to NaOH extracts, sequential extraction with NaOH and $Na_4P_2O_7$ seems to additionally separate SOM stabilized by polyvalent cation bridges on mineral surfaces or in clay microstructures (< 20 µm). The pyrophosphate treatment gives no indication of the metals involved in the bridging effect, and the amount of pyrophosphate required to quantitatively displace polyanionic SOM from oxide surfaces or from associations with highly amorphous minerals is not clear (Oades 1984). In summary, alkaline extraction procedures simultaneously affect several stabilization mechanisms that are of different relevance in different soils and soil horizons. Systematic investigations with soils of different texture and mineralogy would be necessary to investigate the influence of the extraction procedure on the amounts of SOM and to differentiate between the stabilization mechanisms involved.

Soil organic matter contains HS non-humic substances produced by plants and microbes. The HS are fractionated into fulvic acid (FA), humic acid (HA), humin (H) and after total SOM extraction (Baldock and Nelson 2000). As shown in the fractionation scheme (Figure 1.3), the FA fraction is soluble in alkali (e.g. 0.1M NaOH + 0.1M $Na_4P_2O_7$) and in mineral acids (e.g. HCl), the HA fraction is soluble in alkali, not insoluble in acids and the H fraction is insoluble in water at any pH (Stevenson 1994).

Electrophoretical and spectroscopic techniques have been used for studying specific proteins or specific aspects of proteins extracted from soil.

The pyrolysis-mass spectrometry is among the most used chemical methods for SOM studies, either by direct pyrolysis , which produced molecular ions through

sof ionization, or indirect pyrolysis (e.g. Curie-point pyrolysis) coupled with gas chromatography followed by mass spectrometry (Py-GC/MS), based on the use of library search for the identification of pyrolysis products due to electron impact ionization. The pyrolysis analysis provide information on the HS structure and carbohydrates, branched phenols, lipids, aliphatics and nitrogenous compounds precursors of the HS and their HA and FA fractions (Schulten and Schnitzer 1992). The pyrolysis analysis has allowed the contruction of the first molecular models of the HS and the HA and FA fractions, in which the aromatic rings are cross linked by aliphatic chains in a molecular network containing voids (Schulten et al. 1991). Such network has been suggested to shape it macromolecular structure

Depending on it chemical composition and solution chemistry in a way that HS form globular structures at higher ionic strength (Myneni et al. 1991), and structural voids which can be occupied by small organic and inorganic molecules.

This classical structural model as been also confirmed by electron microscopy (Stevenson and Schnitzer 1982; Myneni et al. 1991) and NMR and X-ray spectrmetry studies (Schnitzer et al. 1991).

The nuclear magnetic resoncance (NMR) technique has been widely applied for the characterization of Hs and its fractions since the mid 1990s, either using liquid- or solid-state ^{13}C NMR, which indicated the formation of various molecular arrangements, also in combination with mass spectrometric methods.

The NMR has been used for studying the interactions between proteins and HS, relying on the known phenomenon of the reduction of Brownian motion of proteins following the complexation with HS, which can be revealed by monodimensional ^{1}H experiments. The molecular size of the humic aggregates, implies signal broadening effects because of a less efficient minimization of molecular dipolar couplings. When weak interactions occur, molecule mobility decreases with a significative reduction in relaxation time. This phenomenon leads to a broadering of the NMR signal. In addition, the change of overall

magnetic field experienced by the observed nuclei is univocally shown by a drift of chemical shifts. For instance, hydrophobic interactions may exert an up-shift effect due to the increase of electron density surrounding the observed nucleus, while electron-attractive nuclei involved in hydrogen bonds, can exert a de-shielding effect and a consequent downshift drift.

Of relevant importance for the aims of this thesis, is that the HS models based on pyrolysis and NMR studies do not provide information on the type of interaction between HS and proteins, although it was reported that protein-derived moieties in HS could account for 10% of HS (Schulten and Schnitzer 1992).

Fan et al. 2000 provides evidences for the presence of peptidic structures in HS; they studied organic matter structural motifs by pyrolysis-GCMS and FT-IR spectroscopy and HS-ligand interaction. The authors found that the abundance order of aminoacids in the studied HS was atypical for proteins in live organisms, and they hypothesized this deviation from the general protein composition could be due to a differential degradation of protein components during the diagenesis of the HS. Fan et al. (2004) confirmed the previous fiundings analyzing HS labeled with ^{13}C and ^{15}N by NMR, CD and pyrolysis gas chromatography-mass spectrometry (pyro-GCMS). They got evidence for random-coil peptidic structures in the labeled HS from 2-D nuclear magnetic resonance (NMR), pyrolysis gas chromatography-mass spectrometry (pyro-GCMS), and circular dichroism data; and, moreover, the detailed ^{13}C and ^{15}N labeling patterns of amino acid residues in the acid hydrolysates of HS acquired from NMR and GC-MS revealed two pools of peptides: one extant (unlabeled) and the other, newly humified with little isotopic scrambling (fully labeled). The authors concluded that the persistence of pre-existing peptidic structures indicates their resistance to degradation while the presence of fully labeled peptidic amino acids suggests wholesale incorporation of newly synthesized peptides into HS.

Tomaszewski et al. (2011) reported the encapsulation of lysozyme and ribonuclease A by HS using a quartz microbalance. Quartz microbalance

technique represent a piezoelectric biosensor, a direct method, label free, suitable to detect interaction between two partner molecules. In the present thesis the described technique has been used but, because of a false positive, gave no useful results.

Schindler et al. (2007) reported the presence of humic substances in the glomalin protein extracted from soil and analysed by NMR; possible protein degradation by HS was hypothesized by Solaiman et al. (2007). Complexation with high affinity between humic acids and specific domains human serum albumin were reported by Ding et al (2011).

Through studies of HS by infrared spectrometry Nanny and Ratasuk (2002), Calace and Petronio (1997), Artiola-Fortuny and Fuller (1982) reported no absorption band corresponding to C-O stretching in polysaccharides. Fan et al. (2004) used the circular dichroism technique to study the chemical properties of peptides after contact with soil borne humic substances.

1.5.5 A peculiarity of soil proteomics: missing proteins and low quality of identification

Despite the large progresses in protein detection technologies, low rate of protein identification and the low quality of protein identification parameters are nowadays major drawbacks in soil proteomics.

The theoretical complexity of the soil metaproteome can be exemplified by the fact that relatively simple prokaryotic and eukaryotic unicellular organisms such as *E. coli* and *S. cerevisiae* possess $4.3 \cdot 10^3$ and $6.1 \cdot 10^3$ open reading frames, respectively, coding for proteins with molecular weights between 1.7 and 559 kD, with a number of aminoacids in the order of 10^6. By considering the extreme microbial diversity of most soils, it is clear that no techniques are currently available for resolving the soil metaproteome. Moreover, some microbial species may be ubiquitous in soil whereas other may inhabit only specific soil niches such as aggregates and rhizosphere. Furthermore, microbial activity is generally

localized in hot spots characterized by sufficient availability of nutrients and H_2O. Such specific soil features require specific soil protein sampling procedures as compared to nucleic acids, because it is likely that soil organisms of the same species may differ for their metabolic activity and therefore for their responses at proteomic level.

One of the major limitations of the past soil proteomic studies is the lack of information on the microbial community composition, and on the presence of key microbial species with annotated proteomes in the analyzed soils. Although the soil metagenomic studies were launched since the 2000's (Rondon et al. 2000), joint soil genomic-proteomic research, which could provide indications on the protein extraction efficiency and recovery of specific functional proteins, were not conducted. In early proteomic studies, soil microbial communities were implicitly assumed to be evenly distributed and active, without considering the uneven distribution of soil microbial communities, with microbes localized in hot spots of such as soil aggregates and rhizosphere (Brimecombe et al. 2001), as above mentioned. Another important neglected factor is the typical species dominance within the soil microbial communities.

Another current limitation in soil proteomics is the relative poor information of proteomic databases, which is not comparable with the increasing genomic and transpcriptomic databases. A comparison of some available genomic and proteomic information in the NCBI databases is reported in Table 1.3.

Both bacterial and animal databases are more developed (King et al., 2006), than fungal and plant proteomic databases, because the latter have only become active research areas in recent years (Jorrin et al., 2006; Kim et al., 2007). A key aspect in development of soil proteomics, is that the annotation of proteomes of environmental strains, is lower than that of human- or animal-related microbiome (Table 1.3).

Table 1.3. Example of genomic and proteomic information available in NBCI and Expasy SwissProt database for living organisms of various kingdoms (October 2012), of specific interest for soil proteomics

	Genomes (NBCI)	**Proteins** (NBCI)	**Proteomes** (SwissProt)
	Total	**Entries**	**Total**
Eukaryote	$4.30 \cdot 10^{3}$	$1.10 \cdot 10^{7}$	$3.05 \cdot 10^{2}$
Opisthokonta	$3.47 \cdot 10^{3}$	$7.70 \cdot 10^{6}$	$2.37 \cdot 10^{2}$
Metazoa	$4.28 \cdot 10^{3}$	$5.52 \cdot 10^{6}$	$8.50 \cdot 10^{1}$
• Mesozoa	None	$8.70 \cdot 10^{-1}$	none
• Eumetazoa	$3.01 \cdot 10^{3}$	$5.23 \cdot 10^{6}$	$8.30 \cdot 10^{1}$
○ *Bilateria*	$3.02 \cdot 10^{2}$	$5.18 \cdot 10^{6}$	$8.20 \cdot 10^{1}$
• Nematoda	$7.1 \cdot 10^{1}$	$2.97 \cdot 10^{5}$	9
• Arthropoda	$4.85 \cdot 10^{2}$	$1.66 \cdot 10^{6}$	3
• Anellida	9	$17.67 \cdot 10^{3}$	none
Fungi	$3.56 \cdot 10^{2}$	$2.49 \cdot 10^{6}$	$1.49 \cdot 10^{2}$
• Dikarya	$3.17 \cdot 10^{2}$	$2.28 \cdot 10^{6}$	$1.40 \cdot 10^{2}$
○ *Ascomycota*	$2.50 \cdot 10^{2}$	$1.90 \cdot 10^{6}$	$1.18 \cdot 10^{2}$
○ *Basidiomycota*	$6.70 \cdot 10^{1}$	$5.08 \cdot 10^{5}$	$2.20 \cdot 10^{1}$
Green plants (Viridiplantae)	$4.55 \cdot 10^{2}$	$2.23 \cdot 10^{6}$	$2.00 \cdot 10^{1}$
Green algae (Chlorophyta)	$9.8 \cdot 10^{1}$	$1.68 \cdot 10^{5}$	7
Amoebozoa	$2.20 \cdot 10^{1}$	$1.20 \cdot 10^{5}$	6
Bacteria (eubacteria)	$5.50 \cdot 10^{3}$	$4.37 \cdot 10^{5}$	$1.58 \cdot 10^{3}$
Proteobacteria	$2.47 \cdot 10^{3}$	$2.29 \cdot 10^{5}$	$7.30 \cdot 10^{2}$
Alphaproteobacteria	$5.63 \cdot 10^{2}$	$2.81 \cdot 10^{6}$	$1.67 \cdot 10^{2}$
Bproteobacteria	$3.76 \cdot 10^{2}$	$2.25 \cdot 10^{6}$	$1.12 \cdot 10^{2}$
Gammaproteobacteria	$1.33 \cdot 10^{3}$	$9.90 \cdot 10^{6}$	$3.46 \cdot 10^{2}$
Firmicutes	$1.47 \cdot 10^{3}$	$7.84 \cdot 10^{6}$	$3.40 \cdot 10^{2}$
Cyanobacteria (blue-green algae)	$1.45 \cdot 10^{2}$	$6.56 \cdot 10^{6}$	$4.10 \cdot 10^{1}$
Fusobacteria	$3.6 \cdot 10^{1}$	$1.59 \cdot 10^{5}$	5
Actinobacteria	$4.83 \cdot 10^{2}$	$3.77 \cdot 10^{6}$	$1.73 \cdot 10^{2}$
Nitrospirae	4	$2.87 \cdot 10^{4}$	2
Archaea	$2.43 \cdot 10^{2}$	$1.25 \cdot 10^{3}$	$1.23 \cdot 10^{2}$
Viruses	$3.96 \cdot 10^{3}$	$1.17 \cdot 10^{6}$	$1.17 \cdot 10^{3}$
Plastids	$1.05 \cdot 10^{5}$ (nucleotides)	$4.83 \cdot 10^{3}$	None
Mitocondria	9	$9.32 \cdot 10^{5}$	1

In addition, the current rate of protein and proteome annotation of protozoa and earthworms in the databases is also low, and does not allow the proteomic study of soil fauna. Therefore, the present poor database does allow a limited proteomic study of soil microorganisms and soil fauna.

The proteomic databases can be used for the identification of specific functional proteins and for the identification of soil microorganisms. As one protein is theoretically sufficient for organism identification via databases (Aebersold and Mann, 2003), soil metagenomics should theoretically precede soil proteomic to first reconstruction of the main metabolic pathways present and potentially active in the studied soils (Torsvik et al., 2002). In this way proteomics could assess what functions are actually expressed by the soil microbial communities and their reflections at ecosystem level. From this point of view, the strain resolution level approach followed by Luo et al. (2007) for the study of the microbial community of a mine drainage environment, where the genomic data were used to identify proteins from dominant community members coupled with multidimensional proteomics, appears as the most promising to be applied to more complex microbial communities, like those of soils. The proteomic responses of known microorganisms with annotated proteome may be useful to increase the yields of whole soil protein extraction and provide new information on the relationship between microbial diversity and soil functions. Quantitative (sub)proteomic approaches based on stable isotope metabolic tagging using ^{13}C-glucose or, $^{15}NH_4^+$ or ^{13}C-labelled amino acids (Gygi et al. 1999; Ong and Mann 2006), followed by separation and identification of the labeled proteins by LC-MS/MS, can be also helpful in detecting synthesized proteins, although expensive and labor intensive. Stable isotope probing may also reveal low-abundance proteins and distinguish between low expression rates from low recovery. Targeted proteomics may be a suitable technique for detecting and quantifying proteins in soils since it can detect specific peptides or protein fragments of relevant interest for soil studies (e.g. stress biomarker protein, nutrient hydrolyzing enzymes). In addition, targeted proteomics is a quantitative approach as target stable isotope

labeled specific peptides can be prepared and added as internal standards to the soil extracts. It must be also noted that of the papers quoted in Table 1, protein identification parameters were generally not published before 2007 (Benndorf et al. 2007), thus making the obtained information on the potentials of the adopted analytical approaches of limited interest.

In the work of Schultze et al. (2005) soil proteins were identified by LC-MS/MS analysis. Mass spectra were searched against NBCI database, using the following search parameters: maximum of one missed trypsin cleavage, cysteine carbamidomethylation, methionine oxidation, and a maximum 0.2 Da error tolerance in both the MS and MS/MS data (40 ppm after dynamic recalibration). This setup brought to the identification of just 75 proteins and less than 50% was of bacterial origin and on average, 30% of the proteins were identified by a single tryptic peptide. Bendorf et al. (2007) identified only 4 proteins conducting a metaproteomic analyses on soil microcosms to investigate the degradation of 2,4-dichlorophenoxy acetic acid (2,4-D). Protein identification was carried out using nano-LC-electrospray ionization (ESI) source-MS/MS, and searching the resulting mass spectra against the NCBI database, no parameters are reported about the search in the database.

In the work of Chourey et al. (2010), through a complex off gel approach identified 716 proteins from soils inoculated with genome-sequenced bacteria. The protein search was performed using the *P. putida* genomic database, and the MS/MS spectra were searched via SEQUEST using the following parameters: system parent mass tolerance, 3.0; fragment ion tolerance, 0.5; up to four missed cleavages allowed; fully tryptic peptides only. Common contaminants such as trypsin, keratin, etc. were also included. Wang et al. (2011) carried out a metaproteomic analysis of crop rhizospheric soil separating proteins by 2D electrophoresis, 287 spots were analyzed by MALDI-TOF/TOF followed by MS/MS. Spectra were searched against the database of the National Center for Biotechnology Information (NCBI) in SwissProt

(http://www.matrixscience.com/search_form_select.html), applying the following parameters: one missed cleavage site, carbamidomethyl as fixed modification of cysteine and oxidation of methionine as a variable modification, MS tolerance of 100 ppm, MS/MS tolerance of 0.6 Da. Authors specify that known contaminant ions, such as keratin, were excluded. This is an important technical clarification, because keratins represent a class of proteins that are identified by mass spectrometry in the majority of biological samples. Their abundance can overwhelm the analysis capacity of the LC-MS system and obscure the peptides of interest because the mass spectrometer usually prefers detecting precursor ions with higher intensity (Xiao et al. 2003). In the case of soil protein samples, if low extraction yields occur, this phenomenon can be particularly problematic when performing analysis as MS/MS, it will focus on the peptides from the more abundant keratins, providing less or no information about the proteins of interest.

Wu et al. (2011) reported that only six bacterial proteins and one fungal protein were differentially expressed in their *R. glutinosa* rhizosphere proteomic study. For example, Wu et al (2011) identified 26 plant proteins from a rhizosphere metaproteome experiment, and recognized proteins of eight KEGG categories: carbohydrate metabolism and energy (38%), amino acid metabolism (23%), stress/defense response (11%), and proteins from the glycan metabolism, protein metabolism and signal transduction all representing 4% of the total.

Low rates and quality of protein identification is also common to other environments. For example, Kan et al. (2005) reported that of the 140 proteins extracted and detected by the marine microbial communities of the Chesapeake Bay, only few of them were identified with high quality scores (i.e. two or more peptides).

The above mentioned published data on soil proteomic studies show that only a limited fraction of the theoretical soil metaproteome can be extracted and identified proteins, by considering the extremely microbial diversity of soil (Torsvik et al. 2002) and thinking that relatively simple prokaryotic and eukaryotic unicellular organisms such as *E. coli* and *S. cerevisiae* possess $4.3 \cdot 10^3$

and $6.1 \cdot 10^3$ open reading frames, respectively. Although the low rate of protein identification in soil metaproteomic studies can be partly explained by the poor annotation of environmental microbial strains in the proteomic databases, low rates of protein identification have been reported also in model studies based on the proteomic analysis of bacterial strains with known genome and proteome (Giagnoni et al 2011, 2012).

1.6 The rationale of the PhD project

The high reactivity of the proteins and the hypothesis that HS may have a supramolecular structure may explain, at least in part the low protein identification rates common to soil proteomic studies (see paragraph **1.5.6** of this thesis. In fact, even if soils possess a strong proteolytic activity (Renella et al. 2002), proteolysis as well as chemical denaturation is unlikely responsible for protein degradation or denaturation at such an high extent to cause the observed low rate of protein recovery and identification in most of the soil proteomic studies. Complete proteolysis in soil is prevented by several physico-chemical factors which can be resumed as: lack of contact between proteins and proteases due to protein sorption by soil colloids and protein intracellular location, unsuitable pH values, lack of protease co-factor. Moreover, commonly the used lysis buffers contain broad spectrum protease inhibitors.

As above mentioned (see paragraphs 1.3 and 1.4) proteins may react and be stabilized by soil colloids maintaining or modifying their native conformation (Miller et al. 2006). Therefore, it is reasonable to hypothesize that the low protein identification rates may also be due to protein interactions with low molecular weight organic ligands (LMWOLs) present in soil or released before or during protein extraction from soil HS, and such interactions may induce pitfalls in the current protein identification MS analysis. In fact, as above introduced (paragraph **1.5.2** of this thesis), MS protein identification is based on the probability score for ranking possible peptide sequences that best fits an observed tandem mass

spectra, automatically computed by the proteomic databases (e.g. Mascot). The basic assumptions behind this model is that if a peptide sequence produce an established number of fragment ions, and a sufficient number of MS peaks match the predicted ones, then the "random-chance probability is computed as to have a 5% chance of yielding correct scores. Therefore, the use of the probability score makes search engine results easier to interpret. However, other scores based on other mathematical procedures can be used [e.g. SEQUEST's cross-correlation score (XCorr), Mowse score] but these bioinformatics procedures require more basic knowledge (Gasteiger et al., 2005). Diffrently from the SEQUEST, the Mascot database uses a probability based algorithm for the assignment of MS/MS spectra to pepetides, and can be coupled with statistical models such as ProteinProphet and PeptideProphet, allowing a more robust and database-independent assessment of the validity of the protein identification (Keller et al., 2002). While it is common to generate thousands of MS/MS spectra, the major challenge in protein MS analysis is the correct peptide assignment, particularly in the case of low quality of spectra. However, in all models, all the mass-to-charge ratios (m/z) peaks are mathematically assumed to be 'independent', but this is not always the case in MS data, so as to be reliable accurate for scoring every matched MS peak. Therefore, the results depend on clean spectra and well-fitting data, with problems in assigning peak probability to ion fragments having low quality spectra. Although different MS manufacturers may require site specific optimal set up, uniformity in the MS spectra analysis may be useful, especially for soil proteomics due to the high diversity of protein sources and potentials of protein extracellular modification or re-arrangements.

An interesting observation on this specific aspect was reported by Nebbioso et al. (2010) whose showed potential drawbacks in the use of ESI-MS analysis in the presence of humic like substances was illustrated whose reported that the detection of a model tripeptide was drastically reduced after contact with hydrophobic organic acids when analyzed by ESI due to differences in ionization of the mixture.

Relying on the supramolecular model of the HS and on the available data on identification of protein extracted from soil, we hypothesized that contact between proteins and HS could lead to the establishment of protein-HS chemical interactions and also destabilize the weak forces of the HS supramolecular structure leading to the binding of low molecular weight organic ligands (LMWOLs) to proteins accompanied by entropy gain caused by protein conformational changes during the adsorption and greater ligand stabilization, as schematized in Figure 1.7. Such an interaction may influence not only the protein extraction from soil, but also their purification and identification by MS, due to uncertain and unpredictable induced modifications, thus partially explaining the limited protein identification in the published soil proteomic studies.

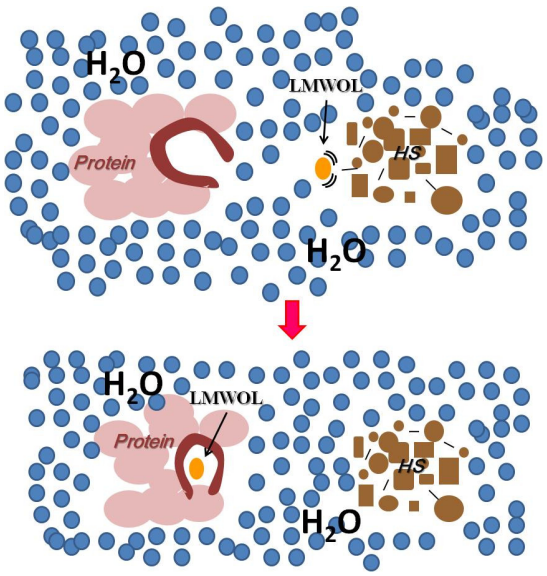


Figure 1.7. Representation of possible protein - humic molecules interaction

1.7 Aim of the PhD project

The aim of my PhD project was to study the effects of interactions between soil borne HS and various proteins. In my PhD project, the protein HS interactions

were studied with complementary techniques to assess the establishment of HS-protein interactions, the type and strength of such interactions, and eventual protein modification after contact with soil borne HS.

To fulfil the aim of the project, I selected proteins with different molecular properties and tested the interactions with purified HS or soil samples using a quartz microbalance, electrophoresis coupled to MS, and nuclear magnetic resonance (NMR).

Chapter 2: Materials and Methods

2.1 The selected proteins

Four commercially available proteins with different structural properties and chemical reactivity, namely α- and β-glucosidase (βg) from *A. Niger*; myoglobin from horse muscle (Mb) and a recombinant ovine prion protein (RecPrP), were selected for the electrophoresis and MS experiments. A fifth protein, CopH from *C. metallidurans*, containing ^{13}C- and ^{15}N-labeled aminoacids, was a custom preparation for the NMR studies.

2.1.1 The α- and β-glucosidase

The α-glucosidase (αg) and β-glucosidase (βg) was purchased by the Sigma Aldrich company (cat. n. 49291). The main characteristics of the α-glucosidase (αg) are: number of amino acids 640, Molecular weight 68.3 kDa, and theoretical pI: 4.25.

The **α-glucosidase** is a dimeric protein with 25.5% carbohydrate, the sequence is:

```
        10         20         30         40         50         60
MSFRSLLALS GLVCTGLANV ISKRATLDSW LSNEATVART AILNNIGADG AWVSGADSGI

        70         80         90        100        110        120
VVASPSTDNP DYFYTWTRDS GLVLKTLVDL FRNGDTSLLS TIENYISAQA IVQGISNPSG

       130        140        150        160        170        180
DLSSGAGLGE PKFNVDETAY TGSWGRPQRD GPALRATAMI GFGQWLLDNG YTSTATDIVW

       190        200        210        220        230        240
PLVRNDLSYV AQYWNQTGYD LWEEVNGSSF FTIAVQHRAL VEGSAFATAV GSSCSWCDSQ

       250        260        270        280        290        300
APEILCYLQS FWTGSFILAN FDSSRSGKDA NTLLGSIHTF DPEAACDDST FQPCSPRALA

       310        320        330        340        350        360
NHKEVVDSFR SIYTLNDGLS DSEAVAVGRY PEDTYYNGNP WFLCTLAAAE QLYDALYQWD

       370        380        390        400        410        420
KQGSLEVTDV SLDFFKALYS DAATGTYSSS SSTYSSIVDA VKTFADGFVS IVETHAASNG

       430        440        450        460        470        480
SMSEQYDKSD GEQLSARDLT WSYAALLTAN NRRNSVVPAS WGETSASSVP GTCAATSAIG

       490        500        510        520        530        540
TYSSVTVTSW PSIVATGGTT TTATPTGSGS VTSTSKTTAT ASKTSTSTSS TSCTTPTAVA

       550        560        570        580        590        600
VTFDLTATTT YGENIYLVGS ISQLGDWETS DGIALSADKY TSSDPLWYVT VTLPAGESFE

       610        620        630        640
YKFIRIESDD SVEWESDPNR EYTVPQACGT STATVTDTWR
```

The main characteristics of the **β-glucosidase** were: number of amino acids: 860, molecular weight 93.2 kDa, theoretical pI 4.65. The β-glucosidase is dimeric or trimeric protein ,10% carbohydrate content, the sequence is:

```
10         20         30         40         50         60
MRFTLIEAVA LTAVSLASAD ELAYSPPYYP SPWANGQGDW AEAYQRAVDI VSQMTLAEKV

        70         80         90        100        110        120
NLTTGTGWEL ELCVGQTGGV PRLGIPGMCA QDSPLGVRDS DYNSAFPAGV NVAATWDKNL

       130        140        150        160        170        180
AYLRGQAMGQ EFSDKGADIQ LGPAAGPLGR SPDGGRNWEG FSPDPALSGV LFAETIKGIQ

       190        200        210        220        230        240
DAGVVATAKH YIAYEQEHFR QAPEAQGYGF NITESRSANL DDKTMHELYL WPFADAIRAG

       250        260        270        280        290        300
AGAVMCSYNQ INNSYGCQNS YTLNKLLKAE LGFQGFVMSD WAAHHAGVSG ALAGLDMSMP

       310        320        330        340        350        360
GDVDYDSGTS YWGTNLTISV LNGTAPQWRV DDMAVRIMAA YYKVGRDRLW TPPNFSSWTR

       370        380        390        400        410        420
DEYGFKYYYV SEGPYEKVNQ FVNVQRNHSE LIRRIGADST VLLKNDGALP LTGKERLVAL

       430        440        450        460        470        480
IGEDAGSNPY GANGCSDRGC DNGTLAMGWG SGTANFPYLV TPEQAISNEV LKNKNGVFTA

       490        500        510        520        530        540
TDNWAIDQIE ALAKTASVSL VFVNADSGEG YINVDGNLGD RRNLTLWRNG DNVIKAAASN

       550        560        570        580        590        600
CNNTIVIIHS VGPVLVNEWY DNPNVTAILW GGLPGQESGN SLADVLYGRV NPGAKSPFTW

       610        620        630        640        650        660
GKTREAYQDY LYTEPNNGNG APQEDFVEGV FIDYRGFDKR NETPIYEFGY GLSYTTFNYS

       670        680        690        700        710        720
NLQVEVLSAP AYEPASGETE AAPTFGEVGN ASDYLYPDGL QRITKFIYPW LNSTDLEASS

       730        740        750        760        770        780
GDASYGQDAS DYLPEGATDG SAQPILPAGG GAGGNPRLYD ELIRVSVTIK NTGKVAGDEV

       790        800        810        820        830        840
PQLYVSLGGP NEPKIVLRQF ERITLQPSKE TQWSTTLTRR DLANWNVETQ DWEITSYPKM

       850        860
VFAGSSSRKL PLRASLPTVH
```

(http://www.brenda-enzymes.org)

2.1.2 The horse muscle myoglobin

The horse muscle myoglobin from (mb) was purchased from Sigma Aldrich (cat n. M0630). The main characteristics of the mb were: number of amino acids: 154, molecular weight, 17.1 kDa, theoretical pI 7.20. The sequence of the mb was:

```
        10         20         30         40         50         60
GLSDGEWQQV LNVWGKVEAD IAGHGQEVLI RLFTGHPETL EKFDKFKHLK TEAEMKASED

        70         80         90        100        110        120
LKKHGTVVLT ALGGILKKKG HHEAELKPLA QSYATKHKIP IKYLEFISDA IIHVLHSKHP

       130        140        150
GDFGADAQGA MTKALELFRN DIAAKYKELG FQ
```

2.1.3 Recombinant ovine prion protein (RecPrP)

The recombinant prion protein (RecPrP) was made available thanks to Dr Luigi D'Acqui from the Institute for Ecosystem Study, Florence CNR (ISE) of Florence, and was obtained by the Virologie et Immunologie Molèculaires INRA, (Jouy-en-Josas, France). The main characteristics of the mb were: number of amino acids 256, molecular weight 27.9, theoretical pI: 9.47. The sequence of the RecPrP was:

```
        10         20         30         40         50         60
MVKSHIGSWI LVLFVAMWSD VGLCKKRPKP GGGWNTGGSR YPGQGSPGGN RYPPQGGGGW

        70         80         90        100        110        120
GQPHGGGWGQ PHGGGWGQPH GGGWGQPHGG GGWGQGGSHS QWNKPSKPKT NMKHVAGAAA

       130        140        150        160        170        180
AGAVVGGLGG YMLGSAMSRP LIHFGSDYED RYYRENMYRY PNQVYYRPVD QYSNQNNFVH

       190        200        210        220        230        240
DCVNITVKQH TVTTTTKGEN FTETDIKIME RVVEQMCITQ YQRESQAYYQ RGASVILFSS

       250
PPVILLISFL IFLIVG
```

The ovine RecPrP variant (ARQ212) used for the experiments was expressed in *E. coli* and had the following modifications as compared to the pathogenic wild type ovine prion protein: alanine (Ala, A) in positions 136, arginine (Arg, R) in position 154, and glutamine (Gln, G) in position 171, and did not contain the glycosyl parts. The used RecPrp was structurally well characterized Leclerc

et al. 2001, Eghiaian et al. 2004) and represented a good model of the natural prion proteins (Eghiaian et al. 2004, Somerville R. A. 2002).

The overall fold of the crystallized part of the C-terminal domain (residues 114–234) of the ARQ variants analyzed consists of a short two stranded β-sheet (residues 129–134 and 163–167) and three α-helices (residues 146–158, 174–196, and 203–228), linked by loops with no regular secondary structure (Figure 3 chap 1); residues 114–126 and 229–234 are disordered and not seen in the electron density.

Protein mix

The protein mix was obtained mixing Myoglobin from Horse Muscle and Glucosidases from A. Niger.

The reported protein sequences have been obtained analyzing by HRMS the control protein lane in the electrophoretic gel, loading each model protein dissolved in sterile milliQ water 1 mg/ml, against the SwissProt protein database as described below.

2.1.4 CopH from C. metalidurans

The CopH protein from *C. metallidurans* was purchased by Giotto Biotech (Florence, Italy). The main characteristics of CopH were: number of amino acids: 127, molecular weight 13.2 kDa, theoretical pI 5.60. the CopH sequence was:

```
        10         20         30         40         50         60
ADKLESTNPA GWSQQAQIQV AAQHDHGARA GLSSLSAIKP GPEASLFGSE APAGSTSRVV

        70         80         90        100        110        120
DVAPGLKYVN VDSGETVTFK SGASEITFAF AKLDRNKAVA LNVLFPELPG GQGVWVYIEQ

SRLYIGG
```

The CopH protein was labeled with ^{13}C and ^{15}N, specifically for efficient NMR analysis of protein HS interactions. CopH is a metal binding protein with high affinity for Cu(II) , zinc and nickel, probably involved in copper homeostasis (Sendra et al. 2006).

2.2 Humic Substances

2.2.1 Commercial humic acid

Commercial humic acids (cHA) used for the preliminary experiments with the quartz microbalance were purchased by Sigma (cat. N. H16752). Chemical characteristics of cHA were available in literature (Aeschbacher et al. 2012, Tan et al. 2008).

2.2.2. Soil borne humic substances

Soil borne humic substances (HS) were extracted from a natural sandy loam soil, classified as Eutric Cambisol (WRB 1998). The HS were extracted using 0.1M $Na_2P_4O_7$ at pH 11, at 1:10 w:v ratio for 4 h at 60°C (Italian official method, G.U. n° 248, 21-10-1999). The soil slurry was centrifuged (10^4 rpm) and filtered through a Millipore 0.22 µm membrane. Afterward, the pH value was adjusted to 7 using 50% H_2PO_4.

The HS were then fractionated in humic and fulvic acids according to Parsons (1988). The extract was acidified with 50% H_2SO_4 to pH 1 to allow humic acids (HA) precipitation. The supernatant was then loaded onto acidified polyvinylpyrrolidone (PVP) columns to separate the fulvic acids (FA) retained from the insoluble fraction (humin); the FA fraction was eluted by the PVP columns by 0.1 M NaOH. The C content of HS, FA and HA was determined by the wet oxidation method, using $K_2Cr_2O_7$ as oxidant in a H_2SO_4 (Italian official method, G.U. n° 248, 21-10-1999). All HS, FA and HA fractions were neutralized to pH 7 using diluted H_2SO_4 or NaOH solutions as appropriate and

stored at 4° for maximum 7 d and filtered on 0.22 μm membranes prior to each HS-protein contact experiments. The C content of the extracted HS was 5.15 g C Kg^{-1}, 2.64 g C Kg^{-1} for FA, and 2.05 g C Kg^{-1} for HA.

2.2.3 Soil borne humic acids

Humic acids used for the NMR experiments were extracted from from a Typic Fulvuland soil (Lake Vico, Lazio, Italy) modifying Stevenson (1994). The raw material was shacked overnight in a 0.5 M NaOH and 0.1 M Na4P2O7 solution under N_2 atmosphere. The suspension was centrifuged at 10^4 rpm and the supernatant filtered through glass wool and treated with 6N HCl to pH 1 to allow precipitation of humic acids during one night at 4°C. The precipitated humic acids were separated by centrifugation and purified from coextracted inorganic particles first by three cycles of dissolution in 0.5 M NaOH followed by flocculation in 6 M HCl, and then shaking humic acids twice in a 0.25 M HF/HCl solution for 24 h. The purified humic acids was then dialyzed against distilled water until chloride-free and freeze-dried. The 30 mg of humic acids were suspended in H_2O, titrated to pH 7 with NaOH 0.1 M by an automatic titrator and freeze dried again. Finally humic acids were then redissolved in 0.5 M NaOH and passed through a strong cation-exchange resin (Dowex 50) to further eliminate divalent and trivalent metals, and freeze-dried again.

2.2.4 Soil for protein inoculation experiments

The soil used (Eutric Cambisol) for evaluating the persistence and modification of αg and βg proteins was the same as that used for HS extractions. Soil sample was collected from the top soil (1–15 cm) of a natural forest located in Tuscany (Italy), and it was classified as sandy loam (USDA). Soil sample main characteriscs were: total organic carbon (TOC) 1,21%; total nitrogen (TN) 0,07%; pH 5.52; electrical conductivity 0,12 dS/m.

2.3 Analytical techniques

2.3.1 Quartz microbalance

The quartz microbalance device used for the HS-protein experiments was an Elba Tech (Tombelli et al. 2000). The quartz microbalance was used for contact experiments between cHA and αG and βG.

The instrument consisted of a A 10 MHz AT cut quartz crystal (14 mm) covered with evaporated gold (42.6 mm^2 area) on both sides (Mistral, Latina, Italy), located into a polyvinyl chloride cell in the way that only one side of the crystal was in contact with the solution. The crystal resonance frequency were continuously recorded using a quartz crystal analyzer (Model QCA917, Seiko EG&G, Chiba, Japan), connected to a PC connected to the QCA917 interface. The electrode surface was cleaned with a boiling solution of H_2O_2 (33%), NH_3 (33%) and milliQ water at a 1:1:5 ratio for 10 min prior to use, and always used immediately after cleaning.

The gold on the quartz was functionalized with thiol and carboxylated dextran to immobilize the enzyme proteins, by immersing it in 1 mM ethanolic solution of 11-mercaptoundecanol at room temperature in the dark, for 48 h. The crystal was then washed with ethanol and milliQ water and sonicated for 10 min in ethanol to remove the excess thiol. The hydroxyl surface was then treated with a 600 mM solution of 1:1 epichlorohydrin-400 mM NaOH : bis-2-methoxyethyl ether (diglyme) for 4 h. After washing with water and ethanol, the crystal was immersed for 20 h in a basic dextran solution, consisting of 3 g of dextran in 10 ml of 100mM NaOH 100. The surface was further functionalized with a carboxymethyl group using 1M bromoacetic acid solution in 2 M NaOH for 16 h). All the reactions were carried out at room temperature. The crystal was then washed with milliQ water and put in the PVC cell. The surface of the crystal was further activated with 200 ml of a solution of 50 mM N-

hydroxysuccinimide (NHS) and 200 mM 1-ethyl-3-(3-dimethylaminopropril) carbodiimide (EDAC) in H_2O. After 5 min the activating solution was replaced by the αg and βg (200 mg/ml) in 10 mM acetate buffer at pH 5. After 20 min, the residual reacting sites were blocked with 200 ml of a 1M solution of ethanolamine hydrochloride at pH 8.6). After washing with the last immobilization buffer, the cHA solution (1 mg/ml), and the reaction was allowed to proceed for 20 min under static conditions.

2.3.2 Electrophoresis

The electrophoretical analysis was conducted under denaturing or native conditions. The denaturing gel electrophoresis was performed using a 12.5% resolving polyacrylamide gel (PAGE) containing sodium dodecyl sulfate (SDS-PAGE), topped by 4% polyacrylamide stacking gel.

The native-PAGE (N-PAGE) was performed using a 10% resolving polyacrylamide gel topped by a 4% polyacrylamide stacking gel. All electrophoretical separations of proteins, HS, HA and FA were performed in 5X Laemmli buffer (Laemmli et al. 1970), and under 150 and 100 V tension for the SDS-PAGE and N-PAGE, respectively, until the tracking dye front reached the bottom of the gel.

After both electrophoresis type, the protein banding patterns were detected using Coomassie Brilliant Blue G as staining. Gel images were acquired by Biorad Optical Densitometer GS-800.

Detection of β-glucosidase catalytic activity in N-PAGE was done by incubating the gel bands with 1 mM 4-methylumbelliferil-β-d-glucopiranoside (MUG) in 50 mM succinate buffer at pH 4 for 20 min (Kim et al. 2007), prior to coomassie staining. The release of 4-methylumbelliferone, a fluorescent product of the enzymatic hydrolysis, was visualized by an UV transilluminator (Vilber Lourmat) and the image acquired by a digital camera (Canon A40).

2.3.3 Mass spectrometry analysis

The protein bands and the respective negative controls were de-stained using a 1% H2O - 1% CH3COOH solution and were manually excised and processed as described by Dani et al. (2011).

Bands of interest (in the range of 10–20 kDa) were excised from the gel and individually transferred to a 1.5-mL microcentrifuge tube. Bands were washed 3 times for about 10 min in 120 µl of acetonitrile and then in 100 µl of a 0.1 M ammonium bicarbonate water solution. The solution was removed and 100 µl of a 1 ng/lL of modified trypsin (Promega) in 10mM ammoniumbicarbonate was added to each band. Tube were kept at 4 °C for 30 min, then the solution was removed and replaced with 100 µl of 10 mM ammonium bicarbonate. After overnight digestion at 37 °C, the supernatant was recovered and the reaction blocked by addition of 10 µl 10% TFA.

- **ESI-MS**

The peptide mixture was submitted to nano-HPLC-ESI FTMS analysis using an Ultimate 3000 (Dionex, San Donato Milanese, Milano, Italy) coupled to a LTQ Orbitrap mass spectrometer (Thermo Fisher, Bremen, Germany). Peptides were concentrated on a precolumn cartridge PepMap100 C18 (300µm i.d. × 5mm, 5µm, 100Å, LC Packings Dionex) and then eluted on a C18 PepMap100 column (75µm i.d. × 5cm, 3µm, 100Å, LC Packings Dionex) at 300 nl min-1. The mobile phases composition was: H2O 0.1% formic acid/CH3CN 97/3 (phase A) and CH3CN 0.1% formic acid/ H2O 97/3 (phase B). The gradient program was: 0 min, 4% B; 10 min, 40% B; 30 min, 65% B; 35 min, 65% B; 36 min, 90% B; 40 min, 90% B; 41 min, 4%B; 60 min, 4% B.

The mass spectra were acquired in positive ion mode, setting the spray voltage at 2 kV, the capillary voltage and temperature respectively at 45 V and 200 °C, and the tube lens at 130 V. Data were acquired in data-dependent mode with

dynamic exclusion enabled (repeat count 2); survey MS scans were recorded in the LTQ Orbitrap analyzer in the mass range 300-2000 Th at a 15.000 nominal resolution, then up to three most intense ions in each full MS scan were fragmented and analyzed in the IT analyzer.

The acquired spectra were searched against nrNCBI (http://www.ncbi.nlm.nih.gov) using MASCOT as search software (Perkins et al. 1999) against the SwissProt protein database (http://www.uniprot.org/), the following search parameters were applied: maximum two missed trypsin cleavages, cysteine carbamidomethylation as fixed modification, methionine oxidation as variable modifications, and detection of bi-, tri and quadri charged ions and a maximum 0.6 Da error tolerance in MS/MS data and 1.2 Da in MS.

- **PEP 3D software analysis to evaluate the ESI signals**

The quality of the protein spectra and criteria for protein identification were checked using the Trans Proteomic Pipeline (TPP) tool of the PEP 3D software (Deutsch et al. 2010). The TPP software is an open-source proteomic data analysis pipeline developed, maintained jointly by the Institute for Systems Biology (ISB) of the ETH in Zurich and the Seattle Proteome Center (http://localhost/tpp-bin/tpp_gui.pl). The use of the TPP software permitted the analysis of the ESI-MS spectra in order to evaluate sample and chromatography quality, and to further validate the results obtained by PeptideProphet search engines (Keller et al. 2002), that assigned an additional probability of correctness to each interpretation.

- **MALDI-MS**

MALDI-MS measurements were performed using a MALDITOF Ultrafle III time-of-flight mass spectrometer (Bruker Daltonics, Bremen, Germany), equipped with 1 kHz smartbeam II laser (l = 355 nm) and operating in the positive reflectron ion modes, mass range 860m/z a 4000m/z. External mass

calibration was carried out using the Peptide Calibration Standard (Bruker Daltonics) with the same mass range.

- **Protein grand average of hydropathicity and Kyte-Doolittle Plots**

Some of the protein properties were investigated using the grand average of hydropathicity (GRAVY index) and the Kyte-Doolittle plots (Kyte and Doolittle, 1982), which can be obtained by the protein identification data. The GRAVY index provide indications on the protein solubility, as positive GRAVY index values indicate protein hydrophobic domains whilst negative GRAVY index values indicate protein hydrophilic domains. The Kyte-Doolittle plots can be generated by the http://gcat.davidson.edu/DGPB/kd/kyte-doolittle.htm website. In all cases, the window size chosen was 9, so that strong negative peaks indicate external regions of globular proteins whereas positive peaks indicate internal regions.

2.3.4 Nuclear magnetic resonance experiments

The nuclear magnetic resonance (NMR) experiments were carried out at the Center for Nuclear Magnetic Resonance of the Federico II University of Naples (CERMANU), under the supervision of Prof A. Piccolo and P. Mazzei.

A Bruker Avance 9.4 Tesla (400 MHZ) equipped with 1.7 mm Inverse Quadrupole resonance (μTXI) probe, working at the frequencies 40.531, 100.57 e 400.13 MHz for ^{15}N, ^{13}C e ^{1}H, respectively, was employed to conduct all liquid-state NMR measurements at a temperature of 298 +/- 1 K. ^{1}H NMR spectra were acquired with 2 s of thermal equilibrium delay, 90° pulse length ranging between 4.5 and 5.5 μs, 32768 time domain points and 260 transient.

An inversion recovery and a Carr-Purcell-Meiboom-Gill (CPMG) pulse sequence, with 16 increments and 5 s delay, were adopted to measure ^{1}H longitudinal (spin-lattice) relaxation time constants (T1) and transverse (spin-

spin) relaxation time constants (T2). In the case of the inversion recovery experiments the delays were 0.02, 0.04, 0.06, 0.08, 0.09, 0.1, 0.2, 0.4, 0.6, 0.7, 0.85,1, 1.3, 1.5, 3 e 5 s, while in the case of Carr-Purcell-Meiboom-Gill (CPMG) the delays were 0.004, 0.012, 0.02, 0.06, 0.08, 0.1, 0.252, 0.5, 0.6, 0.848, 1, 1.24, 1.6, 1.9, 2.5 e 3 s. In both cases a constant 1 ms spin-echo delay was employed.

The ^{1}H DOSY (Diffusion Ordered SpectroscopY) NMR spectra were obtained by choosing a stimulated echo pulse sequence with bipolar gradients, and combined with two spoil gradients and an eddy current delay before signal acquisition. This sequence was selected to reduce signals loss due to short spin-spin relaxation times. The acquisition (32768 points) was executed by 1750 μs long sine-shaped gradients (δ), that linearly ranged from 0.674 to 32.030 G cm-1 in 32 increments, and selecting a 0.24 s delay (Δ) between the encoding and the decoding gradients. In order to observe short-range correlation between proton and carbon (^{13}C) and nitrogen (^{15}N) nuclei, 2D HSQC (Hetero Single Quantum Correlation), optimized with coupling constant J_{H-X} di 145 e 95 Hz, respectively. The ^{1}H spectral width was 16.66 ppm (10 kHz) and the residual water signal was removed from ^{1}H -NMR spectra by pre-saturation technique, whereas the Watergate 3-9-19 pulsed train sequence was preferred for DOSY experiments.

Spectral width for all proton acquisitions were 16 ppm (6410.26 Hz), the FID (Free Induction Decay) was processed by Fourier Transform without zero filling technique and with line broadening of 0.8, 4 ad 8 Hz, respectively, for the proton acquisition, for relaxation times and diffusion experiments. All spectra were baseline corrected and processed by both Bruker Topspin Software (v.1.3), MestReC NMR Processing Software (v. 4.9.9.9) and Origin (v.6.1). Correlation times were obtained by T1 and T2 relaxation times following Carper and Keller 1997.

2.4 Experiments

The protein-humic substances contact experiments aimed at testing the following varaibles: protein-to-HS ratios, protein contact with HA and FA fractions, contact time, contact pH values, shaking vs static incubations.

2.4.1 Contact experiment 1: incubation of proteins – non fractionated humic substances (HS)

Proteins were solubilized in sterile MilliQ water in the concentration 1 mg ml^{-1} prior the incubation with HS. In the case of protein mix, Mb and glucosidase preparation were solubilazed in sterile MilliQ water in the final concentration of 1 mg ml^{-1}. Control proteins were represented by the model protein, dissolved 1 mg ml^{-1} in sterile water, incubated with HS extractant (0.1M $Na_2P_4O_7$ pH 11) neutralized to pH 7 using diluted H_2SO_4.

The factors tested to evaluate the effects of HS on protein stability and β-glucosidase catalytic activity were:

- contact ratio- 1:1, 1:2, 1:4 protein:HS
- contact time- 24 h, 72 h, 6 d
- contact mode- shaking, static incubation
- contact pH- 3.6, 7, 10

The proteins-HS contact experiments were carried out using 2 ml sterile Eppendorf plastic tubes, in the different protein:HS ratios indicated, and in the different contact periods indicated, either in static mode or under shaking at 200 reverse per minute, in triplicates. The different incubation pH were evaluated by incubation only on 1:10 pC:hC ratio and for 1 d. At the end of the incubation period SDS-PAGE and N-PAGE were performed for each sample, gel bands were cut for MS analysis.

In the case of N-PAGE, the β-glucosidase in gel activity was detected as described in 2.2.2 paragraph prior coomassie staining.

2.4.2 Contact experiment 2: proteins incubation in soil

Resistance of selected model proteins in soil was evaluated by inoculating 5 mg of protein for 1 d and 6 d in 100 g of the Eutric Cambisol, previously kept in the dark at 25°C and 60% soil water holding capacity for 7 d. At the end of the incubation period the protein was extracted accorded to Doni et al. 2012 (annex II) using 0.5 M K_2SO_4 at pH 6.6 (1:3 w:v ratio) at room temperature for 1 h under shaking at 200 rpm, with the slight modification of the use of 10 mM EDTA as metal-protease inhibitor. The soil extracts were filtered through a 0.22 μm cellulose nitrate filter (Sartorius) to remove microbial cells and appropriately diluted. The solutions were dialysed against sterile distilled water over night at 4°C using 3500 MW cellulose tubular membrane (Cellu Sep T1, Orange Scientific) until an electrical conductivity value lower than 0.5 dS m^{-1} was reached. Each dialyzed sample was concentrated by using Amicon PM-10 diaflomembrane, molecular cut-off 10 KDa (Millipore) until a volume of 10 ml under inert (nitrogen) atmosphere and subsequently by Amicon Ultra-15 centrifugal filter devices (molecular cutoff 10 KDa) until 50 μl for the native gel electrophoresis (native-PAGE) and 1 ml for the denaturing gel electrophoresis (SDS PAGE). The 1 ml extract was precipitated by adding a trichloroacetic acid Na-deoxycholate mixture (TCA-DOC) for 1 h at 4°C, followed by overnight precipitation with TCA at -20°C. Protein solutions were then centrifuged at 14600 x *g* for 15 min at 4°C, the supernatant was removed and the pellet was washed with 1 ml of cold acetone. The supernatant was evaporated and the protein pellet was air-dried for 5 min and stored frozen prior to use.

SDS-PAGE and N-PAGE were performed for each sample, gel bands were cut for MS analysis.

In the case of N-PAGE, prior coomassie staining β-glucosidase in gel activity was detected as described in 2.2.2 paragraph.

2.4.3 NMR experiments

- CopH and Humic acids

Lyophilized **CopH** protein and lyophilized humic acids (HA) were dissolved 16 mg/ml in 0.1 M phosphate buffer pH 7 (15% D_2O) in the following ratios:

- 1:0 (CopH:HA)
- 1:0.5 (CopH:HA)
- 1:1 (CopH:HA)

The incubations were carried out for 24 h before the analysis, each test had two replicates.

In order to review all the experiments done a summarizing table in reported below (Table 2.1).

Table 2.1 Summarizing experiments table

	TECHNIQUES					
PROTEINS		**PAGE**	**ESI-MS**	**MALDI-MS**	**Microbalance**	**NMR**
	αg	X	X		X	
	βg	X	X		X	
	Mb	X	X	X		
	RecPrP	X	X			
	CopH					X

Chapter 3: Results

3.1 Quartz microbalance analysis

The reaction partners for this experiment were the αg and βg from *A. niger* and the commercial humic acids (cHA).

The results of the quartz crystal vibrational frequency after washing, functionalization, enzyme sorption and saturation of the sorptive sites with ethanolamine are shown in Figure 3.1.

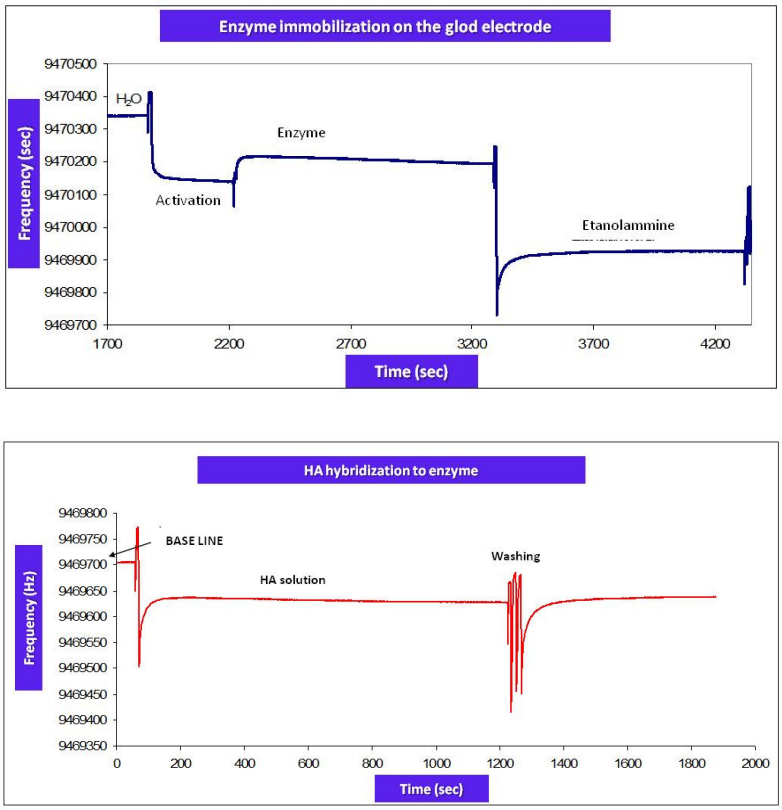


Figure 3.1. Variations in the quartz crystal vibrational frequency *(Top)* during the enzyme immobilization phase. *(Bottom)* during the treatment with HA

As expected, the vibration frequency decreased upon functionalization, enzyme sorption and site saturation, due to the increase of the molecular mass recorded by

the quartz crystal. After incubation of the cHA and the final washing, the average of the recorded vibrational frequency in the replicate samples decreased by 49.8 (±28.6) Hz.

However, the blank measurement carried out without the immobilized enzyme showed a decrease in the vibrational frequency in the same order of magnitude (78 Hz) as the enzyme-cHA contact experiments. Therefore, the measurement signal provided by the instrument could not unequivocally be interpreted as a protein-cHA interaction but was a false positive result, probably due to interactions between the cHA and the ethanolamine used to saturate the free carboxylic groups after enzyme sorption.

3.2 Electrophoresis experiments

The analysed αg, βg, Mb either singly or as protein mixture, and Mix proteins were all well detected in all the SDS-PAGE experiments. No alteration of the electrophoretical migration of the analyzed proteins could be detected after the incubation with total HS, regardless of the incubation ratio (1:1, 1:2, 1:4), incubation time (24 h, 72 h, 6 days), contact pH value, static *vs* dynamic incubation, in the presence of cations, and if used singly or as αg + βg + Mb protein mixture (Figure 3.1, 3.2, 3.3, 3.4).

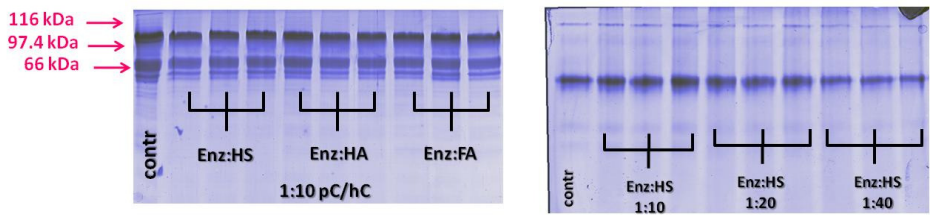


Figure 3.2 Left: SDS-PAGE gel, glucosidases incubated with HS, HA, FA 1:1 incubation ratio, for 6 days under static condition. Right: N-PAGE gel, glucosdases incubated in 1:1, 1:2, 1:4 incubation ratios on shacking

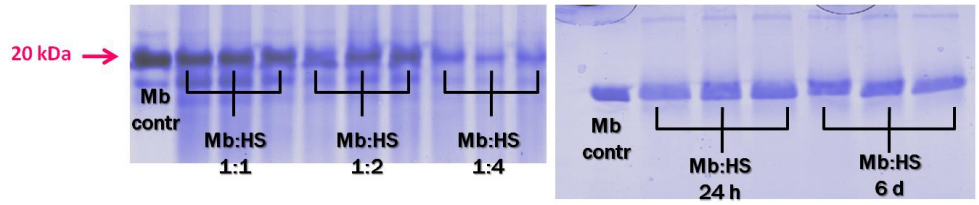


Figure 3.3 Left: SDS-PAGE gel, Mb incubated in 1:1, 1:2, 1:4 incubation ratios with HS for 72 h on shacking. Right: N-PAGE gel, Mb incubated in ratio 1:1 with HS for 24 h and 6 days in static condition

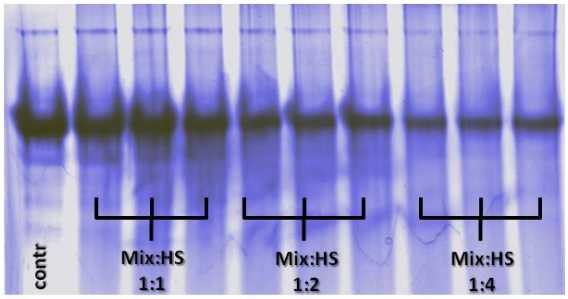


Figure 3.4 N-PAGE gel, protein mix incubated in 1:1, 1:2, 1:4 incubation ratios for 24 h under shacking

The N-PAGE β-glucosidase mobility and catalytic activity were not affected by contact with total HS, HA and FA, regardless of contact time and pH value, and static *vs* dynamic incubation (Figure 3.4).

The electrophoretical analysis of αg, βg and Mb incubated into soil showed marked differences between the tested proteins, as the αg and βg were recovered from soil either after 24 h and 6 days after incubation, whereas the Mb could not be detected after extraction (Figure 3.6). Moreover, the βg also displayed β-glucosidase activity after 24 h and 6 days of incubation into soil (Figure 3.6).

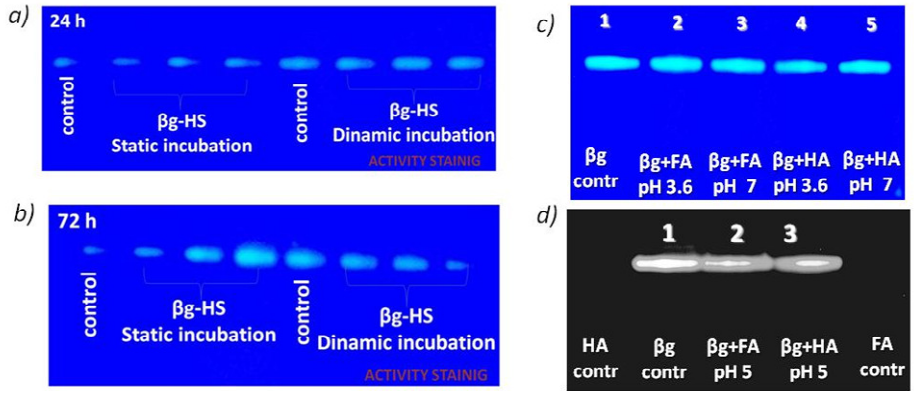


Figure 3.5 Mobility and catalytic activity of β-glucosidase in N-PAGE. *a)* βg:HS 1:1, 1:2, 1:4 incubation ratios, for 24 h, under static and dynamic contact mode; *b)* βg:HS 1:1, 1:2, 1:4 incubation ratios, for 72 h under static and dynamic contact mode; *c, d)* βg incubation with FA and HA and $CaCl_2$ 1 mM at pH 3.6, 5, 7 (1:1 ratio, 24 h, static incubation).

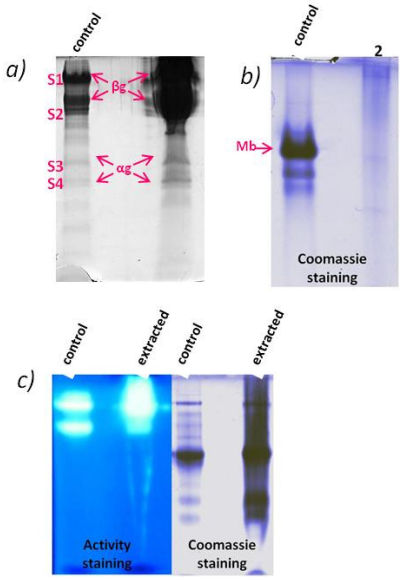


Figure3.6 Glucosidases and Mb N-PAGE profile and βg catalytic activity after 24 h and 6 days of incubation into soil.

3.3 Protein Mass spectrometry - ESI MS

Eletrophoretical bands corresponding to the tested proteins were cut and analyzed by MS.

- *Control proteins identification: SDS-PAGE and Native-PAGE coupled to MS*

ESI-MS control protein identification values are reported in table 3.1

Table 3.1 Identification parameters of αg, βg, Mb and RecPr in SDS-PAGE and N-PAGE (MASCOT algorithm). (ND=not detected)

βg						
	SDS-PAGE			**N-PAGE**		
Coverage %	45	39	45	16	16	17
Identification	α-glucosidase *A. niger*	α-glucosidase *A. niger*	α-glucosidase *A. niger*	β-glucosidase *A. niger*	β-glucosidase *A. niger*	β-glucosidase *A. niger*
Identified peptides	ND	ND	ND	11	11	11
αg						
	SDS-PAGE			N-PAGE		
Coverage %	49	47	52	44	51	51
Identification	α-glucosidase *A. niger*	α-glucosidase *A. niger*	α-glucosidase *A. niger*	α-glucosidase *A. niger*	α-glucosidase *A. niger*	α-glucosidase *A. niger*
Identified peptides	22	19	23	16	18	18
Mb						
	SDS-PAGE			**N-PAGE**		
Coverage %	89	89	89	89	89	89
Identification	Myoglobin *Equus caballus*	Myoglobin *Equus caballus*	Myoglobin *Equus caballus*	Myoglobin *Equus caballus*	Myoglobin *Equus caballus*	Myoglobin *Equus caballus*
Identified peptides	17	17	18	17	18	17
RecPr						
Coverage %	30	28	30	27	27	28
Identification	*Budorcas taxicolor*	*Budorcas taxicolor*	*Budorcas taxicolor*	*Budorcas taxicolor*	*Budorcas taxicolor*	*Budorcas taxicolor*
Identified peptides	8	7	8	5	5	5

All of the tested proteins were consistently identified by MS, after SDS- and N-PAGE, with variable parameters between the two electrophoresis technique, likely due to the different peptide patterns produced by the tryptic digestion.

3.3.1 Protein identification after contact with HS

The MS identification parameters of αg incubated for 24 h under static conditions, pH 7, 1:1, 1:2, 1:4 incubation ratios after SDS-PAGE and N-PAGE are reported in Tables 3.2 and 3.3.

- *α-glucosidase*

ESI-MS α-glucosidase identification values are reported in table 3.2

Table 3.2 Identification parameters (MASCOT algorithm) of **αg** incubated with HS at various ratios, and after **SDS-PAGE** and **N-PAGE**

αg			
SDS-PAGE			
Incubation ratio	**1:1**	**1:1**	**1:1**
Coverage % and identification	43 α- glucosidase *A. niger*	46 α- glucosidase *A. niger*	43 α- glucosidase *A. niger*
Identified peptides	19	20	19
Incubation ratio	**1:2**	**1:2**	**1:2**
Coverage % and identification	42 α- glucosidase *A. niger*	51 α- glucosidase *A. niger*	43 α- glucosidase *A. niger*
Identified peptides	19	19	15
Incubation ratio	**1:4**	**1:4**	**1:4**
Coverage % and identification	29 α- glucosidase *A. niger*	32 α- glucosidase *A. niger*	23 α- glucosidase *A. niger*
Identified peptides	12	15	12
N-PAGE			
	1:1	**1:1**	**1:1**
Coverage % and identification	51 α- glucosidase *A. niger*	51 α- glucosidase *A. niger*	58 α- glucosidase *A. niger*
Identified peptides	18	18	21
	1:2	**1:2**	**1:2**
Coverage % and identification	44 α- glucosidase *A. niger*	44 α- glucosidase *A. niger*	44 α- glucosidase *A. niger*
Identified peptides	15	18	17
	1:4	**1:4**	**1:4**

Coverage % and identification	45 α- glucosidase *A. niger*	39 α- glucosidase *A. niger*	34 α- glucosidase *A. niger*
Identified peptides	16	15	12

- *β-glucosidase*

ESI-MS β-glucosidase identification values are reported in table 3.3.

Table 3.3 Identification parameters (MASCOT algorithm) of **βg** incubated with HS at various ratios, and after **SDS-PAGE** and **N-PAGE.** (ND= not detected)

SDS-PAGE			
Incubation ratio	**1:1**	**1:1**	**1:1**
Coverage % and identification	39 α- glucosidase *A. niger*	45 α- glucosidase *A. niger*	48 α- glucosidase *A. niger*
Identified peptides	ND	ND	ND
Incubation ratio	**1:2**	**1:2**	**1:2**
Coverage % and identification	39 α- glucosidase *A. niger*	41 α- glucosidase *A. niger*	45 α- glucosidase *A. niger*
Identified peptides	ND	ND	ND
Incubation ratio	**1:4**	**1:4**	**1:4**
Coverage % and identification	27 α- glucosidase *A. niger*	37 α- glucosidase *A. niger*	37 α- glucosidase *A. niger*
Identified peptides			
N-PAGE			
Incubation ratio	**1:1**	**1:1**	**1:1**
Coverage % and identification	17 β- glucosidase *A. niger*	15 β- glucosidase *A. niger*	20 β- glucosidase *A. niger*
Identified peptides	11	8	11
Incubation ratio	**1:2**	**1:2**	**1:2**
Coverage % and identification	15 β- glucosidase *A. niger*	15 β- glucosidase *A. niger*	15 β- glucosidase *A. niger*
Identified peptides	7	9	7
Incubation ratio	**1:4**	**1:4**	**1:4**
Coverage % and identification	16 β- glucosidase *A. niger*	16 β- glucosidase *A. niger*	14 β- glucosidase *A. niger*
Identified peptides	9	8	6

- *Myoglobin*

ESI-MS identification parameters of Mb incubated for 24 h under static conditions at pH 7, incubation ratios with HS 1:1, 1:2, 1:4 in both SDS-PAGE and N-PAGE are reported in Table 3.4. Figure 3.7 shows the differences in protein coverage % among one replicate per sample incubated in different ratio with HS.

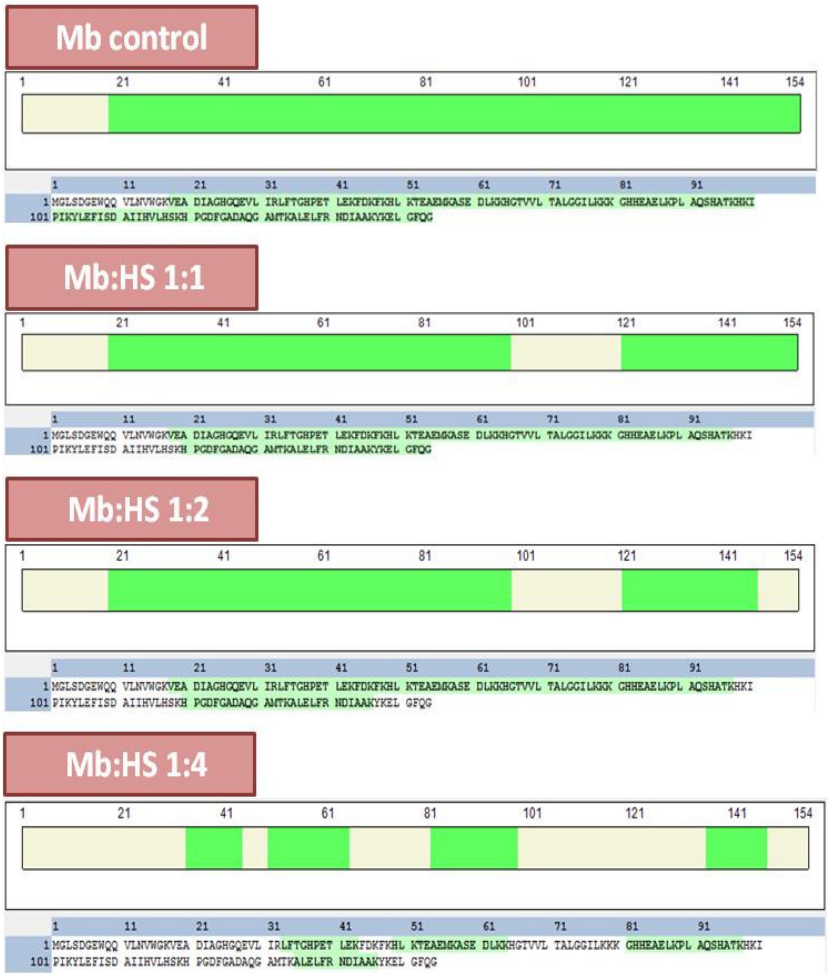


Figure. 3.7 Mb coverage % by ESI-MS analysis

Table 3.4 Identification parameters (MASCOT algorithm) of **Mb** incubated with HS at various ratios, and after **SDS-PAGE** and **N-PAGE**

SDS-PAGE			
Incubation ratio	**1:1**	**1:1**	**1:1**
Coverage % and identification	89 Myoglobin *Equus caballus*	89 Myoglobin *Equus caballus*	89 Myoglobin *Equus caballus*
Identified peptides	17	17	18
Incubation ratio	**1:2**	**1:2**	**1:2**
Coverage % and identification	89 Myoglobin *Equus caballus*	89 Myoglobin *Equus caballus*	89 Myoglobin *Equus caballus*
Identified peptides	17	17	17
Incubation ratio	**1:4**	**1:4**	**1:4**
Coverage % and identification	89 Myoglobin *Equus caballus*	89 Myoglobin *Equus caballus*	89 Myoglobin *Equus caballus*
Identified peptides	17	18	18
N-PAGE			
Incubation ratio	**1:1**	**1:1**	**1:1**
Coverage % and identification	74 Myoglobin *Equus caballus*	78 Myoglobin *Equus caballus*	88 Myoglobin *Equus caballus*
Identified peptides	13	13	15
Incubation ratio	**1:2**	**1:2**	**1:2**
Coverage % and identification	77 Myoglobin *Equus caballus*	74 Myoglobin *Equus caballus*	77 Myoglobin *Equus caballus*
Identified peptides	12	12	12
Incubation ratio	**1:4**	**1:4**	**1:4**
Coverage % and identification	54 Myoglobin *Equus caballus*	47 Myoglobin *Equus caballus*	59 Myoglobin *Equus caballus*
Identified peptides	6	4	7

ESI-MS results of Mb-HS incubations carried out for 6 days are not reported because the detected coverage percentages were the same as the ones reported in tab 3.4.

- *Myoglobin MALDI-MS analysis*

MALDI Mb results were not homogeneous and reproducible, significant differences were found among the replicates of the same samples, as shown in figure 3.8.

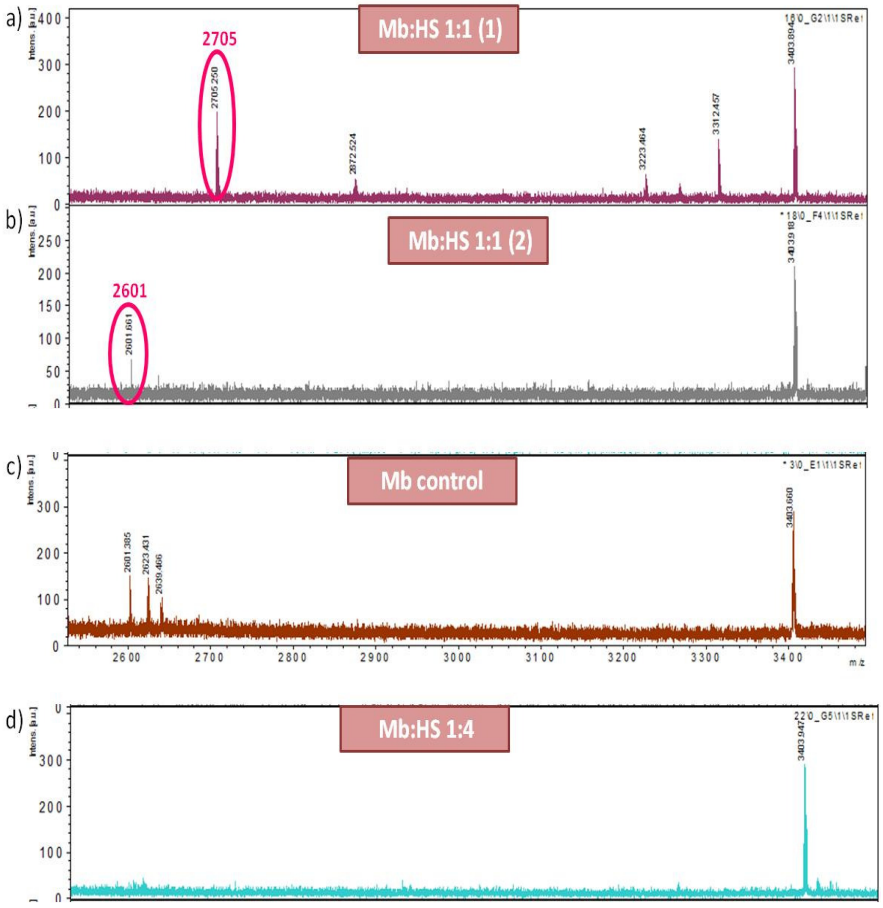


Figure 3.8 Mb MALDI spectra *a)* replicate 1 of incubation ratio 1:1; *b)* replicate 2 of incubation ratio 1:1. Pink rings evidenced the peak possibly subjected to mass shift. *c)* control Mb; *d)* Mb:HS incubation ratio 1:4

- *Mb missed peptide characteristics*

In figure 3.8 b) a peak is evidenced at *m/z* 2601 (MALDI generally produces monocharged ions). The mw of 2601 Da corresponded to the first peptide missed in Mb sequence after the contact with the lower HS amount (1:1) (Figure 3.7). Peptide sequence was: HKIPKYLEFISDAIIHVLHSK

- 21 aa
- Mw: 2601,5
- pI: 8,45
- GRAVY: 0,076 (hydrophobic peptide)

- *RecPrp*

The MS identification parameters of the Mix incubated for 24 h under static conditions at pH 7 incubation at 1:1, 1:2, 1:4 incubation ratios after SDS-PAGE and N-PAGE are reported in Table 3.5.

Table 3.5 Identification parameters (MASCOT algorithm) of the **Mix** incubated with HS at various ratios, and after **SDS-PAGE and N-PAGE**

SDS-PAGE			
Incubation ratio	**1:1**	**1:1**	**1:1**
Coverage % and identification	28 *Canis familiaris*	28 *Canis familiaris*	44 *Capra hircus*
Identified peptides	6	8	8
Incubation ratio	**1:2**	**1:2**	**1:2**
Coverage % and identification	28 *Canis familiaris*	28 *Canis familiaris*	28 *Budorcas taxicolor*
Identified peptides	7	7	7
Incubation ratio	**1:4**	**1:4**	**1:4**
Coverage %	28	30	28

and identification	*Canis familiaris*	*Canis familiaris*	*Budorcas taxicolor*
Identified peptides	6	6	5
N-PAGE			
Incubation ratio	**1:1**	**1:1**	**1:1**
Coverage % and identification	28 *Canis familiaris*	27 *Antilope cervicapra*	23 *Antilope cervicapra*
Identified peptides	5	4	4
Incubation ratio	**1:2**	**1:2**	**1:2**
Coverage % and identification	22 *Canis familiaris*	28 *Canis familiaris*	28 *Budorcas taxicolor*
Identified peptides	4	4	5
Incubation ratio	**1:4**	**1:4**	**1:4**
Coverage % and identification	22 *Canis familiaris*	19 *Canis familiaris*	30 *Budorcas taxicolor*
Identified peptides	4	4	6

- *Protein mix*

The MS identification parameters of the protein mixture containing αg, βg and Mb, incubated for 24 h under static conditions at pH 7 prior and after incubation at 1:1, 1:2, 1:4 ratios with total HS, and after N-PAGE are reported in Table 3.6. Also α-amylase was detected in the protein mixture because of commercial preparation impurity.

Table 3.6 Identification parameters (MASCOT algorithm) of the **αg, βg and Mb,** incubated as a protein mixture, in contact or not with HS at various ratios, and after **N-PAGE.** Colored boxes indicate the protein recognized with highest score. (NI = not identified)

Ratios	**Mb**	**αg**	**α-amylase**	**βg**
	Cov %	**Cov %**	**Cov %**	**Cov %**
No (1)	99	59	18	3
No (2)	99	57	18	NI
No (3)	99	62	23	NI
1:1 (1)	99	64	23	8
1:1 (2)	99	55	20	6
1:1 (3)	99	55	20	6
1:2 (1)	99	53	23	4
1:2 (2)	99	41	14	4
1:2 (3)	99	44	18	4
1:4 (1)	98	47	14	NI
1:4 (2)	Bad ESI file, damaged			
1:4 (3)	94	34	10	8

- *αg and βg vs HS substances fraction (HA, FA)*

The MS identification parameters of αg and βg incubated for 24 h under static conditions at pH 7 prior and after incubation at 1:1, 1:2, 1:4 ratios with HS, HA and FA, for 24 and 144 h, and after N-PAGE are reported in Table 3.7.

Table 3.7 Identification parameters (MASCOT algorithm) of the **βg,** incubated with HS, HA and FA at various ratios, and after **N-PAGE**

Humic substances						
	HS		**HA**		**FA**	
Incubation time	24 h	6 d	24 h	6 d	24 h	6 d
Protein	α-glucosidase *A. Niger*					
Coverage	10	11	11	23	11	23

In general, the tested proteins were identified in all types of contact experiments between proteins and HS, HA and FA, with MS identification parameters comparable to those of the respective control proteins. A high reproducibility of protein identification was also observed among the replicates, either prior or after the contact with humic substances. No relevant differences in the protein identification parameters were observed in relation to incubation conditions such as contact time, and mode (shaking vs static).
However, a general decrease in the coverage percentage upon increase of the HS proportion was observed (Tables 3.2-3.7).

3.3.2 Protein identification after the incubation in soil

- *Ag and βg control protein identification: SDS-PAGE coupled to MS*

S1, S2, S3, S4 control bands (Figure 3.6a) were identified as resumed in table 3.8.

Table 3.8 MASCOT identification results: control **αg** and **βg**, **SDS-PAGE**

Band	Protein identification	Coverage %	Identified peptides
SDS PAGE			
S1	β-glucosidase *A. Niger*	33	25
S2	β-glucosidase *A. Niger*	9	11
S3	α-glucosidase *A. niger*	26	9
S4	α-glucosidase *A. niger*	20	7

- *αg and βg vs soil: SDS-PAGE coupled to MS*

The glucosidases preparation inoculated and extracted from soil was loaded on SDS-PAGE. The gel bands labelled in Figure 3.6a were identified by MS analysis; table 3.9 shows the identification results.

Table 3.9 MASCOT identification results: **αg** and **βg** incubated in soil, **SDS-PAGE**

SDS-PAGE bands	Protein identification	Coverage %	Identified peptides
24 h			
S1	β-glucosidase of *A. niger*	16	10
S2	β-glucosidase of *A. niger*	12	4
S3	α-glucosidase *A. niger*	6	3
S4	α-glucosidase *A. niger*	3	1
6 days			
S1	β-glucosidase of *A. niger*	30	22
S2	β-glucosidase of *A. niger*	9	6
S3	α-glucosidase *A. kawachii*	32	12
S4	α-glucosidase *A. niger*	25	9

- *Myoglobin*

After 24 h of incubation period, the extraction protocol did not recover appreciable amount of protein; in figure 3.6b it is possible to observe just bands due to natural proteins present in soil and extracted by the applied protocol.

3.4 PEP 3D analysis

Representative images generated by PEP 3D software to evaluate sample quality are reported in annex I (p 131-146). Ulterior applications of this software, such as % of identified peptides upon the precursor ions selected by the instrument for the identification (CID), incurred in several problems due to damaged ESI files and

occurrence of unknown errors during the processing. For this reason only Mb and protein mix images are reported (table 4.1 and 4.2).

Table 4.1 Mb, Peptide/CID calculated by PEP 3D

Sample	Peptide/CID (%)
Control Mb (1)	4.5
Control Mb (2)	4.1
Control Mb (3)	0.1
Mb:HS 1:1 (1)	4.7
Mb:HS 1:1 (2)	6.6
Mb:HS 1:1 (3)	4.9
Mb:HS 1:2 (1)	Processing failed
Mb:HS 1:2 (2)	Processing failed
Mb:HS 1:2 (3)	Processing failed
Mb:HS 1:4 (1)	Processing failed
Mb:HS 1:4 (2)	Processing failed
Mb:HS 1:4 (3)	Processing failed

Table 4.2 Protein mix, Peptide/CID calculated by PEP 3D

Sample	Peptide/CID (%)
Control Mix (1)	17
Control Mix (2)	12
Control Mix (3)	15
Mix:HS 1:1 (1)	13
Mix:HS 1:1 (2)	17
Mix:HS 1:1 (3)	12
Mix:HS 1:2 (1)	15
Mix:HS 1:2 (2)	Unreadable file

Mix:HS 1:2 (3)	Unreadable file
Mix:HS 1:4 (1)	10
Mix:HS 1:4 (2)	Unreadable file
Mix:HS 1:4 (3)	Unreadable file

3.5 Nuclear magnetic resonance (NMR) protein-humic acids contact experiments

The NMR experiments were performed to further understand the establishment of interactions between the a protein produced by a soil bacterium CopH and HA extracted from a volcanic soil.

3.5.1 ^{1}H NMR spectrum of the control CopH ^{1}H CopH control protein

The ^{1}H NMR spectrum of CopH and the buckets, obtained dividing the spectrum in several regions, are reported in Figure 5.1 and tab 5.1 respectively.

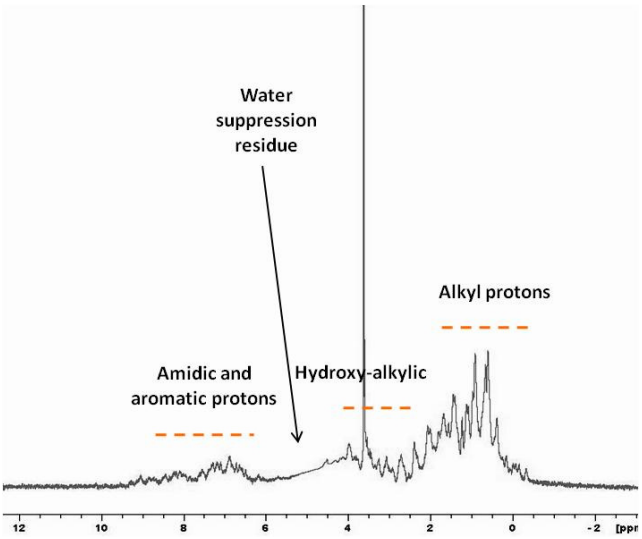


Figure 5.1 ^{1}H CopH spectrum

Amidic and aromatic regions were from 9.1 to 6.4 ppm. Slightly deshielded protons region were from 4 to 1.9 ppm, and the alkyl region was from 1.9 to 0.5

ppm. The CopH spectrum evaluation was performed not for single signal but associating proton intervals to numbered bucket areas. The spectrum was divided in the following fifteen discrete regions: 1-6 Amidic and aromatic, 7-10 slightly deshielded protons, 11-15 alkyl protons. Detailed results for the fifteen discrete regions are reported in Table 5.1.

Table 5.1. Control NMR spectral data for CopH at concentration of 8 mg ml^{-1} at pH value 7 buckets

Object	from (ppm)	to (ppm)
1	9.199	8.995
2	8.384	7.898
3	7.691	7.468
4	7.468	7.246
5	6.986	6.764
6	6.759	6.423
7	4.063	3.904
8	2.844	2.58
9	2.504	2.254
10	2.227	1.944
11	1.886	1.537
12	1.533	1.292
13	1.196	1.037
14	1.032	0.832
15	0.799	0.506

3.5.2 Relaxation times analysis

The T1 and T2 relaxation times were calculated by integrating the whole area under each bucket are reported in Table 5.2.

Table 5.2. T1 and T2 relaxation of control protein, protein incubated with HA in ratio 1:0.5 and protein incubated with HA in ratio 1:1

				Protein:HA incubation ratios				
Bucket	***from***	***to***	**Control**	**1:0.5**	**1:1**	**Control**	**1:0.5**	**1:1**
1	*9.199*	*8.995*	0.38	0.33	0.17	0.055	0.031	0.016
2	*8.384*	*7.898*	0.44	0.24	0.30	0.051	0.032	0.043
3	*7.691*	*7.468*	0.50	0.23	0.36	0.061	0.039	0.038
4	*7.468*	*7.246*	0.47	0.27	0.30	0.055	0.032	0.038
5	*6.986*	*6.764*	0.41	0.29	0.30	0.063	0.039	0.035
6	*6.759*	*6.423*	0.34	0.28	0.29	0.077	0.040	0.041
7	*4.063*	*3.904*	0.28	0.25	0.30	0.036	0.027	0.025
8	*2.844*	*2.58*	0.37	0.31	0.33	0.034	0.028	0.025
9	*2.504*	*2.254*	0.36	0.32	0.32	0.044	0.038	0.031
10	*2.227*	*1.944*	0.34	0.29	0.29	0.038	0.028	0.026
11	*1.886*	*1.537*	0.31	0.27	0.25	0.032	0.023	0.021
12	*1.533*	*1.292*	0.30	0.26	0.23	0.046	0.033	0.027
13	*1.196*	*1.037*	0.33	0.28	0.26	0.050	0.036	0.034
14	*1.032*	*0.832*	0.34	0.30	0.23	0.056	0.044	0.034
15	*0.799*	*0.506*	0.34	0.31	0.28	0.041	0.034	0.026

3.5.3 CopH + HA (1:1) spectra, signal broadening analysis

After the incubation with HA, signals broadening in spectra of control CopH protein and CopH incubated with HA for 24 h at 1:1 ratio and pH value of 7 were evaluated. The differences of the NMR spectra of the alkyl region are reported Figure 5.2.

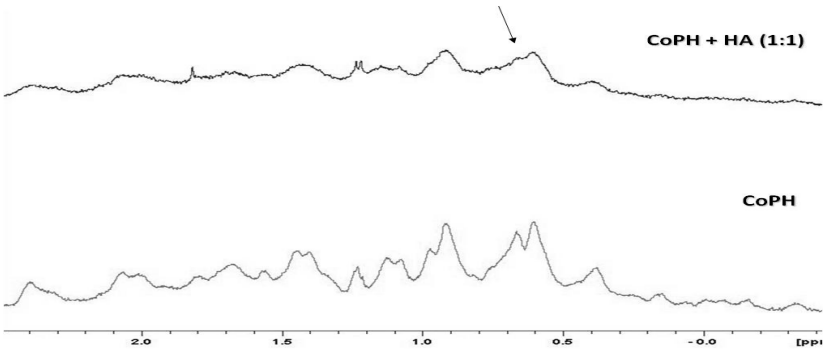


Figure 5.2. Alkyl region of CopH protein incubated with HA (top) for 24h at pH 7 or not incubated (bottom).

The differencs of the NMR spectra of the slightly deshielded protons region are reported Figure 5.3.

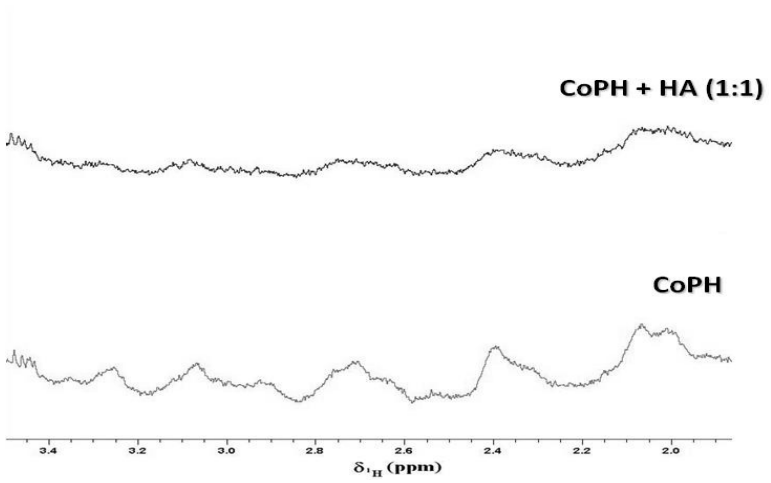


Figure 5.3. Slightly deshielded protons region of CopH protein incubated with HA (top) for 24 h at pH 7 or not incubated (bottom)

The differencs of the NMR spectra of the amidic and aromatic protons region are reported Figure 5.4.

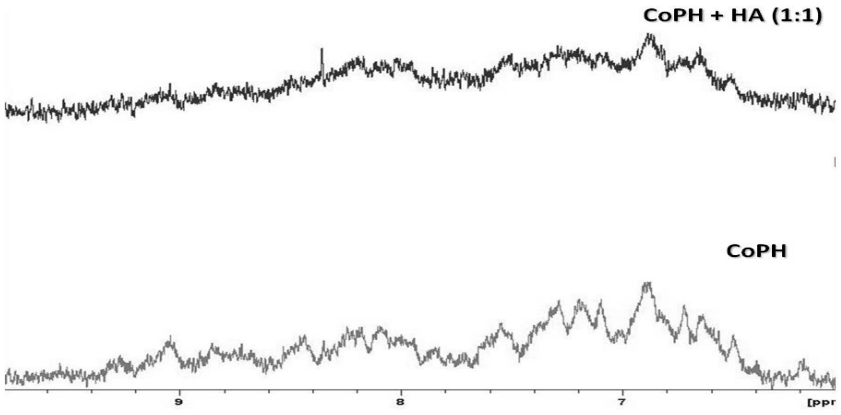


Figure 5.4. Slightly deshielded protons region of CopH protein incubated with HA (top) for 24h at pH 7 or not incubated (bottom)

Residual paramagnetic metals was not detected thanks to a careful purification of the HA and high purity of the protein CopH (>99%).

3.5.4 Correlation times

Correlation times for H nuclei, calculated from T1 and T2, are reported in Table 5.3. Both T1 and T2 values in solution are dependent on the correlation time (τ_C), that is defined as the effective average time needed for a molecule to rotate through one radian. Therefore, the larger the τC values, the slower the molecular motion.

Table 5.3. ^{1}H correlation times (τ_C) of CopH control protein, and CopH protein incubated with HA at 1:0.5 and 1:1 protein:HA ratios

	Protein:HA incubation ratios		
Bucket	**Control**	**1:0.5**	**1:1**
1	1.06	1.36	1.40
2	1.22	1.12	1.06
3	1.18	0.97	1.28
4	1.22	1.20	1.14
5	1.02	1.12	1.20
6	0.77	1.06	1.08
7	1.14	1.26	1.44
8	1.39	1.40	1.52
9	1.17	1.18	1.33
10	1.24	1.36	1.39
11	1.30	1.43	1.44
12	1.02	1.15	1.20
13	1.02	1.16	1.13
14	0.97	1.05	1.05
15	1.19	1.27	1.37

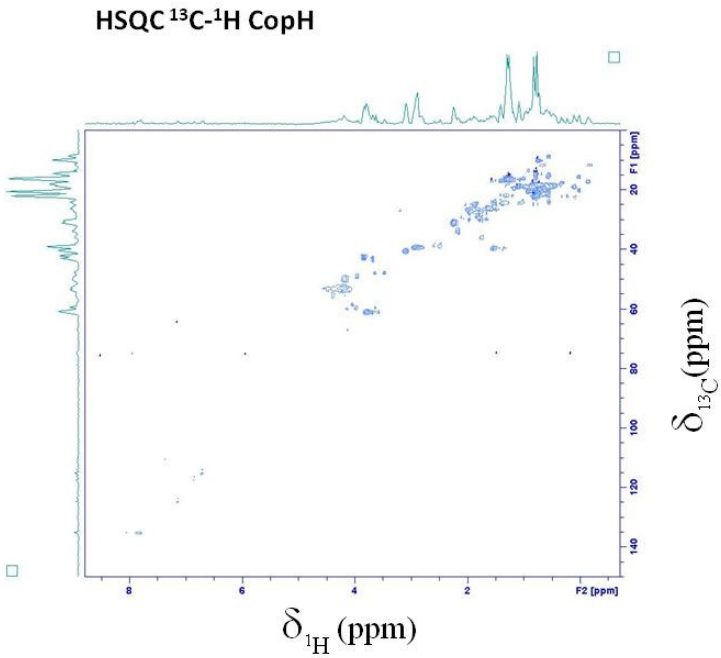
3.5.5 DOSY (Diffusion Ordered SpectroscopY)

The NMR diffusion experiments were conducted to measure possible changes in CopH diffusivity after the contact with HA. The low reproducibility between the replicates hampered the reliability of the experiment.

3.5.6 Heteronuclear Single Quantum Correlation (HSQC) of ^{13}C and ^{15}N CopH spectra

The 2D spectra for the HSQC ^{13}C-^{1}H and ^{15}N-^{1}H HSQC for the CopH protein are reported in Figure 4.5. This type of analysis showed no significant differences between the ^{13}C and ^{15}N CopH HSQC spectra of the CopH control protein and protein incubated with HA.

a)



b)

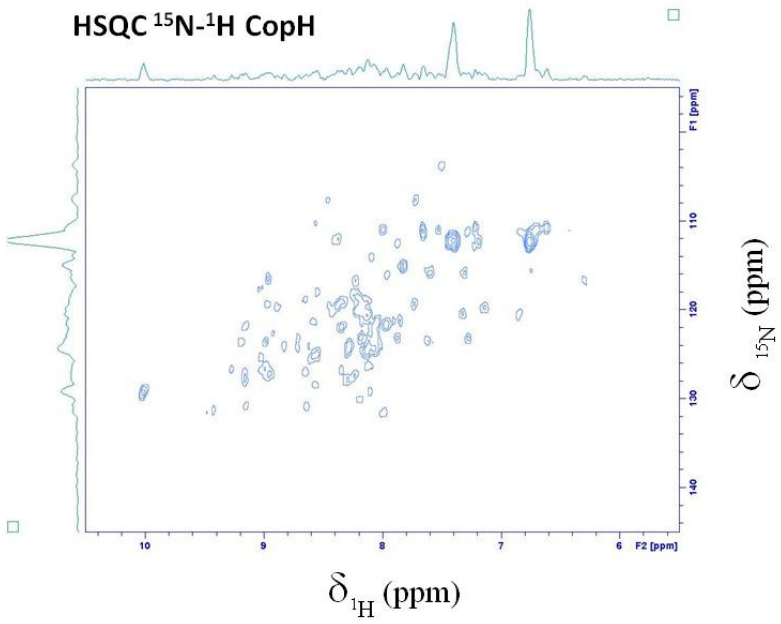


Figure. 5.5. CopH 2D spectra. a) HSQC ^{13}C-^{1}H; b) ^{15}N-^{1}H HSQC

No variations in chemical shift were observed regardless of the tested incubation conditions. For this reason no further detailed analyses were conducted on the spectra.

Chapter 4: Discussion

The chemical properties and high variability of the proteins and the still unclear composition of the HS have so far limited our knowledge about the potential interactions between proteins and HS, especially in the soil environment. This project aimed to determine the establishment of chemical interactions between proteins and HS, relying on the hypothesis that HS are supramolecular associations of relatively low molecular weight organic molecules. To address the main aims of the project five proteins with distinct molecular properties were analyzed before and after contact with HS and their FA and HA fractions and prior and after inoculation into soil.

In the present work quartz microbalance device, used to evaluate the establishing of interaction between the βg from *A. niger* and the commercial HA from Sigma, provided no reproducible results due to false positives. In particular, a decrease in the vibration frequency, even in control samples (only HA), was measured in all cases. A possible explanation of the false positive results could be the reaction of HA with the ethanolamine layer used to block the free COOH groups of undecanoic acid, employed as bridge to link the enzyme. This could had also led to some HA precipitation due to partial HA solubilization, causing the decrease in vibration frequency. Studies on protein-HS interactions based on the use of the microbalance have proven the suitability of this technique (Tomaszewki et al. 2011, 2012, Sander et al. 2012), and showed a strong adsorption of the protein Cry1 Ab to the SOM. A main difference between our study and the quoted ones, was that in our case the βg protein was added as first partner molecule to establish a specific protein orientation, exposing the C-term free for reacting with HS, and to test the presence of low molecular weight organic molecules (LMWOLs) from the cHA with affinity for specific βg domains, whether in the quoted studies the HS were immobilized first, then the proteins could react with HS.

Potentials and limitations of various chemosensors were recently reviewed by Puglisi et al. (2012, see annex 2 of this thesis).

Protein electrophoretical studies date back to 1950s and have allowed the determination of main protein molecular properties such as weight, net electric charge, and denaturation. The use of charged surfactants able to bind and denature proteins such as SDS also allowed to identify some protein properties, due to studies on the formation of protein-surfactant interactions.

Electrophoretic results in our project showed no detectable changes in proteins electrophoretic mobility, either in SDS- or N-PAGE, and under different contact conditions such as pH, temperature, time and static *vs* dynamic incubation modes. Protein net charge is the main factor affecting the migration (i.e. the apparent molecular weight) of proteins during N-PAGE, whereas the peptide mass is the main factor influencing the peptide mobility in SDS-PAGE. An important aspect to consider is that interactions between the tested proteins and the SDS, which is in turn affected by the ionic strength and chemical composition of the solutions, could also have influenced as these factors can reduce the electrostatic repulsion between the charged groups (Giehm et al., 2010). The N-PAGE has been widely used to study protein-protein interactions (Krause 2006), therefore, no changes in electrophoretical mobility of the tested proteins in N-PAGE, upon changes in solution pH and HS concentration, could be interpreted as no protein interactions with other molecules, or that protein interactions with other molecules did not change the protein electrophoretic mobility as compared to the control protein. In the case of SDS-PAGE, no differences in protein mobility, could be explained by the disruption of possible interactions because of the denaturation treatment. Overall, using the electrophoresis technique no changes in the molecular conformation and molecular weight could be detected after contact with HS. Further confirmation of the absence of protein substantial modifications after contact with HS were provided by the in-gel enzyme β-glucosidase activity in N-PAGE, which was not inhibited by any of the treatments. This result is in agreement with those obtained from enzymes-surfactants contact experiments. For example, it was reported catalytic activation of microbial β-glycosidase,

catalase, glucose-6-phosphatase, phospholipases in the presence of low concentrations of different surfactants such as SDS, Triton X-100 and deoxycholate glucose-6-phosphatase (D'Auria et al 1997; Muga et al., 1993; Beyhlm 1987). The lack of βg inhibition after the contact with HS (Figure 3.5) also implied that the soil-borne HS did not contain enzymatic inhibitors, as it is known that this enzyme activity can undergo to competitive and non-competitive inhibition by various mono and oligosaccharides (Melo et al. 2006).

The protein MS data revealed that protein identification was affected by the contact with HS in terms of decrease in protein coverage % (percent of the protein sequence represented by the detected peptides in the dataset) and number of identified peptides (peptides identified by MS and matched to identify the protein). The protein MS and MS/MS analysis highlighted that this phenomenon could be due to protein-HS interactions at molecular level, explained by the HS supramolecular model proposed by Piccolo (2001). If the HS are associations of LMWOLs stabilized by weak dispersive forces such as hydrophobic interactions and H-bonds, such relatively weak intermolecular forces may be destabilized by the chemical affinity of HS-borne organic ligands for specific protein domains. Binding of small humic molecules to proteins, while not causing detectable changes in the protein electrophoretic mobility, could cause the observed decrease in protein coverage % by MS due to shifts in the *m/z* ratio.

The used MS analytical approach showed that probably non-covalent interactions between proteins and HS-borne could cause a reduced peptide coverage patterns for all the studied proteins to a higher extend at higher protein:HS ratios. The effect was particularly evident for the Mb owing to its relatively small molecular dimensions and structure, whereas lower effects on the peptide coverage % was observed for αg and βg, likely due to their larger molecular weight and greater structural complexity. The αg and βg, in particular, are relatively large and highly glycosylated proteins, condition that could modulate and affect intermolecular interactions (Shental-Bechor et al. 2009). The ESI-MS results by protein mixture (αg, βg, Mb) did not show significant differences among control proteins and

proteins incubated with HS at various ratios. The αg and also α-amylase (detected in the commercial glucosidase preparation from *A. niger* as an impurity) showed a decrease in coverage % of 30% and 40% respectively, as compared to their controls, more than the Mb protein. A possible explanation of this result could reside in the increased number of affinity domains of αg and α-amylase that could actively compete with Mb for the HS-borne LMWOLs.

"HS effect" was detected also for the RecPrp protein regardless of its peculiar conformation, particularly at the highest incubation ratios, with a decrease of about 20% in coverage % observed for proteins after N-PAGE electrophoresis. Moreover, while the control RecPrp protein was consistently identified as from *Budorcas taxicolor*, it was misidentified after contact with HS as other mammalian species (*Canis familiaris, Capra hircus, Antilope cervicapra*) that differs in aminoacid sequences from few to several positions.

These results are important because soil is no doubt a reservoir environmental compartment for prion proteins, but very little is known about the specific between prion proteins and SOM, particularly HS, and the importance of the presented results is even more interesting due to the fact that HS are ubiquitous in nature and mostly of unknown structure. It is important to underline that the tested RecPrp protein shares the same aminoacid sequence as the human pathogenic *Ovis* prion protein.

In the last decade, MS has been used for studying non-covalently bound complexes thanks to "soft" ionization methods, such as electrospray ionization (ESI) and matrix assisted laser desorption (MALDI), that can transfer non-covalently bound molecules into the gas phase. Most of the studies have concerned interactions between drugs active ingredients and specific proteins (e.g. cell receptors) or protein-protein interactions (Preston et al. 2012). Pramanik et al. (1998) studied the detection of non-covalent complexes of a protein-ligand, an inhibitor-protein-ligand, a protein-protein and a protein-metal complex using electrospray ionization-mass spectrometry (ESI-MS) technology, and reported

determinations of mass accuracy of better than 0.01% achieved for protein complexes in the range molecular 19-34 kDa, which is far superior to electrophoresis and chromatographic methods. Li et al. (1994) compared the relative binding energies of rapamycin and four of its analogs to the cytoplasmic receptor FKBP and reported that the gas-phase binding reflects the aqueous solution behavior in these complexes. The use of ESI method for studying non-covalent interactions has been reviewed by Loo (1997), and Daniel et al. (2002); the technique is flexible enough to allow the generation of ions from large, nonvolatile analytes such as proteins and peptides without significant analyte fragmentation (Aebersold and Goodlett 2001).

In this PhD project, additional information on possible interactions between proteins and HS-borne LMWOLs was obtained by the comparison of MALDI and ESI spectra of Mb. In fact, the mass shift of 104 Da mass of the peptide with m/z 2601, corresponding to the first missing peptide of the Mb sequence (Figures 3.7, 3.8) detected in the MALDI MS analysis, was compatible with the establishment of a chemical association between this specific peptide with an unknown LMWOL, which altered its *m/z* value. Drastic changes in the MALDI spectra upon increasing HS the protein:HS ratio could also be explained by similar associative mechanisms, thus providing some 'labeling' of protein chemically reactive domains towards the used HS, and possibly explaining the dramatic reduction of Mb ESI based identification after contact with HS.

Generally, MALDI is considered a harsher ionization mode than ESI, but our results showed that some non-covalent interactions could be stronger than the laser desorption/ionization events. The relatively high variability in the MALDI Mb spectra and in peptide coverage % among different Mb analytical replicates (Figure 3.8) was in the expected range for this technique, but allowed to confirm the dramatic reduction of ESI-MS based Mb identification after contact with HS.

The Kyte-Doolittle plots (p. 150-152, annex I), compareded with protein coverage % sequences (p. 131-146, annex I) revealed that first peptides missed in the identification of proteins after contact with lower amount of HS were mainly

hydrophobic. This result pointed towards the fact that molecular interactions between proteins and HS could involve hydrophobic interactions, at least in the early stages, indicating that the potential protein binding domains for HS-borne ligands could be both internal and external, and that HS-borne LMWOLs could reach protein inner sphere. There is increasing evidences that interactions between proteins and humic substances are mainly of hydrophobic nature (De Kruif et al., 2004). Such hydrophobic interactions may be also due to dehydration which may destabilize and partially disassembly humic substances in contact with the proteins, as hypothesized by Tomaszewski et al. (2011). The fact that the four proteins studied by MS showed similar trends was also compatible with the importance of hydrophobic interactions. The, αg and βg had isoelectric points of 4.25 and 4.65, respectively, different from Mb (pI 7.2) and RecPrP (pI 9.47) and for this reason, theoretically, they should have different interaction potentials towards HS, being more negatively charged at pH 7. However, hydrophobic interactions are relatively independent on the net charge of the partner molecules, and thus proteins could interact with humic substances also at pH values above the protein isoelectric point, when both proteins and HS had a net negative charge. De Kruif et al. (2004) reported that such forces could overcompensate electrostatic repulsion of the negatively charged αg or βg and HS at pH 7. However, thermodynamically favorable hydrophobic interactions following electrostatic complexation between lysozyme and synthetic humic acids were reported by Tan et al. (2008).

Another important aspect to take into account for the evaluation of the obtained results is that all interaction energy equations contain a term (1/ε) that considers the characteristics of the medium in which interactions are established. This is an important aspect because interaction energies based on charges, dipoles and other molecular characteristics are expected to increase from a solution into a gas phase. Hydrophobic and hydrophilic interactions and H-bonds, all believed

involved in protein-HS interaction mechanism, are strongly solvent-dependent and in its absence (MS gas-phase) the interaction is substantially weaker.

Hydrophobic interactions between proteins and HS-borne LMWOLs could be accompanied by entropy gain caused by conformational changes of the protein (Haynes and Norde, 1995). However, some small conformational changes may be not detectable by the electrophoresis as the organic ligands may bind high affinity protein sites with weak interactions, without profound influences on the protein tertiary structure.

Hydrophobic ligands may therefore have little impact on the overall protein conformational stability, and this may explain why contact with HS did not inhibit the catalytic activity of βg proteins in N-PAGE (Figure 3.6). The ESI-MS results from samples obtained by SDS-PAGE provided more information about the possible interactions established between the model proteins and the HS. In this case, the MS protein coverage % values and number of identified peptides for each protein were higher than the corresponding sample run in N-PAGE (Tables from 3.1 to 3.5). The fact that most of the analyzed proteins showed similar coverage % and number of identified peptides as the control proteins, could be explained by the fact that protein denaturation led to the destruction of the protein binding domains or to some competition between HS-borne LMWOLs and SDS . However, some effects were still detected in some of the αg and βg replicates and the RecPrP protein was identified as belonging to other mammalian species, as compared to the respective control proteins.

Although the MS technique has been proven to be reliable for studying supramolecular protein interactions, to my knowledge this is the first time this analytical approach is used for studying protein interactions with soil-borne HS. Therefore, in order to evaluate the possible artifacts formation during the analysis (e.g. co-elution of precursors because of non optimal chromatographic gradient, presence of contaminants) the ESI-MS signals were evaluated using the PEP 3D tool of the TPP software. The PEP 3D tool permitted to optimize the performance of the used Orbitrap LC ESI- MS instrument, to evaluate the signal quality, and

also to check reproducibility and consistency of analysis. PEP 3D allowed also to visualize the precursor ions automatically selected by the instrument for identification and checking which precursor ions are recognized and assigned in protein sequence. Despite the problems occurred during this type of investigation, it was pobbile to check that in the protein mix the percentage of assigned precursors decreased in samples incubated with higher proportions of HS (tabs 4.1 and 4.2 Results, p. 11 annex I), in agreement with the above described reduction in peptide coverage % during the mass MS analysis. The change in *m/z* ratio supposedly due to the binding of HS-borne LMWOLs, likely hampered the assignment of the precursor ions to specific protein sequences. In other cases, the software was not able to provide this information, giving *failure messages,* especially for samples incubated in higher ratios with HS. The current interpretation of these results is that a bad quality of the ESI results produce the failure of the software test. However, these considerations need further analysis.

The MS results obtained after protein incubation into soil were in line with those obtained from the protein-HS incubation experiments. In fact, after 1 day of incubation a significant decrease in protein coverage % was observed with respect to the controls and samples incubated with HS. However, after 6 days of incubation, coverage % values closer to those of the control proteins were obtained. Despite the extensive literature on the subject, these results highlights crucial gaps in our knowledge about protein fate and turnover in soil. Interpretation of these results are not straightforward. Possibly, after the initial protein-HS interactions, (e.g. hydrophobic/hydrophilic interactions, H-bonds), other factors (e.g. soil moisture, microbial activity) predominated. Moreover, good proteins identification by MS could be due to the fact that a relatively high amount of protein was incubated into soil, and some further protein concentration during extraction with protocol by Doni et al. 2012 (see annex II) could have occurred. These results may help to interpret some of the results in soil proteomics, in which the concentration of proteins in soil is impossible to exactly

quantify (Roberts and Jones et al. 2008). This factor along with the possible interferences due to protein-HS interactions identified in this study may be responsible for the low protein identification rates in most of the soil proteomic studies. Possibly, the Mb could be not detected after incubation into soil (Figure 3.6) because less resistant to soil proteases. It is long known that soil has a strong proteolytic activity (Zeller et al. 1991; Renella et al. 2002), but αg and βg from A. niger were likely more persistent in soil because reach in glycosylic groups (Abdel-Naby et al. 1999; Rudick et al. 1979). Glycosylation plays a number of roles, including prevent aggregation, intrinsic stabilization of protein structure and increasing the lifetime of proteins by conferring resistance to proteolytic degradation (Varki et al. 1993, Shental-Bechor et al. 2009) that could explain the resistance in soil (Nannipieri et al. 2012).

ESI-MS investigations resulted reliable and reproducible for probing difficult biological problems as the study of the interaction between proteins, the most versatile biopolymers, and HS, ubiquitous and nearly unknown molecules.

The NMR technique has been used for studying the HS chemical properties since the 1960s and although being a non routine method, implementation of applications using ^{1}H, ^{13}C, ^{31}P, and ^{15}N-labelled organics either in liquid or solid-state, have allowed to assess the effects of several soil and environmental conditions on the HS chemical composition and properties (Preston 1996).

The NMR results confirmed the establishing of interactions between the CopH protein and the HA extracted and purified from a volcanic soil. The strongest evidences were obtained by ^{1}H-NMR spectra analysis (signal broadening, relaxation and correlation times). In fact, a progressive signal broadening of CopH ^{1}H-NMR spectra was observed upon increasing the CopH:HA ratio the in alkyl, hydroxyl-alkyl, aromatic and amid proton regions (Figures 3.8-3.10, Table 3.9). This phenomenon is related to the reduction of Brownian motion of CopH following the complexation with HA. The signal broadening is due to a less efficient minimization of molecular dipolar couplings. The width of NMR signals

is inversely proportional to the spin-spin relaxation times, that are progressively reduced when molecular mobility is diminished (Bakhmutov, 2004). In our NMR experiments, the reduction in mobility of CopH could be interpreted as the formation of non-covalent interaction with HA. Mazzei and Piccolo (2012) reported a significant ^{1}H signal broadening after the contact of the herbicide glyphosate with fulvic acids. A progressive decrease of spin-lattice (T1) and spin-spin (T2) relaxation times was observed for CopH after the contact with the highest proportions of HA, with the decrease in both T1 and T2 up to 85% in some buckets. Spin-lattice (T1) and spin-spin (T2) relaxation times are NMR parameters related to changes in protein flexibility, and follow perturbations of the magnetic field on a studied nucleus (Zhang and Forman-Kay 1995, Tollinger et al. 2001). The measurement of relaxation times has proven to be useful to examine the overall motion of proteins or protein domains involved in formation of complexes with other molecules (Pickford and Campbell, 2004) and, the detected changes in T1 and T2 relaxation times of CopH with increasing amount of HA, confirmed the formation of the non-covalent complexes.

The addition of progressive amount of HA to CopH resulted in an increase of the protein correlation time τC values in all spectral buckets, with few exceptions (Table 3.10), confirming a reduced molecular mobility for CopH in the presence of HA, as compared to the free protein. The CopH relaxation times in solution are dependent on the protein correlation time (τC), which is the average time needed for a nuclear spin to rotate through one radian (Carper and Keller, 1997). Therefore, the larger the τC values the slower the molecular motion (Bakhmutov, 2004). The correlation times have been used as qualitative indexes to indicate the changes in molecular rigidity of host-guest complexes between humic substances and various organic pollutants (Smejkalova and Piccolo 2008).

The results on CopH signals broadening, relaxation and correlation times indicated that the protein-HA interactions were governed by weak bonding forces, to which the CopH hydroxy-alkyl and aromatic protons were mostly responsible.

The hydrophilicity of the used HA purified by the volcanic soil implies the participation of its OH- and COOH- functional groups to the H-bonds formed with complementary hydroxy-alkylic components in CopH. On the other hand, the involvement of aromatic protons in CopH-HA interactions may be attributed to π-π hydrophobic bonds between the phenolic molecules in HA and the aromatic residues in CopH. The chemical characteristics of the used HA could explain why the ^{13}C and ^{15}N spectra did not reveal any detectable changes.

Overall, the findings from this project indicate that hydrophobic interactions between HS-borne LMWOLs and proteins occurred independently from the protein charge, and this may in turn affect protein affinity for the soil solid phases or change the protein aggregation properties while not resulting in loss of protein functional activity (e.g. catalysis). Seen the importance of extracellular proteins for soil ecological functionality, these findings point towards the importance of the soil humic substances in stabilization of extracellular enzymes (Nannipieri 1996; Benitez et al. 2000). Moreover, deeper understanding of protein-HS interactions may improve the extraction and purification protocols for soil proteomics, as low protein identification rates, compared to the potential soil protein diversity, are generally reported in soil proteomic papers.

For example, Wang et al. (2011) reported that one-third of the protein spots could not be identified by MS in their rhizosphere metaproteomic study, and Wu et al. (2011) reported that only six bacterial proteins and one fungal protein were differentially expressed in their *R. glutinosa* rhizosphere proteomic study. By using off-gel LC MS/MS tecnhiques, Schulze et al. (2005) reported that only 75 proteins could be identified, and more recently Chourey et al. (2010) identified 716 redundant and 333 non-redundant proteins belonging to characterized soil microorganisms. Even though soils display a strong proteolytic activity (Renella et al., 2002), proteolysis as well as chemical denaturation may be only partially responsible for the low rate of protein recovery and identification, also by considering that lysis buffers contain protease inhibitors. Low protein

identification rates could be due to their interactions with HS leading to their incorrect identification, by a further analysis of the published MS spectra it is possible deducing that a major challenge in soil proteomics is the correct peptide assignment, due to the low quality of spectra. However, in all models, all the mass-to-charge ratios (m/z) peaks are mathematically assumed to be 'independent', but this is not always the case in MS data. Therefore, the results depend on clean spectra and well-fitting data, with problems in assigning peak probability to ion fragments having low quality spectra. Although different MS manufacturers may require site specific optimal set up, uniformity in the MS spectra analysis may be useful, especially for soil proteomics due to the high diversity of protein sources and potentials of protein extracellular modification or re-arrangements.

Soil-induced protein modifications should be identified and techniques comparable to those commonly adopted in the study of post-translational protein modifications (e.g. phosphorylation, glycosylation), such as enrichment of phosphorylated proteins and peptides and MS set up for protein identification (Mann et al. 2002) may be used in soil proteomics. In this view, the use of MS signal analysis (e.g. TPP) and the analysis of collision-induced dissociation (CID) may lead to the identification of modified ions which may serve as 'reporter ions' for preferential HS-protein interactions.

4.5 Needs and gaps

The obtained results concerned a limited range of protein, HS and solvents, and because of the actual chemical environment, where protein-HS interactions can strongly influence the supramolecular structure (e.g. H-bonding, ionic strength), these findings need to be confirmed with further studies involving more solvents and other chemical variables.

The obtained results should be therefore considered preliminary and can not be generalized at this stage. Although the tested proteins were representative of

various protein types and the extreme protein diversity in terms of hydrophobicity, net charge, secondary structure, that determine the type of reactive sites towards potential HS-borne ligands at given chemical conditions (e.g. pH value, ionic strength, temperature, concentration), requires more research on this aspect. Moreover, most of knowledge on the protein SOM interactions, including those from this project, results from oversimplified *in vitro* experiments; formation of chemical bonds can occur between negatively charged protein and SOM functional groups thanks to divalent cations (e.g. Mg, Ca) present in the soil solution, as reported by Yuan and Zydney (2000).

The experiments in aqueous solutions could influence the protein-HS interactions, particularly because the surrounding H_2O molecules could have 'forced' hydrophobic protein-HS binding as an 'hydrophobic effect' (Southall et al. 2002).

More detailed NMR analyses will be also needed to understand the distribution of polar and non-polar components in the used HS and HA, because this can influence the HS aggregation or dispersion in solution.

Chapter 5: Conclusions

The presented data suggest that the presence of humic molecules altered protein MS identification in term of decrease in coverage % upon increasing HS concentrations, due to missing identification of either hydrophobic or hydrophilic peptides and NMR ^{1}H spectra as confirmed by relaxation times, line broadening analysis and correlation times.

The findings by electrophoresis, MALDI, ESI and NMR suggest that the investigated proteins establish in aqueous solution weak interactions with small humic molecules, whose molecular weight is not enough to cause electrophoeretic delay but changes peptides *m/z* values in ESI-MS and as consequence the peptide identification in protein sequence. The formation of non-covalent complexes is supported by the results comparison of the two kind of ionization methods, ESI, that transfers non-covalently bound linkages into gas phase, and MALDI, that in some cases does not maintain weak interactions. Mass shifts have been observed likely due to binding of unknown ligands originated by HS supramolecular structure destabilization. Also NMR results confirm that CopH established with HS interactions such as van der Waals, π-π and H-bonds. The formation of CopH-HA complexes determine the broadening of ^{1}H CopH signal with increasing HA concentration due the progressively limited mobility of the protein. The reduction of the translation and rotational motion of the protein in the loose newly-formed humic-protein complexes were also shown by the changes of relaxation (T1, T2) and correlation (τc) times as measured from ^{1}H NMR spectra for the protein progressively added with HA.

Protein – HS interactions may be one of the key factors controlling protein fate in the environment. These findings may explain the current limitations in the development of soil proteomics (limited protein identification with the best possible analytical set up). Protein-HS interactions may be one of the factors limiting our knowledge on the protein presence in the environment, this is particularly critical in the case of patogenic proteins.

References

Abdel-Naby M.A., Osman M.Y., Abdel-Fattah A.F. (1999) Purification and properties of three cellobiases from Aspergillus niger A20. Appl Biochem Biotechnol 76, 33-44

Aebersold R, Goodlett D. (2001) Mass spectrometry in proteomics. Chem Rev 101, 269-295

Aebersold R., Mann M. (2003) Mass spectrometry-based proteomics. Nature 422, 198–207

Aeschbacher M., Graf C., Schwarzenbach R.P., Sander M. (2012) Antioxidant properties of humic substances. Environ Sci Technol 46, 4916–4925

Alexandrowa L.N. (1960) The use of sodium pyrophosphate for isolating free humic substances and their organic-mineral compounds from soil. Soviet Soil Sci 2, 190–197

Andreoletti O. et al. (2002) PrP(Sc) accumulation in placentas of ewes exposed to natural scrapie: influence of foetal PrP genotype and effect on ewe-to-lamb transmission. J Gen Virol 83, 2607-2616

Artiola-Fortuny J., Fuller WH (1982) Humic substances in landfill leachates: I. Humic acid extraction and identification. J Environ Qual 11, 663–9

Atalo K., Gashe B.A. (1993) Protease production by a thermophilic Bacillus species (P-001A), which degrades various kinds of fibrous proteins. Biotechnol Lett 15, 1151-1156

Bakhmutov V.I. (2004) Pratical NMR Relaxation for chemists. John Wiley and sons, Ltd.,Chicester, West Sussex, England.

Baldock J.A., Nelson P.N. (2000) Soil organic matter. In: Summer M. (Ed.) Handbook of Soil Science. CRC Press, Boca Raton, FL, pp. 25–84

Baron M.H., Revault M., Servagent-Noinville S., Abadie J., Quiquampoix H. (1999) Chymotrypsin Adsorption on Montmorillonite: Enzymatic Activity and Kinetic FTIR Structural Analysis. J Colloid Interface Sci 214 319-332

Benitez E., Nogales R., Masciandaro G., Ceccanti B. (2000) Isolation by isoelectric focusing of humic-urease complexes from earthworm (Eisenia fetida)-processed sewage sludges. Biol Fert Soils 31, 489-493

Benndorf D., Balcke G.U., Harms H., and von Bergen M. (2007) Functional metaproteome analysis of protein extracts from contaminated soil and groundwater. ISME Journal 1, 224–234

Benndorf D., Vogt C., Jehmlich N., Schmidt Y., Thomas H., Woffendin G., Shevchenko A., Richnow H.H., von Bergen M. (2009) Improving protein extraction and separation methods for investigating the metaproteome of anaerobic benzene communities within sediments. Biodegradation 20, 737–750

Beyhlm E. (1986) Kinetics of horseradish peroxidase in aqueous micelles, IRCS Med Sci 14, 417-418

Blanchet R.M. (1958) The direct and indirect effect of humified, organic matter on the nutrition of vascular plants. Annales agronomiques 9, 499-532

Bos M.A., Nylander T. (1996) Interaction between α-Lactoglobulin and Phospholipids at the Air/Water Interface. Langmuir 12, 2791-2797

Branden C., Tooze J. (1991) Introduction to Protein Structure. New York: Garland Publishing.

Bremner J.M. (1951) A review of recent work on soil organic matter. J Soil Sci 2, 67–82

Brimecombe M.J., De Leij FAAM, Lynch J.M. (2001) Nematode community structure as a sensitive indicator of microbial perturbations induced by a genetically modified Pseudomonas fluorescens strain. Biol Fert Soils 3, 270-275

Burns R.(1986) Interaction of enzymes with soil mineral and organic colloids. In: Huang, P.M., Schnitzer, M. (Eds), Interactions of Soil Minerals with Natural Organics and Microbes, ed. Soil Sci. Soc. Am.Madison, 429-452

Calace N., Petronio B.M. (1997) Characterization of high molecular weight organic compounds in landfill leachate: humic substances. J Environ Sci Health A32(8), 2229–44

Calamai L., Lozzi I., Ristori G.G., Fusi P., Stotzky G. (2000) Interaction of catalase with montmorillonite homoionic to cations with different hydrophobicity: effect on bound enzyme activity. Soil Biol Biochem 32, 815-823

Cameron R.S., Thornton B.K., Swift R.S., Posner A.M. (1972) Molecular weight and shape of humic acid from sedimentation and diffusion measurements on fractional extracts. J Soil Sci 23, 394–408

Carper W.R., Keller C.E. (1997) Direct Determination of NMR correlation times from spin-lattice and spin-spin relaxation times. J Phys Chem A 101, 3246-3250

Ceccanti B., Calcinai M., Bonmati-Pont M., Ciardi C., Tansitano R. (1989) Molecular size distribution of soil humic substances with ionic strength. Sci Total Environ 81, 471-479

Cheshire M.V., Hayes M.H.B. (1990) Compositions, origins, structures, and reactivities of soil polysaccharides. In De Boodt, M.F., Hayes, M.H.B. and Herbillon, A. (Eds). Soil Colloids and Their Associations in Aggregates. Plenum, New York and London 307-336

Chourey K., Jansson J, VerBerkmoes N., Shah M., Chavarria K.L., Tom L.M., Brodie E.L., Hettich R.L. (2010) Direct Cellular Lysis/Protein Extraction Protocol for Soil Metaproteomics Journal of Proteome Research 9, 6615–6622

Craig O.E., Collins M.J. (2000) Digestion and Capture Immunoassay (DACIA): An improved method for the immunological detection of mineral bound protein using hydrofluoric acid and direct capture. J Immunological Met 236, 89-97

Craig O.E., Collins M.J. (2002) An improved method for the immunological detection of mineral bound protein using hydrofluoric acid and direct capture. Journal of Immunological Methods 236, 89-97

Criquet S., Farnet A.M., Ferre E. (2002) Protein measurement in forest litter. Biol Fert Soils, 35, 307–313

Dani F.R., Michelucci E., Francese S., Mastrobuoni G., Cappellozza S., La Marca G., Niccolini A., Felicioli A., Moneti G., Pelosi P.(2011) Odorant-Binding Proteins and Chemosensory Proteins in Pheromone Detection and Release in the Silkmoth Bombyx mori. Chem Sens 36, 335-344

Daniel J.M., Friess S.D., Rajagopalan S., Wendt S., Zenobi R. (2002) Quantitative determination of noncovalent binding interactions using soft ionization mass spectrometry. Int J Mass Spectrom 216, 1–27

D'Auria S., Rossi M., Nucci R., Irace G., Bismuto E. (1997) Perturbation of conformation dynamics, enzymatic activity and thermostability of β-glycosidase from archaeon *Sulfolobus solfataricus* by pH and sodium dodecyl sulfate detergent. Proteins 27, 71-79

David R. (2011) Prions: The fast and the furious. Nat Rev Mol Cell Bio 12, 278-278

De Barjac H., Frachon E. (1990) Classification of Bacillus thuringiensis strains. Entomophaga 35, 233–240

De Kruif C.G., Weinbreck F., DeVries R. (2004) Complex coacervation of proteins and anionic polysaccharides. Curr Opin Colloid Interface Sci 9, 340-349

Deutsch E.W., Mendoza L., Shteynberg D., Farrah T., Lam H., Tasman N., Sun Z., Nilsson E., Pratt B., Prazen B., Martin L., Nesvizhskii A.I., Aebersold R. (2010). A guided tour of the Trans-Proteomic Pipeline. Proteomics 10, 1150-1159

Ding F., Diao J.X., Yang X.L , Sun Y (2011) Structural analysis and binding domain of albumin complexes with natural dietary supplement humic acid. J Luminescence 131, 2244-2251

Doni S., Macci C., Peruzzi E., Arenella M., Ceccanti B., Masciandaro G. (2012) *In situ* phytoremediation of a soil historically contaminated by metals, hydrocarbons and polychlorobiphenyls J Environ Monit 14, 1383-1390

Dragunov C. C., Zhelokhovtseva H. H., Strelkova E. J. (1948) A comparative study of soil and peat humic acid, Pochvovedenie 7, 409-420

Elliott M, Ambrose EJ. (1950) Structure of Synthetic Polypeptides. Nature 165, 921

Eghiaian F., Grosclaude J., Lesceu S ., Debey P., Doublet B ., Treguer E., Rezaei H., Knossow M. (2004) Insight into the Prpc - PrPsc conversion from the structures of antibody-bound ovine prion scrapie -susceptibility variants. Pnas 101, 10254-10259

Ensminger L.E., Gieseking J.E. (1942). Resistance of clay-adsorbed proteins to proteolytic hydrolysis. Soil Sci 53, 205–209

Fan T.W.-M., Higashi R.M., Lane A.N. (2000) Chemical characterization of a chelator-treated soil humate by solutionstate multinuclear two-dimensional NMR with FTIR and pyrolysis-GCMS. Environ Sci Technol 34, 1636–1646

Fan T.W.-M., Lane A.N., Chekmenev E., Wittebort R.J., Higashi R.M. (2004) Synthesis and physico-chemical properties of peptides in soil humic substances. J Peptide Res 63, 253–264

Fischer F. and Schrader H. (1921) Brennstoff-Chem 2, 37-45

Flaig W (1966) Chemistry of humic substance in the use of isotopes in soil organic matter studies. Report of FAOIAEA Technical Meeting, Pergamon, Elmsford, NY 103-127

Fusi P., Ristori G.G., Calamai L., Stotzky G. (1989) Adsorption and binding of protein on ‘clean’ (homoionic) and dirty (coated with Fe oxyhydroxides) montmorillonite, illite and kaolinite. Soil Biol Biochem 21, 911-920

Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch, A. (2005) Protein Identification and Analysis Tools on the ExPASy Server; In: John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press. pp. 571-607

Gianfreda L., Bollag J.M. (1996) Influence of Natural and Anthropogenic Factors on Enzyme Activity in Soil. In: *Soil* Biochemistry, Stozky, G. and J.M. Bollag (Eds.). Vol. 9. Marcel Dekker, New York, ISBN: 9780824794415, pp: 123-193

Giagnoni L., Magherini F., Landi L., Taghavi S., Modesti A., Bini L., Nannipieri P., Van derlelie D., Renella G. (2011) Soil solid phases effects on the proteomic analysis of Cupriavidus metallidurans CH34 Eur J Soil Sci 62, 74–81

Giagnoni L., Magherini F., Landi L., Taghavi S., van der Lelie D., Puglia M., Bianchi L., Bini L., Nannipieri P., Renell, G., Modesti A. (2012) Soil solid phases effects on the proteomic analysis of Cupriavidus metallidurans CH34. Biol Fertil Soils 48, 425-433

Giehm L., Pinto O.C.L., Christiansen G., Pedersen J.S., Otzen D.E. (2010) SDS-Induced Fibrillation of alpha-Synuclein: An Alternative Fibrillation Pathway. J Mol Biol 401, 115-133

Gygi S.P., Rochon Y., Franza B.R., Aebersold, R. (1999) Correlation between protein and mRNA

abundance in yeast. Mol Cell Biol 19, 1720–1730.

Guminski S. (1968) Present-day views on physiological effects induced in plant organism by humic compounds. Sov Soil Sci, 1250-6

Hamence JH. (1946) The determination of auxins in soils, including a note on synthetic growth substances. Analyst 71, 111-6

Hamzehi E., Pflug W. (1981) Sorption and binding mechanism of polysaccharide cleaving soil enzymes by clay minerals. Zeitschrift fur Pflanzenernahrung und Bodenkunde 144, 505- 513

Haynes C.A., Norde W.J. (1995) Structures and stabilities of adsorbed proteins. J Colloid Interface Sci 169 313-328

Hayes M. H. B., Clapp C. E. (2001) Humic substances: Considerations of compositions, aspects of structure, and environmental influences. Soil Sci 166, 723-737

Helassa N., Quiquampoix H., Noinville S., Szponarski W., Staunton S. (2008) Adsorption and desorption of monomeric Bt (Bacillus thuringiensis) Cry1Aa toxin on montmorillonite and kaolinite. Soil Biol Biochem 41, 498–504

Hsu P.H., Hatcher P.G. (2005) New evidence for covalent coupling of peptides to humic acids based on 2D NMR spectroscopy: A means for preservation. Geochimica et Cosmochimica Acta 69, 4521-4533

Hsu P.H., Hatcher P.G. (2006) Covalent coupling of peptides to humic acids: Structural effects investigated using 2D NMR spectroscopy. Organic Geochemistry 37, 1694-1704

Huang Q., Jiang M., Li X. (1998) Effects of iron and aluminum oxides and kaolinite on adsorption and activity of invertase. Pedosphere 8 251-260

Jenny H. (1941) Factors of soil formation: a system of quantitative pedology. McGraw-Hill Book Company, NY

Johnson L.N., Lewis R.J. (2001) Structural basis for control by phosphorylation. Chem Rev 101, 2209 – 2242

Johnson C.J., Bennett J.P., Biro S.M., Duque-Velasquez J.C., Rodriguez C.M., Bessen R.A., Rocke T. E. (2011) Degradation of the Disease-Associated Prion Protein by a Serine Protease from Lichens. PlusOne 6(5): e19836. doi:10.1371/journal.pone.0019836

Kan J., Hanson T.E., Ginter J.M., Wang K., Chen F. (2005) Metaproteomic analysis of the Chesapeake Bay microbial communties. Saline Systems 1, 7-14

Keller A., Nesvizhskii A.I., Kolker E., Aebersold R. (2002) Empirical stattistical model to estimate the accuracy of peptide identifications made by MS/MS database search. Analytical Chemistry 74, 5383-5392

Keller B., Simpson A. (2006) Humic Substances in Soils: Are They Really Chemically Distinct? Environ Sci Technol 40, 4605-4611

Keller M., Hettich R. (2009) Environmental Proteomics: a Paradigm Shift in Characterizing Microbial Activities at the Molecular Level. Microbiol Mol Biol R 72, 62-+

Killham K. (1994). Soil Ecology. Cambridge University Press, Cambridge

Kim K.H., Brown K.M., Harris P.V., Langston J.A., Cherry J.R. (2007) A proteomics strategy to discover β-glucosidases from *Aspergillus fumigatus* with

two-dimensional page in-gel activity assay and tandem mass spectrometry. J Proteome Res 6(12):4749-57

King N.L., Deutsch E.W., Ranish J.A., Nesvizhskii A.I., Eddes J.S., Mallick P. Eng. J., Desiere F., Flory M., Martin D.B., Kim B., Lee H., Raught B., Aebersold R. (2006) Analysis of the Saccharomyces cerevisiae proteome with PeptideAtlas. Genome Biology 7, R106

Kleber M., Sollins P., Sutton R. (2007) A conceptual model of organo-mineral interactions in soils: self assembly of organic molecular fragments into zonal structures on mineral surfaces. Biogeochem 85, 9-24

Klose J. (1975) Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues - novel approach to testing for induced point mutations in mammals. Humangenetik 26, 231-43

Kononova M.M. (1961) Soil organic matter. Its Nature, Its Role in Soil Formation and in Soil Fertility. Pergamon Press, New York

Koskella J., Stotzky G. (1997) Microbial utilization of free and clay-bound insecticidal Toxins from Bacillus thuringiensis and their retention of insecticidal activity after incubation with microbes. Applied Environ Microbiol 63, 3561-3568

Krause F. (2006) Detection and analysis of protein–protein interactions in organellar and prokaryotic proteomes by native gel electrophoresis: (Membrane) protein complexes and supercomplexes. Electrophoresis 27, 2759–2781

Kristie L., Evans J.C., Crowder J., Miller E.S. (2000) Subtilisins of Bacillus subsp. hydrolyze keratin and allow growth on feathers. Can J Microbiol 46, 1004-1011

Kyte J, Doolittle R. (1982) A simple method for displaying the hydropathic character of a protein J Mol Biol 157, 105-132

Kuzyakov Y. (2006) Sources of CO2 efflux from soil and review of partitioning methods. Soil Biol Biochem 38, 851–860

Lambert J.P., Ethier M., Smith J.C., Figeys D. (2005) Review Proteomics: from gel based to gel free. Anal Chem 77, 3771-3787

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature, 227, 680–685

Leclerc E., Peretz D., Ball H., Sakurai H., Legname G., Serban A., Prusiner S.B., Burton D.R., Williamson R.A. (2001) Immobilized prion protein

undergoes spontaneous rearrangement to a conformation having features in common with the infectious form. EMBO J. 20, 1547-1554

Lee L., Saxena D., Stotzky G. (2003) Activity of free and claybound insecticidal proteins from Bacillus thuringiensis subsp. Israelensis against the mosquito Culex pipiens. Applied Environ Microbiol 69, 4111–4115

Li Y.T., Hsieh Y.L, Henion J.D., Ocain T.D., Schiehser G.A., Ganem B., J (1994) Analysis of the energetics of gas-phase immunophilin ligand complexes by ion-spray mass-spectrometry. Am Chem Soc 116 7487

Loo J.A. (1997) Studying Noncovalent Protein Complexes by Electrospray Ionization Mass Spectrometry. Mass Spectrometry Reviews 16, 1–23

Luo X., Bai Y., Zhao L. (2007) Preservation method for different arsenic species in groundwater sample, 6th Annual Meeting of the International-Professionals-for-the-Advancement-of-Chinese-Earth-Sciences 18, 115+-

Macko V., Stegeman H. (1969) Mapping of potato proteins by combined electrofocusing and electrophoresis identification of varieties. Hoppe-Seylers Zeitschrift Fur Physiologische Chemie 350, 917

Maddison, B.C., Owen, J.P., Bishop, K., Shaw, G., Rees, H.C., Gough, K.C., 2010. The interaction of ruminant PrPSc with soils is influenced by prion source and soil type. Environ Sci Technol 44, 8503–8508

Mann M., Ong S.E., Gronborg M., Steen H., Jensen O.N., Pandey A. (2002) Analysis of protein phosphorylation using mass spectrometry: deciphering the phosphoproteome. Trends Biotechnol 20, 261-268

Marschner B., Kalbitz K. (2003) Controls of bioavailability and biodegradability of dissolved organic matter in soils. Geoderma 113, 211–235

Masciandaro G., Macci C., Doni S., Maserti B.E., Leo A.C-B., Ceccanti B., Wellington E. (2008) Comparison of extraction methods for recovery of extracellular β-glucosidase in two different forest soils Soil Biol Biochem 40, 2156-216

Mazzei P., Piccolo A. (2012) Quantitative Evaluation of Noncovalent Interactions between Glyphosate and Dissolved Humic Substances by NMR Spectroscopy. Environ Sci Technol 46, 5939−5946

Melo E., da Silveira Gomes A., Carvalho I. (2006) α- and β-Glucosidase inhibitors: chemical

structure and biological activity. Tetrahedron 62 (2006) 10277–10302

Miller I., Crawford J., Gianazza E. (2006) Protein stains for proteomic applications: which, when, why? Proteomics 6, 5385-408

Miltner A, Zech W. (1999) Microbial degradation and resynthesis of proteins during incubation of beech leaf litter in the presence of mineral phases. Biol Fertility Soils 30, 48-51

Muga A., Arrondo J.L., Bellon T., Sancho J., Bernabeu C. (1993) Structural and functional studies on the interaction of sodium dodecyl sulfate with β-galactosidase. Arch Biochem Biophys 300, 451-457

Muscolo A., Felici M., Concheri G., Nardi S. (1993) Effect of humic substances on peroxidase and esterase patterns during growth of leaf explants of Nicotiana plumbaginifolia. Biol Fertil Soils 15, 127-31

Muscolo A., Panuccio M.R., Abenavoli M.R., Concheri G., Nardi S. (1996) Effect of molecular complexity and acidity of earthworm faeces humic fractions on glutamate dehydrogenase, glutamine synthetase and phosphoenolpyruvate carboxylase in Daucus carota II cells. Biol Fertil Soils 22, 83-8

Murase A., Yoneda M.,Ueno R., Yonebayashi K. (2003) Extraction of extracellular proteins from a greenhouse soil. Soil Biol Biochem 71, 733-736

Nannipieri P. (1996) In: Humic substances in terrestrial ecosystem. Piccolo, A. (ed.), pp. 293-328 (Elsevier, Amsterdam)

Nannipieri P ., Ascher J., Ceccherini M.T., Landi L., Pietramellara G., Renella G. (2003) Microbial diversity and soil functions . Europ J Soil Sci 54, 655-670

Nannipieri P., Paul E. (2009) The chemical and functional characterization of soil N and its biotic components. Soil Biol Biochem, 41, 2357–2369

Nannipieri P.; Giagnoni L.; Renella G., Puglisi, E., Ceccanti B., Masciandaro G., Fornasier F., Moscatelli M.C., Marinari S. (2012) Soil enzymology: classical and molecular approaches

Nanny M., Ratasuk N. (2002) Characterization and comparison of hydrophobic neutral and hydrophobic acid dissolved organic carbon isolated from three municipal landfill leachates

Nebbioso A., Piccolo A., Spiteller M. (2010) Limitations of electrospray ionization in the analysis of a heterogeneous mixture of naturally occurring hydrophilic and hydrophobic compounds. Rapid Commun Mass Spectrom 24, 3163-3170

Norde W., Lykema G. (1978) The adsorption of human plasma albumin and bovine pancreas ribonuclease at negatively charged polystyrene surfaces. J Colloid Interface Sci 66, 257-265

Oades J.M. (1984) Soil organic matter and structural stability: mechanisms and implications for management. Plant and Soil 76, 319–337

O'Farrell PH. 1975. High-resolution 2-dimensional electrophoresis of proteins. J Biol Chem 250(10):4007-21. Ohlmeier S, Kastaniotis AJ, Hiltunen JK, Bergmann U. 2004. The yeast mitochondrial proteome, a study of fermentative and respiratory growth. J Biol Chem 279 (6): 3956-79

Ogunseitan O.A. (1993) Direct extraction of proteins from environmental samples. Journal of Microbiological Methods 17, 273–281

Ogunseitan O.A. (1997)Direct extraction of catalytic proteins from natural microbial communities. J Microbiol Met 28, 55-63

Olk D.C. (2006) A chemical fractionation for structure–function relations of soil organic matter in nutrient cycling. Soil Sci Soc America J 1013–1022

Ong S.E., Mann M. (2006) A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). Nature Protocols 1, 2650-2660

Parsons R., Vandernoot T. (1988) The oxidation of small organic-molecules - a survey of recent fuel-cell related research. J electroanal chem 257, 1-2 pp 9-45

Pan K. M. et al. (1993) Conversion of alpha-helices into β-sheets features in the formation of the scrapie prion proteins. Proc Natl Acad Sci USA 90, 10962-10966

Paungfoo-Lonhienne C., Visser J., Lonhienne T.G.A. , Schmidt S. (2012) Past, present and future of organic nutrients. Plant and Soil 35, 1-18

Perkins D.N., Pappin DJC, Creasy DM, Cottrel JS (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis 20, 3551–3567

Piccolo A., Mirabella A. (1987) Molecular weight distribution of peat humic substances extracted with different inorganic and organic solutions. Sci Total Environ 62, 39–46

Piccolo A. (2001) The supramolecular stricter of humic substances. Soil Sci 16611-810–832

Piccolo A. (2002) The Supramolecular structure of humic substances. A novel understanding of humus chemistry and implications in soil science. Advan Agron 75, 57-134

Pickford A.R., Campbell I.D. (2004) NMR Studies of Modular Protein Structures and Their

Interactions. Chemical Reviews 104, 3557-3565

Pinck L.A., Allison F.E. (1951) Influence Of Green Manures On The Release Of Native Soil Carbon. Soil Sci 71, 67-76

Pramanik B.N., Bartner P.L., Mirza U.A., Liu Y.H., Ganguly A.K. (1998) Electrospray ionization mass spectrometry for the study of non-covalent complexes: An emerging technology. J Mass Spectr 33, 911-920

Preston G.W., Radford S.E., Ashcroft A.E., Wilson A. (2012) Covalent Cross-Linking within Supramolecular Peptide Structures. Analytical Chem 84, 6790-6797

Prusiner SB. (1992) Chemistry and Biology of Prions. Biochem 31 (49), 12277–12288

Prusiner SB. (2004) An introduction to prion biology and diseases. In: Prusiner, S.B. (Ed.), Prion Biology and Diseases, second ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 1-89

Puglisi E., Arenella M., Renella G., Trevisan M. (2012) Biosensors for Ecotoxicity of Xenobiotics: a Focus on Soil and Risk Assessment. In: Biosensors and Environmental Health, Victor R. Preedy, King's College Hospital, London, UK; Vinood Patel, University of Westminister, CRC Press. ISBN 9781578087358.

Quiquampoix H., Ratcliffe R.G. (1992) A ^{31}P NMR study of the adsorption of bovine serum albumin on montmorillonite using phosphate and the paramagnetic cation Mn2+: modification of conformation with pH. J Colloid Interface Sci 148 , 343-352

Quiquampoix H. (2000) Mechanism of protein adsorption on surfaces and consequences for enzyme activity in soil. In: Soil Biochemistry, Bollag J.M. and Stozky G. (Eds.). Vol. 10. Marcel Dekker, New York p 171-206

Quiquampoix H., Burns, R.G. (2007) Interactions between proteins and soil mineral surfaces: environmental and health consequences. Elements 3, 401–406

Quiquampoix H. (2008) Interaction with Soil Constituents Determines the Environmental Impact of Proteins. Revista de la ciencia del suelo y nutricion vegetal 8

Rao M.A., Russo F.,Granatab V.,Berisio R., Zagari A., Gianfreda L. (2006). Fate of prions in soil: Interaction of a recombinant ovine prion protein with synthetic humic-like mineral complexes. Soil Biol Biochem 39, 493-504

Renella G., Landi L., Nannipieri P. (2002) Hydrolase activity during and after the chloroform360 fumigation of soils as affected by protease activity. Soil Biol Biochem 34, 51-60

Rigou P., Rezaei H., Staunton S., Quiquampoix H. (2006) Fate of prions in soil: Adsorption and extraction by electroelution of recombinant ovine prion protein from montmorillonite and natural soils. Envir Sci Technol 40, 1497-1503

Roberts P., Jones D.L. (2008) Critical evaluation of methods for determining total protein in soil solution. Soil Biol Biochem 40, 1485-1495

Rondon M.R., August P.R., Bettermann A.D., Brady S.F., Grossman T.H., Liles M.R., Loiacono K.A., Lynch B.A., MacNeil I.A., Minor C., Tiong C.L., Gilman M., Osburne M.S., Clardy J., Handelsman J.,Goodman R.M. (2000) Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. Appl Environ Microbiol 66, 2541-2547

Rudick MJ, Fitzgerald ZE, Rudick VL. (1979) Intra- and extracellular forms of alpha-glucosidase from Aspergillus niger. Arch Biochem Biophys 1, 509-20

Rupert J.P., Granquist W.T., Pinnavaia T.J. (1987) Catalytic properties of clay minerals: in Chemistry of Clays and Clay Minerals, A. C. D. Newman, ed. Mineralogical Society, Longman Scientific & Technical, London 275- 318

Sander M., Tomaszewski J.E, Madliger M., Pedersen J.A., Schwarzenbach R.P. (2012) Adsorption of Insecticidal Cry1Ab Protein to Humic Substances. 1. Experimental Approach and Mechanistic Aspects. Env Sci Technol 46, 9923-9931

Saunders S.E., Bartelt-Hunt S.L., Bartz J.C. (2008) Prions in the environment: occurrence, fate, and mitigation. Prion 2, 162–169

Saunders, S.E., Yuan, Q., Bartz, J.C., Bartelt-Hunt, S.L. (2011) Effects of solution chemistry and aging time on prion protein adsorption and replication of soilbound prions. PLoS ONE 6, e18752

Saunders S.E., Bartelt-Hunt S.L., Bartz J.C. (2012) Occurrence, Transmission, and Zoonotic potential of Chronic Wasting Disease. Emer Infect Dis 18, http://ndx.doi.org/10.3201/eid1802.110685

Schindel F.V., Mercer E.J., Rice J.A. (2007) Chemical characteristics of glomalin-related soil protein (GRSP) extracted from soils of varying organic matter content. Soil Biol Biochem 39, 320–329

Schmidt M., Torn M., Abiven S., Dittmar T., Guggenberger G., Janssens I. , Kleber M., Kögel-Knabner I., Lehmann J., Manning D., Nannipieri P., Rasse D., Weiner S., Trumbore S. (2011) Persistence of soil organic matter as an ecosystem property. Nature 478, 49–56

Schneider T., Riedel K. (2010) Environmental proteomics: Analysis of structure and function of microbial communities. Proteomics 10, 785-798

Schnitzer M. and Khan S.U. (1972) Humic substances in the environment. Marcel Dekker, New York, 327

Schnitzer M. (1978) Humic substances: chemistry and reactions. In: Khan, S.U. (Ed.), Soil Organic Matter. Elsevier Scientific Publishing 47 Company, New York, pp. 1–64

Schnitzer M., Schuppli P. (1989) Method for the sequential extraction of organic matter from soils and soil fractions. Soil Sci Soc American J 53, 1418–1424

Schulten H.R., Plage B., Schnitzer M. (1991) A chemical-structure for humic substances. Naturwissenschaften 78, 311-312

Schulten H.R., Schnitzer M. (1992) Structural studies on soil humic acids by curie-point pyrolysis-gas chromatography mass-spectrometry. Soil Science 153, 205-224

Schulze W.X., Gleixner G., Kaiser K., Guggenberger G., Mann M., Schulze E.D. (2005) A proteomic fingerprint of dissolved organic carbon and of soil particles. Oecologia 142, 335–343

Shental-Bechor D, Levy Y. (2009) Folding of glycoproteins: toward understanding the biophysics of the glycosylation code. Curr Op Struct Biol1 9, 524–33

Singleton I., Merrington G., Colvan S., Delahunty J.S. (2003) The potential of soil protein-based methods to indicate metal contamination. Applied Soil Ecology, 23, 25–32

Sendra V., Cannella D., Bersch B., FieschiF., Menage S., Lascoux D., Cove`s J. (2006) CopH from CupriaVidus metallidurans CH34. A Novel Periplasmic Copper-Binding Protein. Biochem 45, 5557-5566

Servagent-Noinville S., Revault M., Quiquampoix H., and Baron M.H. (2000) Conformational Changes of Bovine Serum Albumin Induced by Adsorption on Different Clay Surfaces: FTIR Analysis. J Colloid Interface Sci 221, 273–283 Symposium Series No. 602, Am Chem Soc, Washington, DC, p 321

Simonart P., Batistic L., Mayaudon J. (1967) Isolation of protein from humic acid extracted from soil. Plant and Soil 27, 153–161

Simpson A. (2002) Determining the molecular weight, aggregation, structures and interactions of natural organic matter using diffusion ordered spectroscopy Magn Res Chem 40, S72–S82

Smejkalova D., Piccolo A. (2002) Aggregation and disaggregation of humic supramolecular assemblies by NMR diffusion ordered Spectroscopy (DOSY-NMR) . Environ Sci Technol 42, 699-706

Smith C.B., Booth C.J., Pedersen J.A. (2011) Fate of prions in soil: a review. J Environ Qual 40, 449–461

Solaiman Z., Marschner P., Wang D.M., Rengel Z. (2007) Growth, P uptake and rhizosphere properties of wheat and canola genotypes in an alkaline soil with low P availability. Biol fert soil 44(1):143-153

Somerville R. A. (2002) TSE agent strains and PrP: reconciling structure and function. Trends in Biochemical Service 27 n.12, 606-612

Southall N.T., Dill K.A., Haymet A.D.J. (2002) A View of the Hydrophobic Effect. J. Phys. Chem. B, 106, 521–533

Stevenson F.J. (1964) In: V. Sauchelli (Ed), Fertilizer nitrose – its chemistry and technology. Am Chem Soc Monogr 161, 18-39

Stevenson F.J. (1994) Humus Chemistry: Genesis, Composition, Reactions. John Wiley & Sons, New York

Stotzky G (1986) Influence of soil mineral colloids on metabolic processes, growth, adhesion, and ecology ofmicrobes and viruses. In: Interactions of Soil Minerals with natural Organics and Microbes (ed) PM Huang andMS chnitzer. pp 305–428. SSSA Spec. Publ. No. 17. Madison, WI

Sutton R., Sposito G. (2005) Molecular structure of humic substances: the new view. Environ Sci Technol 39, 9009-9015

Swaby R.J., Ladd J.N. (1964) Chemical nature, microbial resistance and origin of soil humus. Chemical Abstracts 60, 4727b

Tan K. H. (2003) Humic Matter in Soil and the Environment: Principles and Controversies; Marcel Dekker: New York

Tan W.F., Koopal L.K., Weng L.P., van Riemsdijk W.H., Norde W. (2008) Humic acid protein complexation. Geochim Cosmochim Acta 72, 2090–2099

Tan W.F., Koopal L., Norde W. (2009) Interaction between humic acid and lysozyme, studied by dynamic light scattering and isothermal titration calorimetry. Environ Sci Technol 43, 591-596

Tate R. L.(1992) Humic and fulvic acids: formation and decomposition, in: Soil Organic Matter Biological and Ecological Effects, Krieger Publishing Company, Malebar, Florida, pp. 147–164

Tollinger M., Skrynnikov N.R., Mulder F.A.A., Forman-Kay J.D., Kay L.E. (2001) Slow

Dynamics in Folded and Unfolded States of an SH3 Domain. J American Chem

Soc 123, 11341-11352

Tomaszewski J.E., Schwarzenbach R.P., Sander M (2011). Protein Encapsulation by Humic Substances. Environ Sci Technol 45, 6003–6010

Tomaszewski J.E, Madliger M., Pedersen J.A., Schwarzenbach R.P., Sander M. (2012) Adsorption of Insecticidal Cry1Ab Protein to Humic Substances. 2. Influence of Humic and Fulvic Acid Charge and Polarity Characteristics. Environ Sci Technol 46, 9932–9940

Tombelli S., Mascini M., Braccini L., Anichini M., Turner AP. Coupling of a DNA piezoelectric biosensor and polymerase chain reaction to detect apolipoprotein E polymorphisms (2000) Biosen Bioelectron 15:363–370

Torsvik V. and Øvreås L. (2002) Microbial diversity and function in soil: from genes to ecosystems. Curr Opinion Microbiol 5, 240–245

Tsiroulnikov K., Rezai H, Bonch-Osmolovskaya E., Nedkov P., Gousterova A., Cueff V., Godfroy A., Barbier G., Metro F., Chobert J.M., Clayette P., Dormont D., Grosclaude J., Haertle T. (2004) Hydrolysis of the amyloid prion protein and nonpathogenic meat and bone meal by anaerobic thermophilic

prokaryotes and Streptomyces subspecies. J Agricultural Food Chem 52, 6353-6360

Unlu M., Morgan M.E., Minden J.S. (1997) Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. Electrophoresis 18, 2071–2077

Varki A. (1993) Biological roles of oligosaccharides - all of the theories are correct. Glycobiology 3, 97–130

Waksman S.A. (1938) Humus, 2nd ed. Bailliere, Tindall &Cox, London, UK

Wang F., Wang X., Yuan C.G., Ma J. (2010) Generating a prion with bacterially expressed recombinant prion protein. Science 327, 1132–1135

Wang HB., Zhang ZX., Li H., He HB., Fang CX., Zhang AJ., Li QS., Chen RS., Guo XK., Lin HF., Wu LK., Lin S., Chen T., Lin RY., Peng XX., Lin WX. (2011) Characterization of metaproteomics in crop rhizospheric soil. J Prot Res 10, 932–940

Wershaw R. L. (1999) Molecular aggregation of humic substances. Soil Sci 164, 803–813

Williams M.A., Taylor E.B., Mula H.P. (2010) Metaproteomic characterization of a soil microbial community following carbon amendment. Soil Biol Biochem 42, 1148-1156

Wright S.F., Upadhyay A. (1996) Extraction of an abundant and unusual protein from soil and comparison with hyphal protein of arbuscular mycorrhizal fungi. Soil Sci 161, 575-585

Wright S.F., Upadhyay A. (1998) A survey of soils for aggregate stability and glomalin, a glycoprotein produced by hyphae of arbuscular mycorrhizal fungi. Plant and Soil 198, 97-107

Wu L., Wang H., Zhang Z., Lin R., Zhang Z., Lin W. 2011. Comparative metaproteomic analysis on consecutively *Rehmannia glutinosa*-monocultured rhizosphere soil. PlosOne, 6, e20611

Xiao J.J., Byrd J., Marcucci G., Grever M., (2003) Identification of thiols and glutathione conjugates of depsipeptide FK228 (FR901228), a novel histone protein deacetylase inhibitor, in the blood. Rapid Com In Mass Spectrom Volume: 17 Issue: 8 Pages: 757-766

Yu C.H., Norman M.A., Newton S.Q., Miller D.M., Teppen B.J., Schafer L. (2000) Molecular dynamics simulations of the adsorption of proteins on clay mineral surfaces. J Mol Structure 556, 95-103

Yuan W., Zydney L. (2000) Humic acid fouling during ultrafiltration Environ Sci Technol 34, 5043-5050

Zhang O., Forman-Kay J.D. (1995) Structural Characterization of Folded and Unfolded States of

an SH3 Domain in Equilibrium in Aqueous Buffer. Biochem 34, 6784-6794

Zang X., van Heemst J.D.H., Dria K.J., Hatche P.G. (2000) Encapsulation of protein in humic acid from a histosol as an explanation for the occurrence of organic nitrogen in soil and sediment. Organic Geochem 31, 679-695

Zhang J., Du J., Zhang F., Yu Y., Zhang J. (2011) Chemical characterization of humic substances isolated from mangrove swamp sediments: The Qinglan area of Hainan Island, China. Estuar Coast Shelf S 93, 220-227

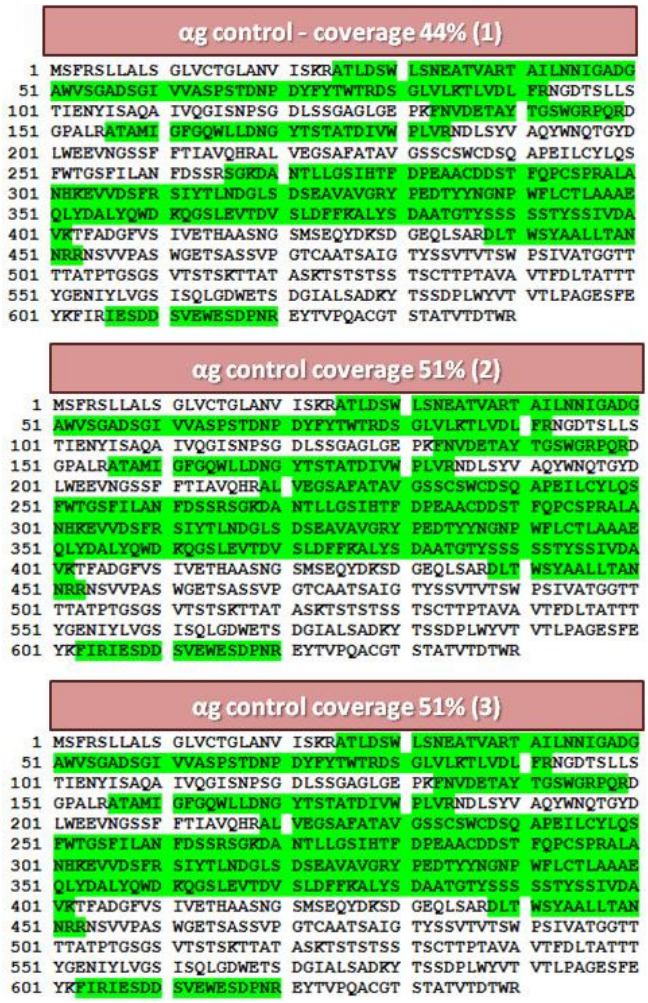
Ziechmann W. (1994) Humic Substances. Mannheim: Wissenschaftsverlag

ANNEX I

1. Model proteins sequence coverage %

- ***α-glucosidase from A. niger (αg)***

N-PAGE, control



- ***α-glucosidase from A. niger (αg) + HS***

N-PAGE, 1:1

αg :HS 1:1 - coverage 51% (1)

```
  1 MSFRSLLALS GLVCTGLANV ISKRATLDSW LSNEATVART AILNNIGADG
 51 AWVSGADSGI VVASPSTDNP DYFYTWTRDS GLVLKTLVDL FRNGDTSLLS
101 TIENYISAQA IVQGISNPSG DLSSGAGLGE PKFNVDETAY TGSWGRPQRD
151 GPALRATAMI GFGQWLLDNG YTSTATDIVW PLVRNDLSYV AQYWNQTGYD
201 LWEEVNGSSF FTIAVQHRAL VEGSAFATAV GSSCSWCDSQ APEILCYLQS
251 FWTGSFILAN FDSSRSGKDA NTLLGSIHTF DPEAACDDST FQPCSPRALA
301 NHKEVVDSFR SIYTLNDGLS DSEAVAVGRY PEDTYYNGNP WFLCTLAAAE
351 QLYDALYQWD KQGSLEVTDV SLDFFKALYS DAATGTYSSS SSTYSSIVDA
401 VKTFADGFVS IVETHAASNG SMSEQYDKSD GEQLSARDLT WSYAALLTAN
451 NRRNSVVPAS WGETSASSVP GTCAATSAIG TYSSVTVTSW PSIVATGGTT
501 TTATPTGSGS VTSTSKTTAT ASKTSTSTSS TSCTTPTAVA VTFDLTATTT
551 YGENIYLVGS ISQLGDWETS DGIALSADKY TSSDPLWYVT VTLPAGESFE
601 YKFIRIESDD SVEWESDPNR EYTVPQACGT STATVTDTWR
```

αg :HS 1:1 - coverage 51% (2)

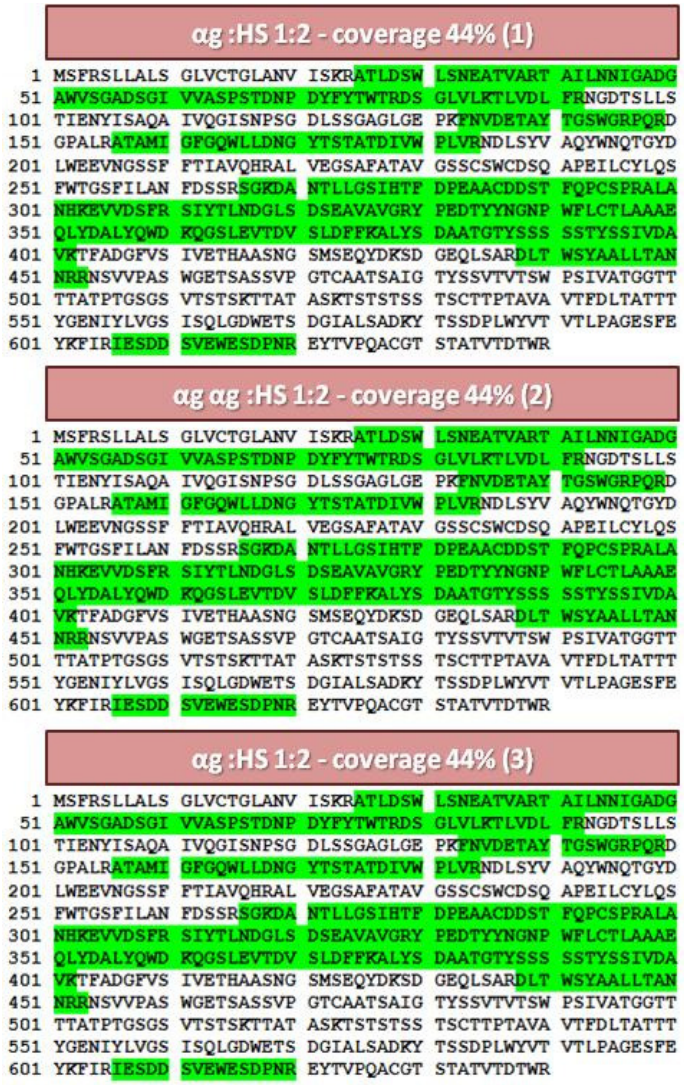
```
  1 MSFRSLLALS GLVCTGLANV ISKRATLDSW LSNEATVART AILNNIGADG
 51 AWVSGADSGI VVASPSTDNP DYFYTWTRDS GLVLKTLVDL FRNGDTSLLS
101 TIENYISAQA IVQGISNPSG DLSSGAGLGE PKFNVDETAY TGSWGRPQRD
151 GPALRATAMI GFGQWLLDNG YTSTATDIVW PLVRNDLSYV AQYWNQTGYD
201 LWEEVNGSSF FTIAVQHRAL VEGSAFATAV GSSCSWCDSQ APEILCYLQS
251 FWTGSFILAN FDSSRSGKDA NTLLGSIHTF DPEAACDDST FQPCSPRALA
301 NHKEVVDSFR SIYTLNDGLS DSEAVAVGRY PEDTYYNGNP WFLCTLAAAE
351 QLYDALYQWD KQGSLEVTDV SLDFFKALYS DAATGTYSSS SSTYSSIVDA
401 VKTFADGFVS IVETHAASNG SMSEQYDKSD GEQLSARDLT WSYAALLTAN
451 NRRNSVVPAS WGETSASSVP GTCAATSAIG TYSSVTVTSW PSIVATGGTT
501 TTATPTGSGS VTSTSKTTAT ASKTSTSTSS TSCTTPTAVA VTFDLTATTT
551 YGENIYLVGS ISQLGDWETS DGIALSADKY TSSDPLWYVT VTLPAGESFE
601 YKFIRIESDD SVEWESDPNR EYTVPQACGT STATVTDTWR
```

αg :HS 1:1 - coverage 58% (3)

```
  1 MSFRSLLALS GLVCTGLANV ISKRATLDSW LSNEATVART AILNNIGADG
 51 AWVSGADSGI VVASPSTDNP DYFYTWTRDS GLVLKTLVDL FRNGDTSLLS
101 TIENYISAQA IVQGISNPSG DLSSGAGLGE PKFNVDETAY TGSWGRPQRD
151 GPALRATAMI GFGQWLLDNG YTSTATDIVW PLVRNDLSYV AQYWNQTGYD
201 LWEEVNGSSF FTIAVQHRAL VEGSAFATAV GSSCSWCDSQ APEILCYLQS
251 FWTGSFILAN FDSSRSGKDA NTLLGSIHTF DPEAACDDST FQPCSPRALA
301 NHKEVVDSFR SIYTLNDGLS DSEAVAVGRY PEDTYYNGNP WFLCTLAAAE
351 QLYDALYQWD KQGSLEVTDV SLDFFKALYS DAATGTYSSS SSTYSSIVDA
401 VKTFADGFVS IVETHAASNG SMSEQYDKSD GEQLSARDLT WSYAALLTAN
451 NRRNSVVPAS WGETSASSVP GTCAATSAIG TYSSVTVTSW PSIVATGGTT
501 TTATPTGSGS VTSTSKTTAT ASKTSTSTSS TSCTTPTAVA VTFDLTATTT
551 YGENIYLVGS ISQLGDWETS DGIALSADKY TSSDPLWYVT VTLPAGESFE
601 YKFIRIESDD SVEWESDPNR EYTVPQACGT STATVTDTWR
```

- ***α-glucosidase from A. niger (αg) + HS***

N-PAGE, 1:2



- ***α-glucosidase from A. niger (αg) + HS***

N-PAGE, 1:2

αg :HS 1:4 - coverage 44% (1)

```
  1 MSFRSLLALS GLVCTGLANV ISKRATLDSW LSNEATVART AILNNIGADG
 51 AWVSGADSGI VVASPSTDNP DYFYTWTRDS GLVLKTLVDL FRNGDTSLLS
101 TIENYISAQA IVQGISNPSG DLSSGAGLGE PKFNVDETAY TGSWGRPQRD
151 GPALRATAMI GFGQWLLDNG YTSTATDIVW PLVRNDLSYV AQYWNQTGYD
201 LWEEVNGSSF FTIAVQHRAL VEGSAFATAV GSSCSWCDSQ APEILCYLQS
251 FWTGSFILAN FDSSRSGKDA NTLLGSIHTF DPEAACDDST FQPCSPRALA
301 NHKEVVDSFR SIYTLNDGLS DSEAVAVGRY PEDTYYNGNP WFLCTLAAAE
351 QLYDALYQWD KQGSLEVTDV SLDFFKALYS DAATGTYSSS SSTYSSIVDA
401 VKTFADGFVS IVETHAASNG SMSEQYDKSD GEQLSARDLT WSYAALLTAN
451 NRRNSVVPAS WGETSASSVP GTCAATSAIG TYSSVTVTSW PSIVATGGTT
501 TTATPTGSGS VTSTSKTTAT ASKTSTSTSS TSCTTPTAVA VTFDLTATTT
551 YGENIYLVGS ISQLGDWETS DGIALSADKY TSSDPLWYVT VTLPAGESFE
601 YKFIRIESDD SVEWESDPNR EYTVPQACGT STATVTDTWR
```

αg :HS 1:4 - coverage 39% (2)

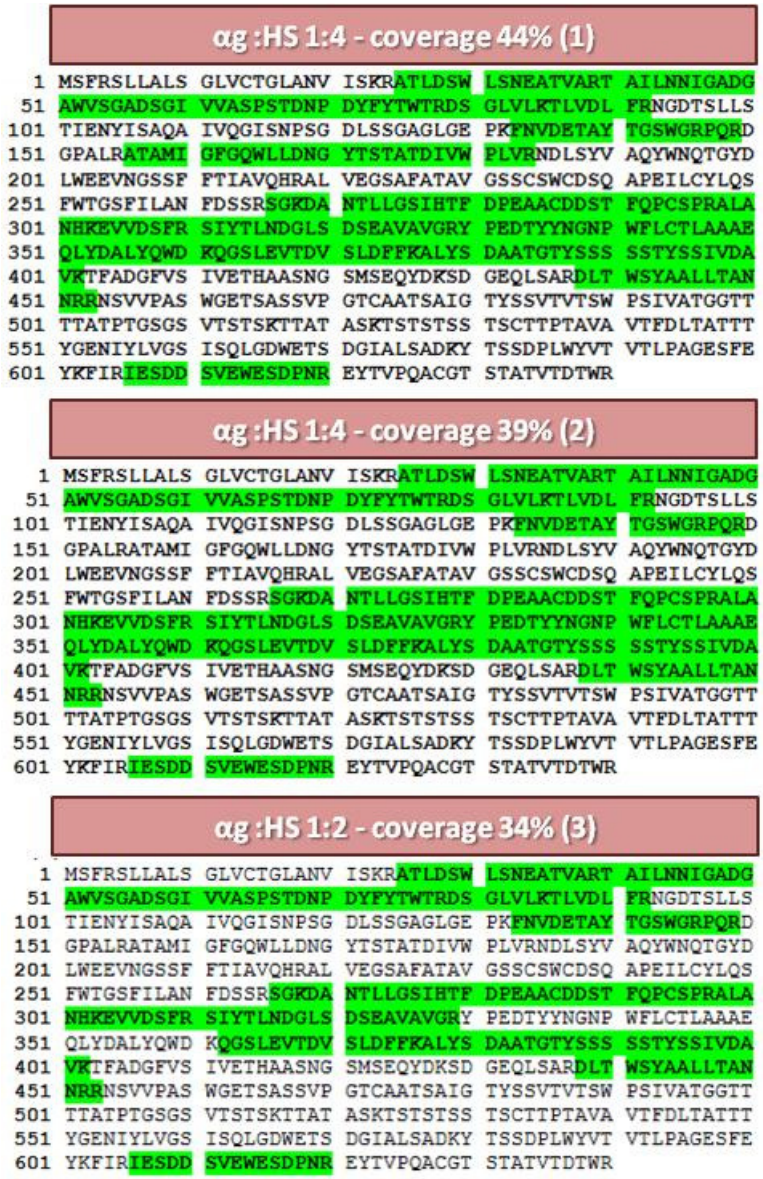
```
  1 MSFRSLLALS GLVCTGLANV ISKRATLDSW LSNEATVART AILNNIGADG
 51 AWVSGADSGI VVASPSTDNP DYFYTWTRDS GLVLKTLVDL FRNGDTSLLS
101 TIENYISAQA IVQGISNPSG DLSSGAGLGE PKFNVDETAY TGSWGRPQRD
151 GPALRATAMI GFGQWLLDNG YTSTATDIVW PLVRNDLSYV AQYWNQTGYD
201 LWEEVNGSSF FTIAVQHRAL VEGSAFATAV GSSCSWCDSQ APEILCYLQS
251 FWTGSFILAN FDSSRSGKDA NTLLGSIHTF DPEAACDDST FQPCSPRALA
301 NHKEVVDSFR SIYTLNDGLS DSEAVAVGRY PEDTYYNGNP WFLCTLAAAE
351 QLYDALYQWD KQGSLEVTDV SLDFFKALYS DAATGTYSSS SSTYSSIVDA
401 VKTFADGFVS IVETHAASNG SMSEQYDKSD GEQLSARDLT WSYAALLTAN
451 NRRNSVVPAS WGETSASSVP GTCAATSAIG TYSSVTVTSW PSIVATGGTT
501 TTATPTGSGS VTSTSKTTAT ASKTSTSTSS TSCTTPTAVA VTFDLTATTT
551 YGENIYLVGS ISQLGDWETS DGIALSADKY TSSDPLWYVT VTLPAGESFE
601 YKFIRIESDD SVEWESDPNR EYTVPQACGT STATVTDTWR
```

αg :HS 1:2 - coverage 34% (3)

```
  1 MSFRSLLALS GLVCTGLANV ISKRATLDSW LSNEATVART AILNNIGADG
 51 AWVSGADSGI VVASPSTDNP DYFYTWTRDS GLVLKTLVDL FRNGDTSLLS
101 TIENYISAQA IVQGISNPSG DLSSGAGLGE PKFNVDETAY TGSWGRPQRD
151 GPALRATAMI GFGQWLLDNG YTSTATDIVW PLVRNDLSYV AQYWNQTGYD
201 LWEEVNGSSF FTIAVQHRAL VEGSAFATAV GSSCSWCDSQ APEILCYLQS
251 FWTGSFILAN FDSSRSGKDA NTLLGSIHTF DPEAACDDST FQPCSPRALA
301 NHKEVVDSFR SIYTLNDGLS DSEAVAVGRY PEDTYYNGNP WFLCTLAAAE
351 QLYDALYQWD KQGSLEVTDV SLDFFKALYS DAATGTYSSS SSTYSSIVDA
401 VKTFADGFVS IVETHAASNG SMSEQYDKSD GEQLSARDLT WSYAALLTAN
451 NRRNSVVPAS WGETSASSVP GTCAATSAIG TYSSVTVTSW PSIVATGGTT
501 TTATPTGSGS VTSTSKTTAT ASKTSTSTSS TSCTTPTAVA VTFDLTATTT
551 YGENIYLVGS ISQLGDWETS DGIALSADKY TSSDPLWYVT VTLPAGESFE
601 YKFIRIESDD SVEWESDPNR EYTVPQACGT STATVTDTWR
```

- ***α-glucosidase from A. niger (αg) + HS***

N-PAGE, 1:4



- ***α-glucosidase from A. niger(ag)***

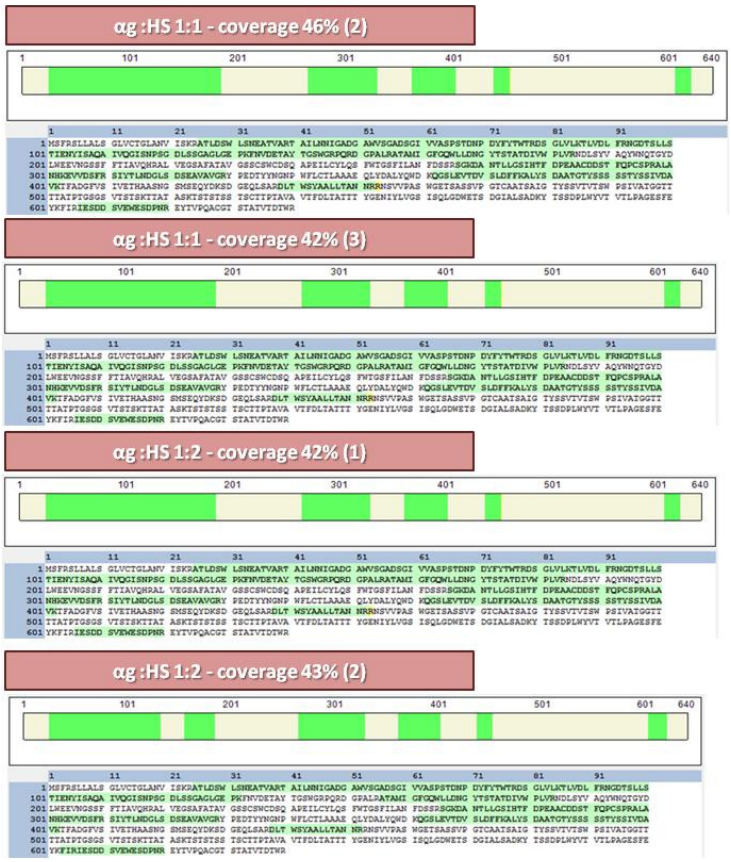
SDS-PAGE

- ***α-glucosidase from A. niger(αg)***

SDS-PAGE

- ***α-glucosidase from A. niger (αg)***

SDS-PAGE



- ***α-glucosidase from A. niger (αg)***

SDS-PAGE

- ***β-glucosidase from A. niger (βg)***

N-PAGE, control

βg control - coverage 16% (1)

```
  1 MRFTSIEAVA LTAVSLASAD ELAYSPPYYP SPWANGQGDW AEAYQRAVDI
 51 VSQMTLAEKV NLTTGTGWEL ELCVGQTGGV PRLGIPGMCA QDSPLGVRDS
101 DYNSAFPAGV NVAATWDKNL AYLRGQAMGQ EFSDKGADIQ LGPAAGPLGR
151 SPDGGRNWEG FSPDPALSGV LFAETIKGIQ DAGVVATAKH YIAYEQEHFR
201 QAPEAQGYGF NITESGSANL DDKTMHELYL WPFADAIRAG AGAVMCSYNQ
251 INNSYGCQNS YTLNKLLKAE LGFQGFVMSD WAAHHAGVSG ALAGLDMSMP
301 GDVDYDSGTS YWGTNLTISV LNGTVPQWRV DDMAVRIMAA YYKVGRDRLW
351 TPPNFSSWTR DEYGFKYYYV SEGPYEKVNQ FVNVQRNHSE LIRRIGADST
401 VLLKNDGALP LTGKERLVAL IGEDAGSNPY GANGCSDRGC DNGTLAMGWG
451 SGTANFPYLV TPEQAISNEV LKNKNGVFTA TDNWAIDQIE ALAKTASVSL
501 VFVNADSGEG YINVDGNLGD RRNLTLWRNG DNVIKAAASN CNNTIVIIHS
551 VGPVLVNEWY DNPNVTAILW GGLPGQESGN SLADVLYGRV NPGAKSPFTW
601 GKTREAYQDY LYTEPNNGNG APQEDFVEGV FIDYRGFDKR NETPIYEFGY
651 GLSYTTFNYS NLQVEVLSAP AYEPASGETE AAPTFGEVGN ASDYLYPDGL
701 QRITKFIYPW LNSTDLEASS GDASYGQDAS DYLPEGATDG SAQPILPAGG
751 GAGGNPRLYD ELIRVSVTIK NTGKVAGDEV PQLYVSLGGP NEPKIVLRQF
801 ERITLQPSKE TQWSTTLTRR DLANWNVETQ DWEITSYPKM VFAGSSSRKL
851 PLRASLPTVH
```

βg control - coverage 17% (2)

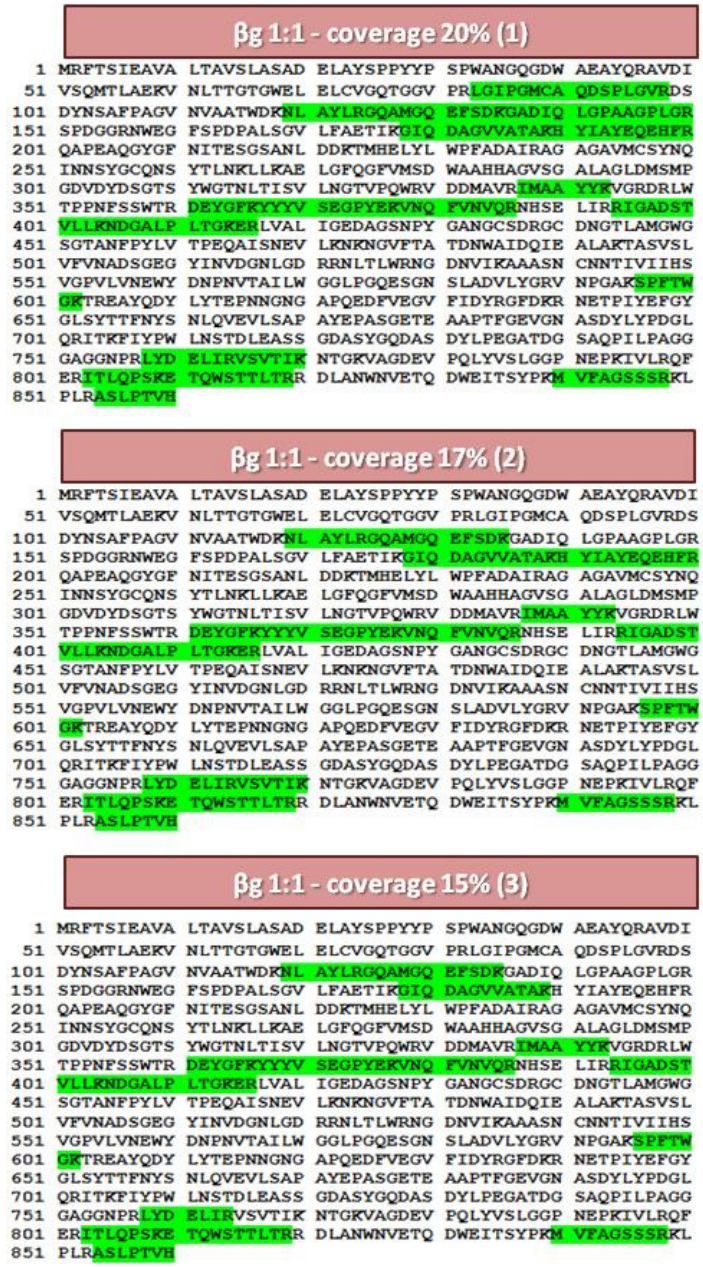
```
  1 MRFTSIEAVA LTAVSLASAD ELAYSPPYYP SPWANGQGDW AEAYQRAVDI
 51 VSQMTLAEKV NLTTGTGWEL ELCVGQTGGV PRLGIPGMCA QDSPLGVRDS
101 DYNSAFPAGV NVAATWDKNL AYLRGQAMGQ EFSDKGADIQ LGPAAGPLGR
151 SPDGGRNWEG FSPDPALSGV LFAETIKGIQ DAGVVATAKH YIAYEQEHFR
201 QAPEAQGYGF NITESGSANL DDKTMHELYL WPFADAIRAG AGAVMCSYNQ
251 INNSYGCQNS YTLNKLLKAE LGFQGFVMSD WAAHHAGVSG ALAGLDMSMP
301 GDVDYDSGTS YWGTNLTISV LNGTVPQWRV DDMAVRIMAA YYKVGRDRLW
351 TPPNFSSWTR DEYGFKYYYV SEGPYEKVNQ FVNVQRNHSE LIRRIGADST
401 VLLKNDGALP LTGKERLVAL IGEDAGSNPY GANGCSDRGC DNGTLAMGWG
451 SGTANFPYLV TPEQAISNEV LKNKNGVFTA TDNWAIDQIE ALAKTASVSL
501 VFVNADSGEG YINVDGNLGD RRNLTLWRNG DNVIKAAASN CNNTIVIIHS
551 VGPVLVNEWY DNPNVTAILW GGLPGQESGN SLADVLYGRV NPGAKSPFTW
601 GKTREAYQDY LYTEPNNGNG APQEDFVEGV FIDYRGFDKR NETPIYEFGY
651 GLSYTTFNYS NLQVEVLSAP AYEPASGETE AAPTFGEVGN ASDYLYPDGL
701 QRITKFIYPW LNSTDLEASS GDASYGQDAS DYLPEGATDG SAQPILPAGG
751 GAGGNPRLYD ELIRVSVTIK NTGKVAGDEV PQLYVSLGGP NEPKIVLRQF
801 ERITLQPSKE TQWSTTLTRR DLANWNVETQ DWEITSYPKM VFAGSSSRKL
851 PLRASLPTVH
```

βg control - coverage 16% (1)

```
  1 MRFTSIEAVA LTAVSLASAD ELAYSPPYYP SPWANGQGDW AEAYQRAVDI
 51 VSQMTLAEKV NLTTGTGWEL ELCVGQTGGV PRLGIPGMCA QDSPLGVRDS
101 DYNSAFPAGV NVAATWDKNL AYLRGQAMGQ EFSDKGADIQ LGPAAGPLGR
151 SPDGGRNWEG FSPDPALSGV LFAETIKGIQ DAGVVATAKH YIAYEQEHFR
201 QAPEAQGYGF NITESGSANL DDKTMHELYL WPFADAIRAG AGAVMCSYNQ
251 INNSYGCQNS YTLNKLLKAE LGFQGFVMSD WAAHHAGVSG ALAGLDMSMP
301 GDVDYDSGTS YWGTNLTISV LNGTVPQWRV DDMAVRIMAA YYKVGRDRLW
351 TPPNFSSWTR DEYGFKYYYV SEGPYEKVNQ FVNVQRNHSE LIRRIGADST
401 VLLKNDGALP LTGKERLVAL IGEDAGSNPY GANGCSDRGC DNGTLAMGWG
451 SGTANFPYLV TPEQAISNEV LKNKNGVFTA TDNWAIDQIE ALAKTASVSL
501 VFVNADSGEG YINVDGNLGD RRNLTLWRNG DNVIKAAASN CNNTIVIIHS
551 VGPVLVNEWY DNPNVTAILW GGLPGQESGN SLADVLYGRV NPGAKSPFTW
601 GKTREAYQDY LYTEPNNGNG APQEDFVEGV FIDYRGFDKR NETPIYEFGY
651 GLSYTTFNYS NLQVEVLSAP AYEPASGETE AAPTFGEVGN ASDYLYPDGL
701 QRITKFIYPW LNSTDLEASS GDASYGQDAS DYLPEGATDG SAQPILPAGG
751 GAGGNPRLYD ELIRVSVTIK NTGKVAGDEV PQLYVSLGGP NEPKIVLRQF
801 ERITLQPSKE TQWSTTLTRR DLANWNVETQ DWEITSYPKM VFAGSSSRKL
851 PLRASLPTVH
```

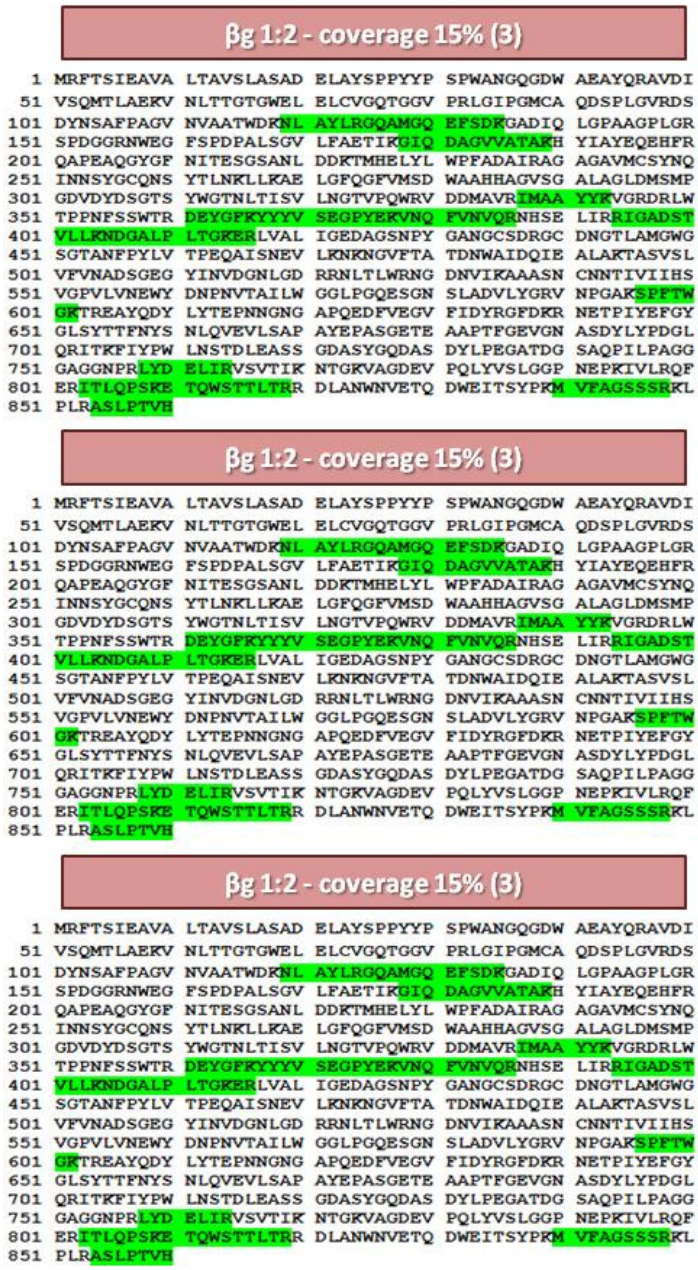
- ***β-glucosidase from A. niger (βg) + HS***

N-PAGE, 1:1



- ***β-glucosidase from A. niger (βg) + HS***

N-PAGE, 1:2



- ***β-glucosidase from A. niger (βg) + HS***

N-PAGE, 1:4

βg 1:4 - coverage 16% (1)

```
  1 MRFTSIEAVA LTAVSLASAD ELAYSPPYYP SPWANGQGDW AEAYQRAVDI
 51 VSQMTLAEKV NLTTGTGWEL ELCVGQTGGV PRLGIPGMCA QDSPLGVRDS
101 DYNSAFPAGV NVAATWDKNL AYLRGQAMGQ EFSDKGADIQ LGPAAGPLGR
151 SPDGGRNWEG FSPDPALSGV LFAETIKGIQ DAGVVATAKH YIAYEQEHFR
201 QAPEAQGYGF NITESGSANL DDKTMHELYL WPFADAIRAG AGAVMCSYNQ
251 INNSYGCQNS YTLNKLLKAE LGFQGFVMSD WAAHHAGVSG ALAGLDMSMP
301 GDVDYDSGTS YWGTNLTISV LNGTVPQWRV DDMAVRIMAA YYKVGRDRLW
351 TPPNFSSWTR DEYGFKYYYV SEGPYEKVNQ FVNVQRNHSE LIRRIGADST
401 VLLKNDGALP LTGKERLVAL IGEDAGSNPY GANGCSDRGC DNGTLAMGWG
451 SGTANFPYLV TPEQAISNEV LKNKNGVFTA TDNWAIDQIE ALAKTASVSL
501 VFVNADSGEG YINVDGNLGD RRNLTLWRNG DNVIKAAASN CNNTIVIIHS
551 VGPVLVNEWY DNPNVTAILW GGLPGQESGN SLADVLYGRV NPGAKSPFTW
601 GKTREAYQDY LYTEPNNGNG APQEDFVEGV FIDYRGFDKR NETPIYEFGY
651 GLSYTTFNYS NLQVEVLSAP AYEPASGETE AAPTFGEVGN ASDYLYPDGL
701 QRITKFIYPW LNSTDLEASS GDASYGQDAS DYLPEGATDG SAQPILPAGG
751 GAGGNPRLYD ELIRVSVTIK NTGKVAGDEV PQLYVSLGGP NEPKIVLRQF
801 ERITLQPSKE TQWSTTLTRR DLANWNVETQ DWEITSYPKM VFAGSSSRKL
851 PLRASLPTVH
```

βg 1:4 - coverage 16% (2)

```
  1 MRFTSIEAVA LTAVSLASAD ELAYSPPYYP SPWANGQGDW AEAYQRAVDI
 51 VSQMTLAEKV NLTTGTGWEL ELCVGQTGGV PRLGIPGMCA QDSPLGVRDS
101 DYNSAFPAGV NVAATWDKNL AYLRGQAMGQ EFSDKGADIQ LGPAAGPLGR
151 SPDGGRNWEG FSPDPALSGV LFAETIKGIQ DAGVVATAKH YIAYEQEHFR
201 QAPEAQGYGF NITESGSANL DDKTMHELYL WPFADAIRAG AGAVMCSYNQ
251 INNSYGCQNS YTLNKLLKAE LGFQGFVMSD WAAHHAGVSG ALAGLDMSMP
301 GDVDYDSGTS YWGTNLTISV LNGTVPQWRV DDMAVRIMAA YYKVGRDRLW
351 TPPNFSSWTR DEYGFKYYYV SEGPYEKVNQ FVNVQRNHSE LIRRIGADST
401 VLLKNDGALP LTGKERLVAL IGEDAGSNPY GANGCSDRGC DNGTLAMGWG
451 SGTANFPYLV TPEQAISNEV LKNKNGVFTA TDNWAIDQIE ALAKTASVSL
501 VFVNADSGEG YINVDGNLGD RRNLTLWRNG DNVIKAAASN CNNTIVIIHS
551 VGPVLVNEWY DNPNVTAILW GGLPGQESGN SLADVLYGRV NPGAKSPFTW
601 GKTREAYQDY LYTEPNNGNG APQEDFVEGV FIDYRGFDKR NETPIYEFGY
651 GLSYTTFNYS NLQVEVLSAP AYEPASGETE AAPTFGEVGN ASDYLYPDGL
701 QRITKFIYPW LNSTDLEASS GDASYGQDAS DYLPEGATDG SAQPILPAGG
751 GAGGNPRLYD ELIRVSVTIK NTGKVAGDEV PQLYVSLGGP NEPKIVLRQF
801 ERITLQPSKE TQWSTTLTRR DLANWNVETQ DWEITSYPKM VFAGSSSRKL
851 PLRASLPTVH
```

βg 1:4 - coverage 16% (3)

```
  1 MRFTSIEAVA LTAVSLASAD ELAYSPPYYP SPWANGQGDW AEAYQRAVDI
 51 VSQMTLAEKV NLTTGTGWEL ELCVGQTGGV PRLGIPGMCA QDSPLGVRDS
101 DYNSAFPAGV NVAATWDKNL AYLRGQAMGQ EFSDKGADIQ LGPAAGPLGR
151 SPDGGRNWEG FSPDPALSGV LFAETIKGIQ DAGVVATAKH YIAYEQEHFR
201 QAPEAQGYGF NITESGSANL DDKTMHELYL WPFADAIRAG AGAVMCSYNQ
251 INNSYGCQNS YTLNKLLKAE LGFQGFVMSD WAAHHAGVSG ALAGLDMSMP
301 GDVDYDSGTS YWGTNLTISV LNGTVPQWRV DDMAVRIMAA YYKVGRDRLW
351 TPPNFSSWTR DEYGFKYYYV SEGPYEKVNQ FVNVQRNHSE LIRRIGADST
401 VLLKNDGALP LTGKERLVAL IGEDAGSNPY GANGCSDRGC DNGTLAMGWG
451 SGTANFPYLV TPEQAISNEV LKNKNGVFTA TDNWAIDQIE ALAKTASVSL
501 VFVNADSGEG YINVDGNLGD RRNLTLWRNG DNVIKAAASN CNNTIVIIHS
551 VGPVLVNEWY DNPNVTAILW GGLPGQESGN SLADVLYGRV NPGAKSPFTW
601 GKTREAYQDY LYTEPNNGNG APQEDFVEGV FIDYRGFDKR NETPIYEFGY
651 GLSYTTFNYS NLQVEVLSAP AYEPASGETE AAPTFGEVGN ASDYLYPDGL
701 QRITKFIYPW LNSTDLEASS GDASYGQDAS DYLPEGATDG SAQPILPAGG
751 GAGGNPRLYD ELIRVSVTIK NTGKVAGDEV PQLYVSLGGP NEPKIVLRQF
801 ERITLQPSKE TQWSTTLTRR DLANWNVETQ DWEITSYPKM VFAGSSSRKL
851 PLRASLPTVH
```

- ***β-glucosidase from A. niger (βg)***

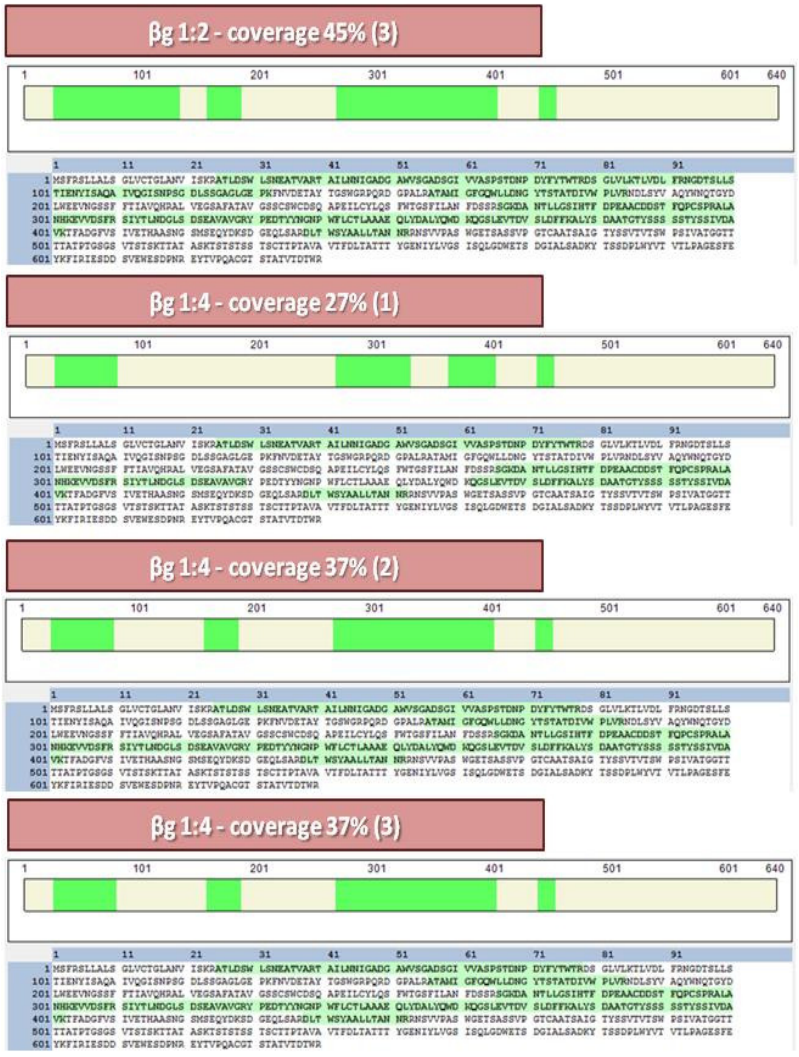
SDS-PAGE

- ***β-glucosidase from A. niger (βg)***

SDS-PAGE

- ***β-glucosidase from A. niger (βg)***

SDS-PAGE



- ***Rec PrP***

N-PAGE, control; 1:1 (+HS)

RecPrP control - Budorcas taxicolor cov 30%

```
  1 MVKSHIGSWI LVLFVAMWSD VGLCKKRPKP GGGWNTGGSR YPGQGSPGGN
 51 RYPPQGGGGW GQPHGGGWGQ PHGGGWGQPH GGGWGQPHGG GGWGQGGSHS
101 QWNKPSKPKT NMKHVAGAAA AGAVVGGLGG YMLGSAMSRP LIHFGSDYED
151 RYYRENMYRY PNQVYYRPVD QYSNQNNFVH DCVNITVKQH TVTTTTKGEN
201 FTETDIKIME RVVEQMCITQ YQRESQAYYQ RGASVILFSS PPVILLISFL
251 IFLIVG
```

RecPrP control - Budorcas taxicolor cov 30%

```
  1 MVKSHIGSWI LVLFVAMWSD VGLCKKRPKP GGGWNTGGSR YPGQGSPGGN
 51 RYPPQGGGGW GQPHGGGWGQ PHGGGWGQPH GGGWGQPHGG GGWGQGGSHS
101 QWNKPSKPKT NMKHVAGAAA AGAVVGGLGG YMLGSAMSRP LIHFGSDYED
151 RYYRENMYRY PNQVYYRPVD QYSNQNNFVH DCVNITVKQH TVTTTTKGEN
201 FTETDIKIME RVVEQMCITQ YQRESQAYYQ RGASVILFSS PPVILLISFL
251 IFLIVG
```

RecPrP control - Budorcas taxicolor cov 30%

```
  1 MVKSHIGSWI LVLFVAMWSD VGLCKKRPKP GGGWNTGGSR YPGQGSPGGN
 51 RYPPQGGGGW GQPHGGGWGQ PHGGGWGQPH GGGWGQPHGG GGWGQGGSHS
101 QWNKPSKPKT NMKHVAGAAA AGAVVGGLGG YMLGSAMSRP LIHFGSDYED
151 RYYRENMYRY PNQVYYRPVD QYSNQNNFVH DCVNITVKQH TVTTTTKGEN
201 FTETDIKIME RVVEQMCITQ YQRESQAYYQ RGASVILFSS PPVILLISFL
251 IFLIVG
```

RecPrP :HS 1:1- Canis familiaris cov 28%

```
  1 MVKSHIGSWI LVLFVAMWSD VGLCKKRPKP GGGWNTGGSR YPGQGSPGGN
 51 RYPPQGGGGW GQPHGGGWGQ PHGGGWGQPH GGGWGQPHGG GGWGQGGTHS
101 QWNKPSKPKT NMKHVAGAAA AGAVVGGLGG YLLGSAMSRP LIHFGNDCED
151 RYYRENMYRY PNQVYYRSVD QYNNQSTFVH DCVNITVKQH TVTTTKGENF
201 TETDIKMMER VVEQMCITQY QRESEAYYQR GASVILFSSP PVILLVSFLI
251 FLIVG
```

RecPrP :HS 1:1- Antilope cercicapra cov 23%

```
  1 MVKSHIGSWI LVLFVAMWSD VGLCKKRPKP GGGWNTGGSR YPGQGSPGGN
 51 RYPPQGGGGW GQPHGGGWGQ PHGGGWGQPH GGGWGQPHGG GWGQPHGGGG
101 WGQGGTHSQW NKPSKPKTNM KHVAGAAAAG AVVGGLGGYM LGSAMSRPLI
151 HFGNDYEDRY YRENMYRYPN QVYYRPVDQY SNQNNFVHDC VNITVKQHTV
201 TTTTKGENFT ETDIKMMERV VEQMCITQYQ RESQAYYQRG ASVILFSSPP
251 VILLISFLIF LIVG
```

RecPrP :HS 1:1- Antilope cercicapra cov 27%

```
  1 MVKSHIGSWI LVLFVAMWSD VGLCKKRPKP GGGWNTGGSR YPGQGSPGGN
 51 RYPPQGGGGW GQPHGGGWGQ PHGGGWGQPH GGGWGQPHGG GWGQPHGGGG
101 WGQGGTHSQW NKPSKPKTNM KHVAGAAAAG AVVGGLGGYM LGSAMSRPLI
151 HFGNDYEDRY YRENMYRYPN QVYYRPVDQY SNQNNFVHDC VNITVKQHTV
201 TTTTKGENFT ETDIKMMERV VEQMCITQYQ RESQAYYQRG ASVILFSSPP
251 VILLISFLIF LIVG
```

- ***Rec PrP + HS***

N-PAGE, 1:2; 1:4

RecPrP :HS 1:2- Canis familiaris cov 28%

```
  1 MVKSHIGSWI LVLFVAMWSD VGLCKKRPKP GGGWNTGGSR YPGQGSPGGN
 51 RYPPQGGGGW GQPHGGGWGQ PHGGGWGQPH GGGWGQPHGG GGWGQGGTHS
101 QWNKPSKPKT NMKHVAGAAA AGAVVGGLGG YLLGSAMSRP LIHFGNDCED
151 RYYRENMYRY PNQVYYRSVD QYNNQSTFVH DCVNITVKQH TVTTTKGENF
201 TETDIKMMER VVEQMCITQY QRESEAYYQR GASVILFSSP PVILLVSFLI
251 FLIVG
```

RecPrP :HS 1:2- Canis familiaris cov 22%

```
  1 MVKSHIGSWI LVLFVAMWSD VGLCKKRPKP GGGWNTGGSR YPGQGSPGGN
 51 RYPPQGGGGW GQPHGGGWGQ PHGGGWGQPH GGGWGQPHGG GGWGQGGTHS
101 QWNKPSKPKT NMKHVAGAAA AGAVVGGLGG YLLGSAMSRP LIHFGNDCED
151 RYYRENMYRY PNQVYYRSVD QYNNQSTFVH DCVNITVKQH TVTTTKGENF
201 TETDIKMMER VVEQMCITQY QRESEAYYQR GASVILFSSP PVILLVSFLI
251 FLIVG
```

RecPrP :HS 1:2- Canis familiaris cov 28%

```
  1 MVKSHIGSWI LVLFVAMWSD VGLCKKRPKP GGGWNTGGSR YPGQGSPGGN
 51 RYPPQGGGGW GQPHGGGWGQ PHGGGWGQPH GGGWGQPHGG GGWGQGGTHS
101 QWNKPSKPKT NMKHVAGAAA AGAVVGGLGG YLLGSAMSRP LIHFGNDCED
151 RYYRENMYRY PNQVYYRSVD QYNNQSTFVH DCVNITVKQH TVTTTKGENF
201 TETDIKMMER VVEQMCITQY QRESEAYYQR GASVILFSSP PVILLVSFLI
251 FLIVG
```

RecPrP :HS 1:4 - Budorcas taxicolor cov 30%

```
  1 MVKSHIGSWI LVLFVAMWSD VGLCKKRPKP GGGWNTGGSR YPGQGSPGGN
 51 RYPPQGGGGW GQPHGGGWGQ PHGGGWGQPH GGGWGQPHGG GGWGQGGSHS
101 QWNKPSKPKT NMKHVAGAAA AGAVVGGLGG YMLGSAMSRP LIHFGSDYED
151 RYYRENMYRY PNQVYYRPVD QYSNQNNFVH DCVNITVKQH TVTTTKGEN
201 FTETDIKIME RVVEQMCITQ YQRESQAYYQ RGASVILFSS PPVILLISFL
251 IFLIVG
```

RecPrP :HS 1:4- Canis familiaris cov 22%

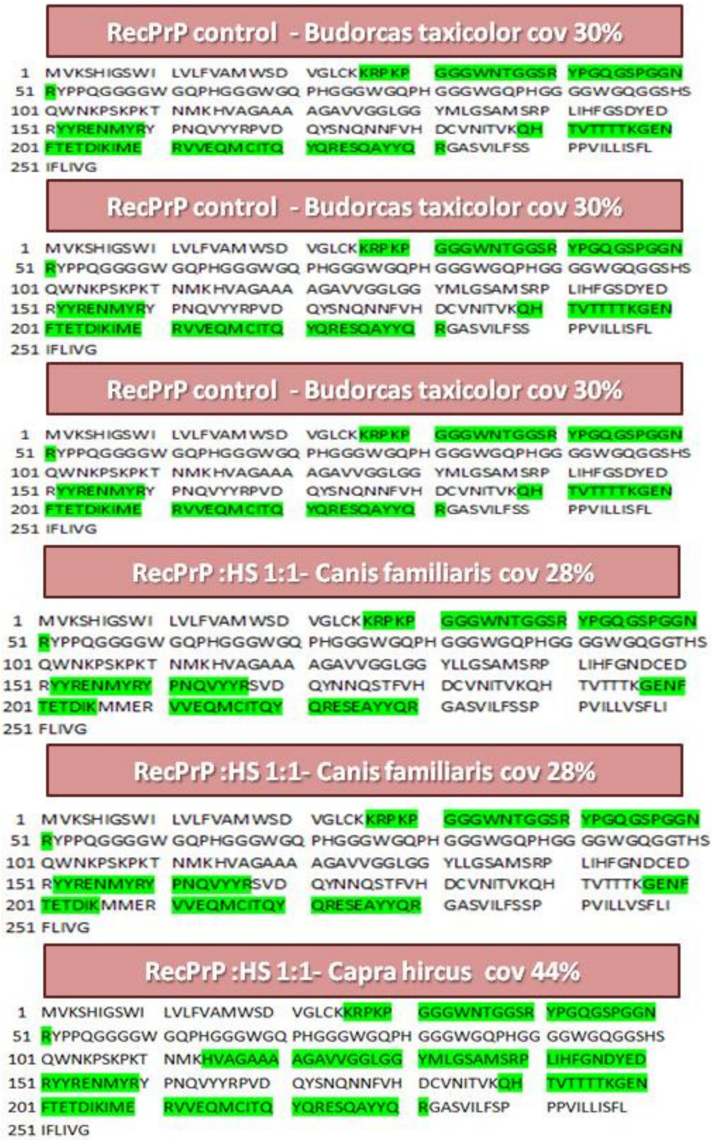
```
  1 MVKSHIGSWI LVLFVAMWSD VGLCKKRPKP GGGWNTGGSR YPGQGSPGGN
 51 RYPPQGGGGW GQPHGGGWGQ PHGGGWGQPH GGGWGQPHGG GGWGQGGTHS
101 QWNKPSKPKT NMKHVAGAAA AGAVVGGLGG YLLGSAMSRP LIHFGNDCED
151 RYYRENMYRY PNQVYYRSVD QYNNQSTFVH DCVNITVKQH TVTTTKGENF
201 TETDIKMMER VVEQMCITQY QRESEAYYQR GASVILFSSP PVILLVSFLI
251 FLIVG
```

RecPrP :HS 1:4- Canis familiaris cov 19%

```
  1 MVKSHIGSWI LVLFVAMWSD VGLCKKRPKP GGGWNTGGSR YPGQGSPGGN
 51 RYPPQGGGGW GQPHGGGWGQ PHGGGWGQPH GGGWGQPHGG GGWGQGGTHS
101 QWNKPSKPKT NMKHVAGAAA AGAVVGGLGG YLLGSAMSRP LIHFGNDCED
151 RYYRENMYRY PNQVYYRSVD QYNNQSTFVH DCVNITVKQH TVTTTKGENF
201 TETDIKMMER VVEQMCITQY QRESEAYYQR GASVILFSSP PVILLVSFLI
251 FLIVG
```

- ***Rec PrP***

SDS-PAGE, control; 1:1 (+HS)



2. Kyte-Doolittle Plots

- Figure 1: ***Myoglobin (Mb),*** *N-PAGE*

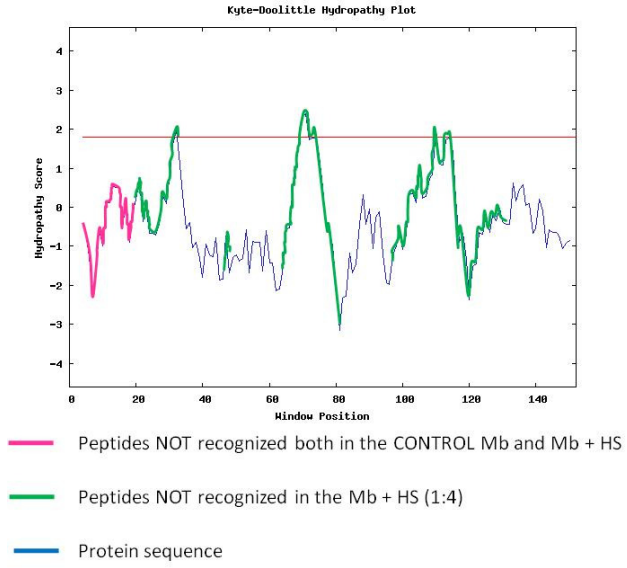


Figure 2: ***α- glucosidase (αg)*** , *N-PAGE*

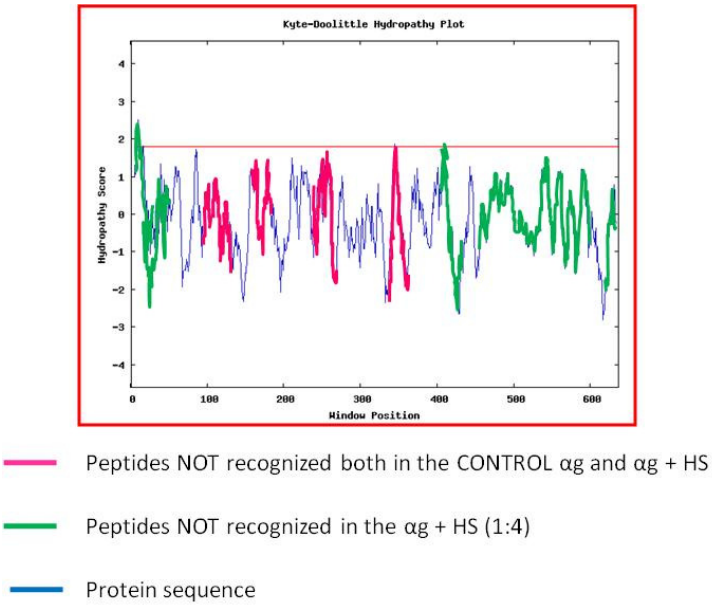
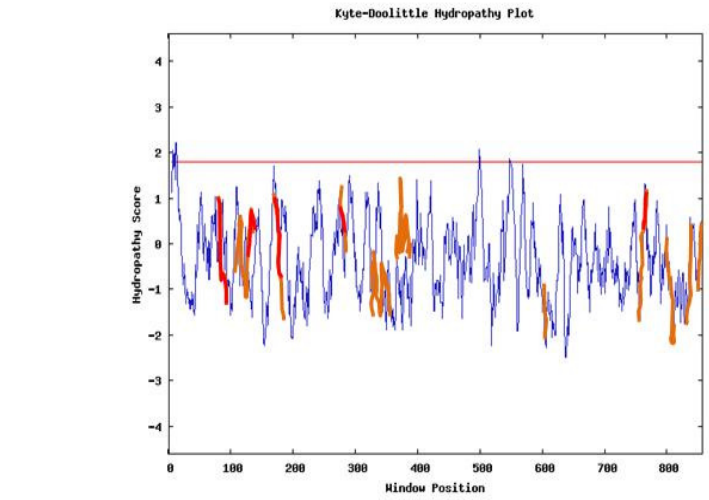


Figure 3: ***β-glucosidase (βg)*** , *N-PAGE*



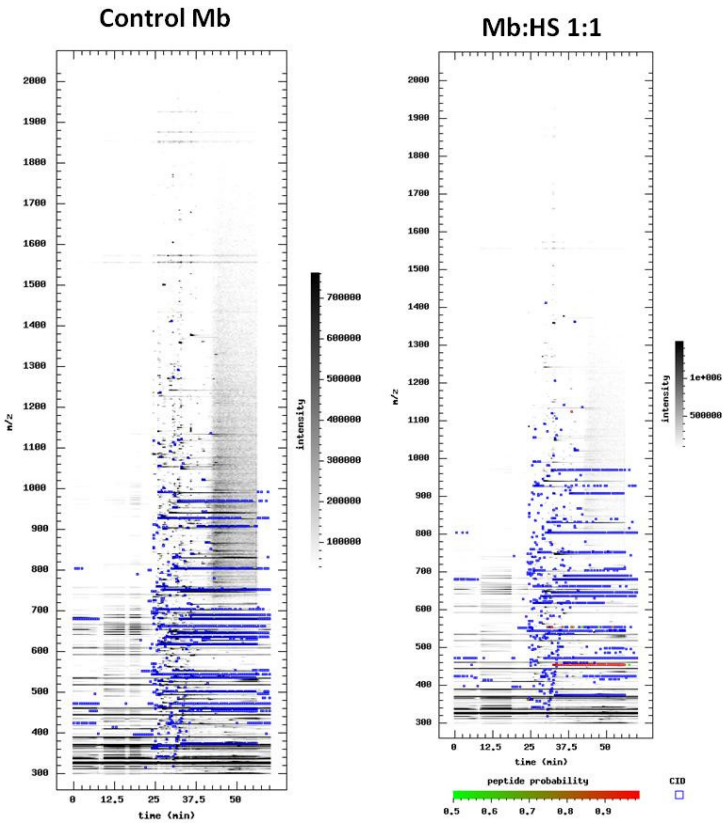
Peptides RECOGNIZED in the CONTROL βg

Peptides RECOGNIZED both in the αg + HS (1:4) and CONTROL βg

Protein sequence

3. PEP 3D IMAGES

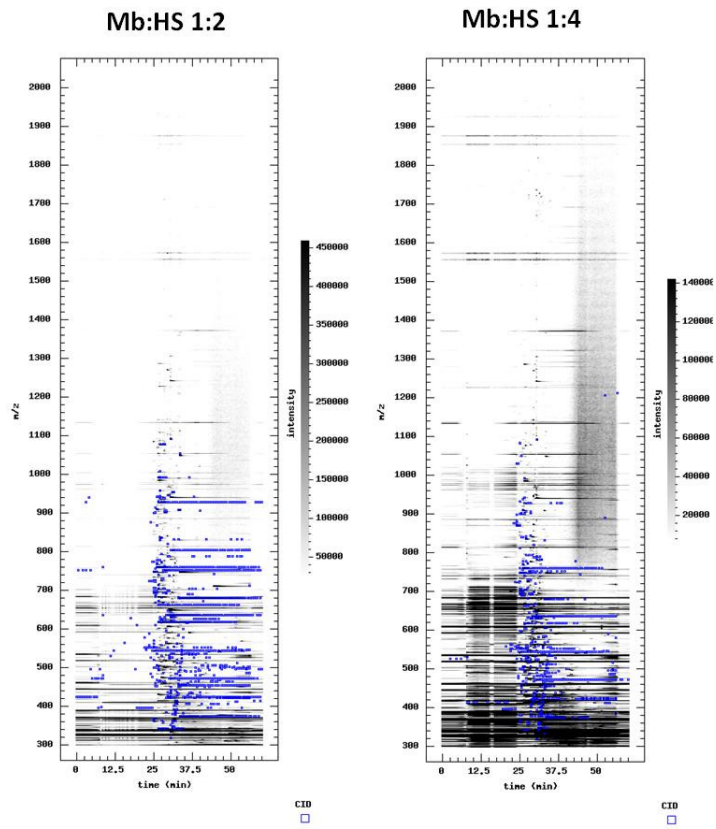
- ***Myoglobin***



Black pixels: precurson ions

Blue pixels: precursor ions selected for the identification by the instrument (CID)

Red/Green pixels: CID identified

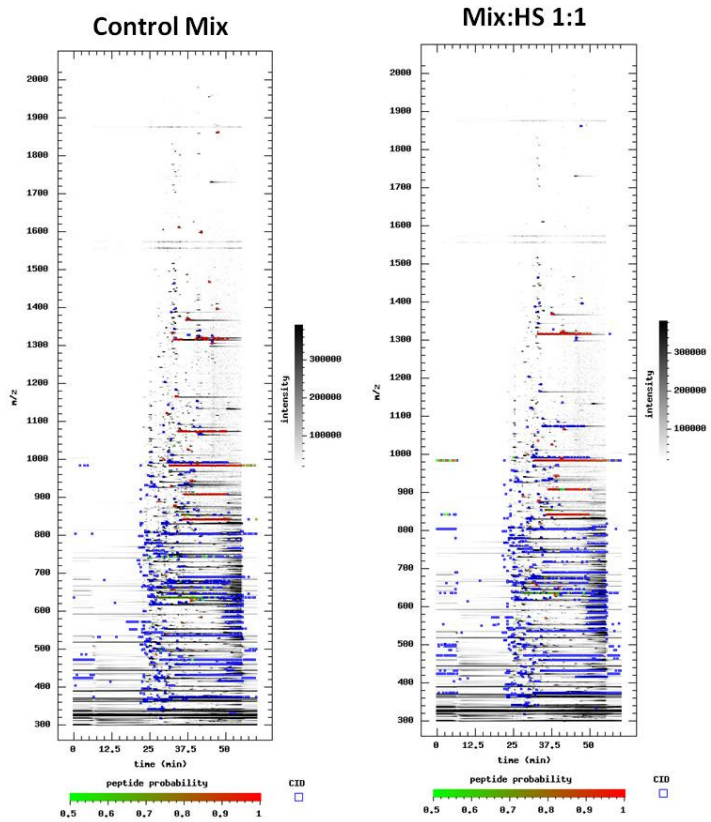


Black pixels: precurson ions

Blue pixels: precursor ions selected for the identification by the instrument (CID)

Red/Green pixels: CID identified

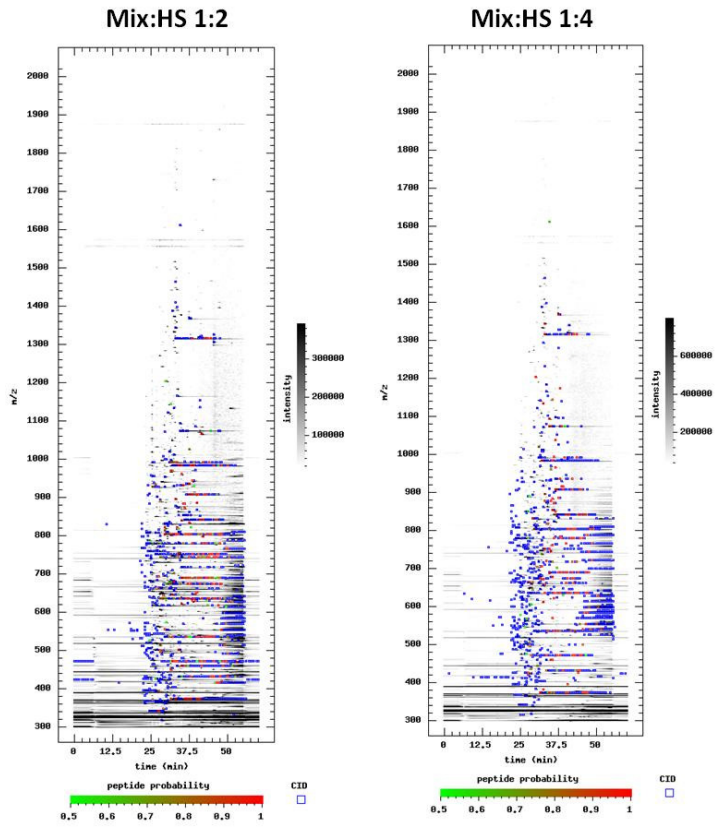
- ***Protein Mix***



Black pixels: precurson ions

Blue pixels: precursor ions selected for the identification by the instrument (CID)

Red/Green pixels: CID identified



Black pixels: precurson ions

Blue pixels: precursor ions selected for the identification by the instrument (CID)

Red/Green pixels: CID identified

ANNEX II

Publications

1. E. Puglisi, M. **Arenella**, G. Renella, M. Trevisan. *Biosensors for ecotoxicity of xenobiotics*. In Biosensors and Environmental Health**,** Victor R. Preedy, King's College Hospital, London, UK; Vinood Patel, University of Westminister, London, UK . Agosto 2012 by CRC Press. ISBN 9781578087358

2. S. Doni, C. Macci, E. Peruzzi, M. **Arenella**, B. Ceccanti, G. Masciandaro**.** *In Situ Phytoremediation of a Historically Contaminated Soil by Heavy Metals and Polychlorobiphenyls*. *Journal of Environmental Monitoring*, 2012, 14, 1383-1390

3

Biosensors for Ecotoxicity of Xenobiotics: A Focus on Soil and Risk Assessment

Edoardo Puglisi,[1] Mariarita Arenella,[2,a] Giancarlo Renella[2,b] and Marco Trevisan[3,*]

ABSTRACT

In this chapter we discuss the potential application of biosensors for the ecotoxicologal risk assessment of xenobiotics in soil environments. In the first part we introduce basic concepts of ecotoxicological risk assessment, ecotoxicology, environmental fate of xenobiotics, with particular focus on bioavailability processes and their importance in exposure and ecotoxicity of xenobiotics in soils. Specifically, all the main processes controlling the environmental fate of xenobiotics in soils are explained.

In the second part we introduce some principles of the functioning of different classes of whole cell biosensors and chemical sensors for

[1]Istituto di Microbiologia, Università Cattolica del Sacro Cuore, Via Emilia Parmense 84, 29122 Piacenza, Italy; E-mail: edoardo.puglisi@unicatt.it
[2]Dipartimento di Scienze delle Produzioni Vegetali, del Suolo e dell'Ambiente Agroforestale, Università di Firenze, Piazza delle Cascine 18, 50144 Firenze, Italy.
[a]E-mail: mariarita.arenella@unifi.it
[b]E-mail: giancarlo.renella@unifi.it
[3]Istituto di Chimica Agraria ed Ambientale, Università Cattolica del Sacro Cuore, Via Emilia Parmense 84, 29122 Piacenza, Italy; E-mail: marco.trevisan@unicatt.it
*Corresponding author

List of abbreviations after the text.

xenobiotics of ecotoxicological interests and present some applications and their contribution to improve the ecoxicological risk assessment of xenobiotics in soils. We show how biosensors can be applied together with chemical assessment of bioavailability of xenobiotics in soil to assess ecotoxicological risk. The application of biosensors to improve the classical risk assessment paradigm is presented, in order to show how this technology can be used to assess the ecotoxicological risk of xenobiotics in soils. Examples of biosensors with potential application for ecotoxicological risk assessment are also presented and discussed. The relevance of these studies of environmental health is also acknowledged.

We conclude that biosensors technology holds relevant potential for ecotoxicological risk assessment of xenobiotics in soils.

INTRODUCTION—PRINCIPLES OF ECOTOXICOLOGICAL RISK ASSESSMENT

Risk assessment for xenobiotics is currently the main regulatory decision-support process behind the measures to prevent, assess and manage environmental pollution. Risk assessment can be defined as a scientific procedure aimed at identifying the hazards correlated with a pollutant and the risks related with their use/presence in the environment. It is possible to distinguish between a human health and an ecotoxicological risk assessment: in both cases the concern is related to toxic substances, but the focus is moved from human to ecological receptors. The US National Academy of Science already defined about 30 years ago a risk assessment "paradigm", that is applicable for both human or ecological receptors, and it is made up of four components (NRC 1983): i) hazard identification, ii) exposure assessment, iii) dose-response assessment and iv) risk characterization (Fig. 3.1). In the first step, all available information about the toxicity (or ecotoxicity) of the chemical of concern are gathered; exposure is aimed at identifying the magnitude of the releases as well as the possible pathways and potential exposures for human and ecological receptors; dose-response assessments are divided in the evaluation of the observable data range and the extrapolation of ranges to toxicological (or ecotoxicological) endpoints; risk characterization finally integrates the information of the previous components in order to assess the potential or existing risk of an adverse effect (Newman and Unger 2003).

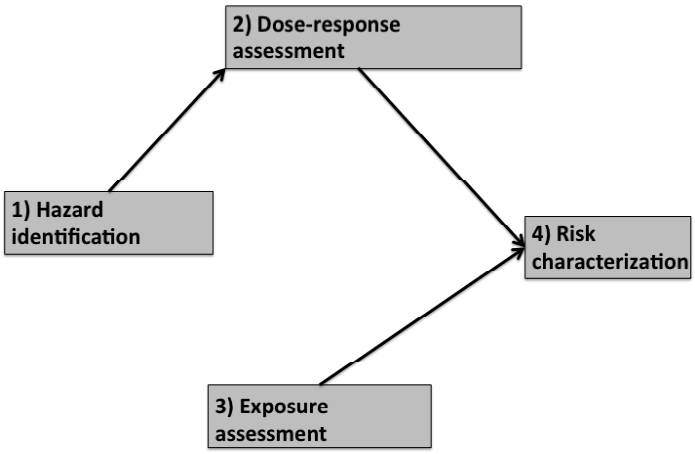


Figure 3.1. The risk assessment paradigm scheme. Modified from NRC (1983). This figure shows that different steps must be undertaken in order to characterize and assess risk of a xenobiotic. First, the hazard (i.e. the toxicity or ecotoxicity) of the xenobiotic must be identified. This hazard is then quantified by a dose-response assessment: increasing doses of the xenobiotic are applied and the response measurement. Finally, exposure is assessed by quantifying the fraction of xenobiotic that is reaching the receptor.

ECOTOXICITY AND ENVIRONMENTAL FATE OF XENOBIOTICS

The term ecotoxicology was firstly used in 1969 by Truhaut (see Truhaut 1977), and aimed at extending toxicology, i.e. the science of studying the effects of poisons on an organism, to an ecosystem level. Ecotoxicology thus takes some approaches used in toxicology (e.g. dose-response assessments, effective doses and other end-points) and extends them to an ecosystem level. As outlined by Moriarty (1983), it would be a mistake to think of ecotoxicology as a simple translation of toxicology, where the only difference is the test species (e.g. a water flea rather than a lab rat): ecotoxicology is concerned with ecosystems health, and thus takes into account effects at population or community level, and considers not only death or carcinogenicity, but also effects that could have a deep impact on ecosystems, such as developmental or endocrine effects. Several definitions of ecotoxicology were proposed in time: a recent and synthetic one defines it as "the study of harmful effects of chemicals on ecosystems" (Walker et al. 2001).

Given the definition above, ecotoxicology is quite a multidisciplinary science, aimed at assessing the distribution and the effects of contaminants in the ecosystems. Knowledge about the environmental distribution of a contaminant is thus at the basis of any ecotoxicological risk assessment, including those based on the use of biosensors. Any ecosystem, either aquatic or terrestrial, is studied by ecotoxicology: here we will focus on xenobiotics in soil ecosystems, especially on agricultural soils.

First of all, it is necessary to make some definition and distinction among terms that are often used as synonyms by non-experts. In this chapter we focus on xenobiotics: the term derives from the Greek, and literary means "stranger for life". Xenobiotics are thus compounds that are not produced in nature and not normally considered constitutive components of a specific biological system (Rand and Petrocelli 1985). Xenobiotics are not to be confounded with contaminants and pollutants: a contaminant is a compound with natural or synthetic origin that is present in a given environment at concentrations higher than the natural ones, while pollutant indicates a compound that is present at concentrations able to cause deleterious effects on living organisms. This implies that a xenobiotic is always a contaminant, but it becomes a pollutant of ecotoxicological concern only when present at concentrations exerting toxic effects.

In order to show how biosensors can be applied to assess the ecotoxicity of xenobiotics, some definitions of environmental fate are also necessary. Once released in the environment, a xenobiotic undergoes a series of processes of transport and, eventually, transformation. These processes depend on the physico-chemical properties of the pollutant and the conditions of the surrounding ecosystem. Focusing on the soil environment, xenobiotics can migrate from it to the surrounding compartments: air (volatilization), water (leaching in the case of groundwater, lateral drainage and runoff and transport with sediment eroded in the case of surface water), or living organisms (uptake processes) (Fig. 3.2). From an ecotoxicological point of view, the uptake by living organisms (whether animals, plants or microorganisms) is a process of major concern, as it represents the necessary (but not sufficient) conditions to have ecotoxicological effects. Xenobiotics with low vapour pressures, high Kow (octanol-water partition coefficients, a measure of chemicals hydrophobicity), and low degradability have a low tendency to escape (i.e. fugacity) towards air and water compartments, and are thus the ones of major interest for soil environments.

Another relevant concept in ecotoxicology is that of bioavailability, which is the capacity of a xenobiotic to be actively or passively absorbed by living organisms. The peculiar problem of assessing soil xenobiotics bioavailability is related to the fact that the soil environment is structured

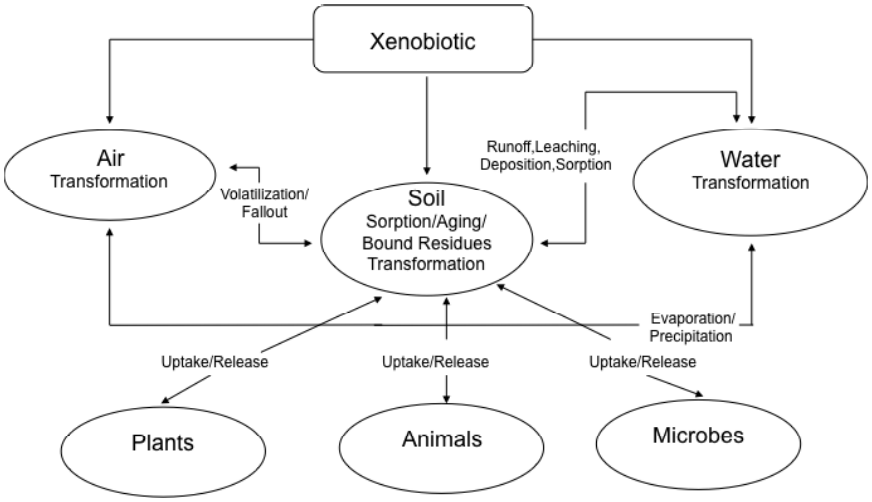


Figure 3.2. Processes controlling the distribution of a xenobiotic in the soil and in the surrounding environments. Once released in the environment, a xenobiotic undergoes a number of processes controlling its fate. In soil, it can be sorbed, bound or transformed in other chemical forms. It can also move to the air through volatilization or to the water through runoff or leaching. A portion of what is remaining ~~the soils,~~ termed the bioavailable, fraction can finally be uptaken by biological receptors (plants, animals or microbes). This bioavailable fraction is the one that can potentially pose a risk. Unpublish[illegible]terial of the authors.

Should this be 'remaining in the soils

in aggregates and pores of various dimensions (Ø µm-cm) and constituted by reactive minerals and organic substances. Once entering into the soil, xenobiotics can indeed be trapped into micropores or be adsorbed onto organic and inorganic particles (Fig. 3.3). Therefore, normally the xenobiotic bioavailable fraction does not correspond to its total concentration, and may not be entirely available for uptake by ecological receptors. Furthermore, entrapment and sorption processes generally increases with time, and the aging phenomenon has to be taken into account for a meaningful definition of xenobiotic bioavailability in soil. Due to the large variability of the soil properties, the bioavailable fraction differs in various types of soils and xenobiotics, but it also depends on the ecological considered receptor.

The xenobiotic availability in soil has been traditionally estimated by chemical extraction methods: a summary of these methods is presented in Table 3.1. In this chapter we show how biosensor technology can integrate the information given by chemical methods with more biologically meaningful information on xenobiotic bioavailability, in order to provide an improved and integrated approach to ecotoxicological risk assessment in soils.

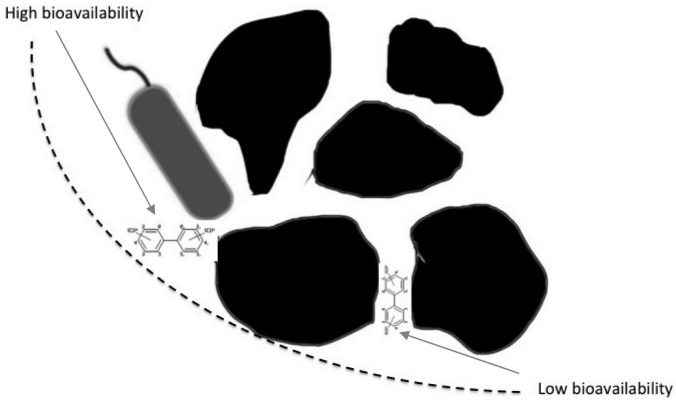


Figure 3.3. Schematization of bioavailability processes in soil environments. Xenobiotics can be trapped within soil particles or within the organic matter, and their biovailability for ecological receptors such as bacteria is reduced. Unpublished material of the authors.

Color image of this figure appears in the color plate section at the end of the book.

Table 3.1. Examples of chemical extraction methods for the assessment of the bioavailable fraction of organic xenobiotics in soils.

Extraction method	Description	Reference(s)
Cyclodextrins	Modified cyclodextrins (e.g. hydroxypropyl-β-cyclodextrin, HPCD) are used to entrap xenobiotics such as PAHs or PCBs, thus increasing their solubility.	Semple et al. 2003 Puglisi et al. 2007
Hydrophobic resins	Resins such as XAD or Tenax are used to desorb and quantify xenobiotics labile fractions	Northcott and Jones 2001
Mild solvents	Use of mild solvents such as buthanol to increase the solubility of hydrophobic xenobiotics	Kelsey et al. 1997
Persulphate oxidation	The fraction of PAHs oxidized by persulphate is used as an estimate of the bioavailable fraction	Cuypers et al. 2001
Supercritical fluid extraction	The bioavailable fraction is extracted by putting samples in pressure and temperature controlled conditions in chambers flushed with supercritical CO_2	Björklund et al. 1999

In ref.--2000

Different extraction methods for the quantification of the bioavailable fraction have been developed. These methods can be coupled with biosensors in order to achieve an integrated ecotoxicological risk assessment.

BIOSENSORS WITH FOCUS ON SOIL APPLICATIONS

A biosensor can be defined as a measurement device or system composed of a biological sensing component, which recognizes a chemical or physical change, coupled to a transducing element that produces a measurable signal

in response to the environmental change (Daunert et al. 2000) (Fig. 3.4). The sensing component classifies the biosensors in one of the three basic types such as: molecular, tissue and cellular (Pancrazio et al. 1999). The sensing elements of molecular biosensors can be subcellular components or macromolecules such as nucleic acids, lipid bilayers, enzymes, ion channels or antibodies. Tissue based biosensors are derived from intact tissue, whereas the whole-cell biosensors are usually bacterial cells producing a measurable signal or product.

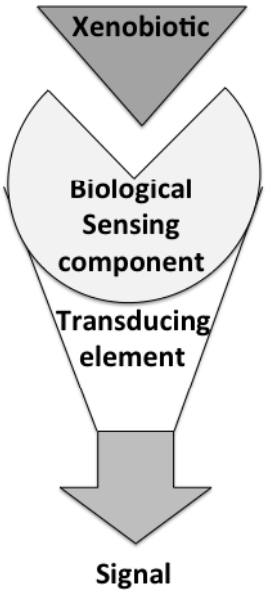


Figure 3.4. Schematic representation of a biosensor for detection of xenobiotics. A biosensor is made up of different components assuring its functioning: a biological sensing component (e.g. a microorganism, a tissue, an enzyme or an antibody) recognizes a specific xenobiotic or class of xenobiotics, and through a transducing element produces a signal (e.g. light or fluorescence) easily measurable. Unpublished material of the authors.

Whole Cell Biosensors

Bacteria have long being used as models for exploring the dose-dependent toxicity and mutagenicity of specific analytes in drug tests and monitoring of environmental contamination. The first use of a bacterial strain dates back to the Microtox test which was introduced in the early 1980s and it is based on the natural bioluminescence of the marine bacterium *Vibrio fischeri* to assess samples toxicity (Bulich and Isenberg 1981). The basic assumption for this test is that bioluminescence is an energy- and cofactor-demanding process and any factor reducing the bioluminescence indicates that bacteria

have been exposed to a toxicant. Microtox is a typical example of the so-called "light-off" or non-specific biosensor: similar "light-off" biosensor have been developed by inserting *lux* genes into enteric bacterial strains such as *Escherichia coli* or soil borne bacteria such as *Pseudomonas fluorescens* (Hamin-Hanjani et al. 1993). The lack of specificity is the major limitation of these 'light off' biosensors, which may also produce false positive responses to organic xenobiotics if they are metabolized and produce cellular reducing potential leading to an increase of bioluminescence.

Not listed in ref.

A major step forward in biosensor technology has been made by the production of the first specific biosensor, constructed by the insertion of a promoterless *lux* gene from *Vibrio fisherii* into the soil borne bacterium *Pseudomonas fluorescens* under the control of the promoter *nahG* gene, specifically activated by salycilate, the metabolite produced after naphthalene uptake (King et al. 1990). This work demonstrated the possibility of stabilizing a genetic construct and transmitting it to the following generations in ecological representative microorganism. It is also the first example of a specific "light-on" biosensor, where the luminescence is emitted only in response to the assimilation of specific chemicals or classes of chemicals. Since then, a large number of biosensors have been created for analysis of the bioavailability of the major classes of organic xenobiotics, but also for trace elements, nutrients and ecological interactions, with important applications for soil systems.

"Light-on" biosensors are characterized by two main elements: a reporter gene and a promoter (or responsive element). The concept of the reporter gene is relatively old in basic biology and biochemistry, and it can be defined as a gene whose phenotypic expression can be easily detected. Reporter genes traditionally used in microbiology and biochemistry are *lacZ*, coding for β-galactosidase, or *xylE* coding for catechol 2,3-dioxygenase, but they are not useful for monitoring microbial responses in soil, because their products cannot be easily recovered or targeted against the high background in soil samples. Differently, bioluminescence-based reporter genes, expressing light emitting molecules, such as the promoterless *lux* (bacterial luciferase), *luc* (firefly luciferase) or *gfp* (green fluorescent protein encoding gene) are more suitable to study activity and location of microorganisms in the soil environment.

Four genes have been successfully used to construct whole cell reporter strains, mainly due to the possibility to detect their products in the soil environment. The *luxCDABE* gene cassette codes for the luciferase enzyme, which emits constant light when supplied with oxygen and intracellular energy, whilst bioluminescence is reduced by uncoupling factors, loss of membrane integrity, or direct enzyme inhibition. The *gfp* (Table 3.2), isolated from the jellyfish *Aequorea victoria*, encodes for the green fluorescent protein

Table 3.2. Key features of fluorescent proteins.

1.	The first fluorescent protein discovered was the green fluorescent protein (GFP), a component of the bioluminescent organs of the jellyfish *Aequoria victoria.*
2.	Three scientists (Osamu Shimomoura, Martin Chalfie and Roger Y. Tsien) received in 2008 the Nobel Prize in Chemistry for the discovery of GFP and development of GFP-related applications.
3.	Other fluoresecent proteins, such as the yellow and the red fluorescent proteins (YFP and RFP) have been also engineered.
4.	Scientists can now clone fluorescent proteins and insert them into living cells of different organisms and visualize the location and the dynamics of the gene product using fluorescence microscopy.
5.	Fluorescent based biosensors are constructed by inserting the fluorescence genes under the control of promoters activated by specific compounds.
6.	The advantage of this approach for soil studies is that fluorescent proteins respond fast, and have no interferences with soil constituents.
7.	It is possible to insert more fluorescent proteins each under the control of a different promoter in a single organism, thus obtaining multiple-specific biosensors.

Fluorescent proteins have very important applications in science, as acknowledged by the Nobel Prize in Chemistry awarded in 2008 to their three discoverers. In the case of biosensors application for soils, fluorescent proteins present several advantages as compared to other reporter systems.

(GFP), a protein of 27 kDa size that converts by chemiluminescence the blue light (395 nm) of the Ca^{2+} -photoprotein (aequorin) into green light (510 nm). The chromogenic part of the protein is a tripeptide, which requires oxygen for maturation. The cDNA of the *gfp* was firstly cloned by Prasher et al. (1992) and afterwards introduced in intact cells and/or organelles by Inouye and Tsuji (1994). The main advantages of the gfp-based whole cell biosensors are that no substrates are required to perform the assays, GFP protein is stable in a broad range of pH (6–12) and resists to different proteases. The latter two protein characteristics are particularly important for soil applications, because soil pH may vary in different soils and because the soil holds an intense proteolytic activity either due to the biontic or abiontic proteases (see chapter X). The disadvantage of GFP is that the formation of the GFP active form requires some hours and that under UV light excitation several organic and inorganic soil constituents are also fluorescent. However, reporters with mutant unstable GFP having a rapid intracellular turnover (Andersen et al. 1998) allows the construction of gfp-based reporter bacteria, which respond faster and no interferences by soil constituents in the fluorescence measurements have been reported.

When a specific biosensor for organic xenobiotics is required, the most useful promoter genes are usually the ones involved in degradative pathways. Specific biosensors have been developed employing genes involved in the degradation of xenobiotic compounds such as phenols (Shingler and Moore 1994), middle-chain alkanes (Sticher et al. 1997), salycilates (King et al. 1990), chlorobenzoates from PCBs degradation (Boldt et al. 2004), and dioxins (Garrison et al. 1996).

Chemosensors

Chemical sensors (chemosensors) consist of two main components: a receptor and a detector. Receptors include enzymes, antibodies, and lipid layers, and are responsible for the selectivity of the sensor. The detector is not selective and acts as transducer, because it translates the physical or chemical changes of the receptor into an electrical signal. Detectors include transduction platforms having electrochemical (potentiometric, amperometric, impedance), piezoelectric, thermal or optical (reflectrometry, interferometry, optical waveguide lightmode spectroscopy, total internal reflection fluorescence, surface plasmon resonance) properties.

The chemosensors for environmental monitoring based on the use of piezoelectric detectors are among the most commonly used. Piezo-electric crystals (e.g. quartz) vibrate under the influence of an electric field, and the variations of the resonant frequency of an oscillating piezoelectric crystal in relation to the mass deposited on the crystal surface are used as an index of interactions between the receptor and the analyte. The main differences in the piezoelectric chemosensors concern the physical dimensions of the quartz plate and the thickness of the deposited electrode.

The piezoelectric DNA-based biosensors are constructed by immobilizing double stranded DNA (dsDNA), and are then placed in contact with the environmental liquid phase or extracts, allowing the contact between DNA and eventual environmental pollutants. As for the other types of sensors, in the analysis of soil and soil solution the extraction conditions are a critical step. A DNA-based biosensor for the qualitative/semiquantitative detection of genotoxic effects of various aromatic xenobiotics in soil was presented by Bagni et al. (2005), and its response in soils polluted by benzene, naphthalene and anthracene derivatives were in agreement with standard plant and animal toxicity, and the comet tests. This paper also well illustrated the different responses obtained using different extractants and extraction conditions.

Supramolecular Sensors

Supramolecular chemistry is the study of the complex multimolecular species formed by aggregation of relatively simpler molecules (Ganjali 2006). Supramolecular chemistry has been used for constructing sensitive membranes, selective for particular analytes, by matching the size and binding properties of the species to enhance the sensor selectivity. Ceresa et al. (2001) used ion-selective membranes based on the Pb-selective ionophore thioacetic acid dimethylamide, in this way they established that potentiometric polymeric membrane electrodes based on electrically neutral

In ref.--Ganjali et al. 2006

ionophores are useful analytical tools for heavy metal ion determinations in drinking water at nanomolar total concentrations.

APPLICATION OF BIOSENSORS FOR ECOTOXICOLOGICAL RISK ASSESSMENT OF XENOBIOTICS IN SOILS

The potential application of biosensors in different stages of the basic risk assessment procedure is graphically illustrated in Fig. 3.5. As shown, the hazard can be identified through the application of a range of different biosensors, by selecting those with wider applicability and more ecological relevance. Wide applicability and ecological relevance are important pre-requisites to allow the biosensors to be accepted as tools in risk assessment future and legislation. From this point of view, owing to the biological complexity, responses of whole cell biosensors offer a more meaningful picture of the exposure and toxicity of xenobiotics than of chemosensors, which are more suitable for detection, particularly specific biosensors which can be applied for a deeper screening of toxicity, and to obtain dose-response curves and ecotoxicological end-points such as EC_{50}s and NOECs. Biosensors can also be very useful to determine the exposure level, since they provide a direct measure of the amount of xenobiotic that is actually able to cross the cell membrane. This approach also shares the possibility of selecting ecologically representative microbial strains as hosts: compared to

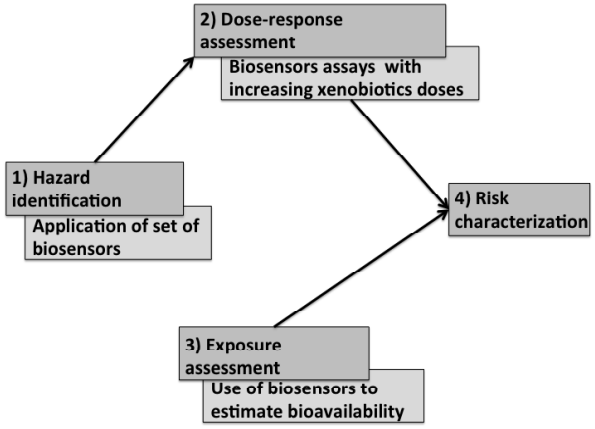


Figure 3.5. Modification of the risk assessment paradigm scheme to take into account the possible contribution of biosensors in soil ecotoxicological risk assessment. **Biosensors can be used to improve the risk paradigm in the case of ecotoxicological risk assessment of xenobiotics in soils.** 1) Different biosensors can be used as a screening tool to identify the major classes of xenobiotics present in a contaminated site; 2) specific biosensors are then used to carry out dose response assessments and obtain ecotoxicological endpoints; 3) biosensors can also be used as tools to estimate the bioavailable fraction of xenobiotics; 4) all the information is used to produce a risk characterization. Unpublished material of the authors.

classical biosensors tools such as the marine bacterium *Vibrio fisherii*, more representative strains (e.g. *Pseudomonas fluorescens* for soil) should be chosen. As a final step, all obtained information is elaborated to produce a risk assessment (Fig. 3.5). Biosensors of course do replace all tools necessary for risk assessment: their role is complementary, by providing information, such as ecotoxicity or real exposure, not obtainable with classical tools such as chemical analyses and environmental modelling. Chemical methods are also very useful to improve the application of biosensors. It is indeed difficult to apply biosensors directly on soil. *In situ* application of biosensors has been successfully achieved for nutrients status report (Puglisi et al. 2008), but for xenobiotics the situation is more difficult because of the quenching phenomena disturbing the expression of the signal. A common approach is thus to extract the xenobiotic from the soil, and then expose the biosensors to the extracts. If this is done for an ecotoxicological risk assessment, the bioavailable and not the total extractable fraction is relevant. At present several extraction methods have been developed and reviews on the topic published (Semple et al. 2003). Examples of some of these methods are reported in Table 3.1.

An example of the application of biosensors coupled to chemical methods for the extraction of the bioavailable fraction, has been given in the case of dioxins. Dioxins and furans are extremely toxic chemicals, formed as a by-product of industrial processes involving Cl such as waste incineration, chemical and pesticide manufacturing, with proven adverse effects on human and animal health in various parts of the world. Dioxins are characterized by a long persistence in the environment. The most known toxic compound is the 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD), and the toxicity of the dioxin family as well as other chemicals with comparable effects (e.g. PCB) are measured in relation to TCDD. Kurosawa et al. (2006) developed an immunosensor with a modified anti-TCDD (2,3,7,8-tetrachlorodibenzo-dioxin) monoclonal antibody as the molecular recognition component on the chemically activated surface of the electrode and a high sensitive bisphenol-A detection method using a signal amplification protocol to solve the problem of low signal emission. The immunosensor method described by Kurosawa has demonstrated its effectiveness as an alternative screening method for environmental monitoring because these results were compared with results obtained through environmental monitoring methods such as GC/MS and ELISA. Another biosensor system for dioxin compounds has been developed by modification of rat hepatoma cells with insertion of luciferase gene under the control of the aryl hydrocarbon receptor: it is known as DR-CALUX (Dioxin Response Chemical Activated Luciferase eXpression) and it is widely used to assess the presence and toxicity of dioxin and dioxin-like xenobiotics in food or environmental matrixes (Garrison et al. 1996; Besselink et al. 2004).

Please check

As compared to the chemical specific immunosensor, the DR-CALUX is more "mechanism" specific: in fact, the light emission is proportional not only to the total concentration of dioxins but also to their relative toxicity. This offers several advantages for risk assessment, although the above mentioned limitations should be born in mind when the total rather than the bioavailable fraction is used.

To evaluate this potential limitation, Puglisi et al. (2007) conducted a study coupling non-exhaustive extraction techniques with resins and cyclodextrins for bioavailability assessment with the DR-CALUX assay, in an ecotoxicological risk assessment of dredged sediments historically contaminated with PCBs and dioxins. Results showed that after a long time a fraction varying from 38 to 70% of the total xenobiotic contents was bioavailable, and that the coupling of non-exhaustive extraction techniques and DR-CALUX bioanalyses lead to a lower risk estimated for as compared to commonly adopted exhaustive extraction techniques.

Biosensors can also be used for ecotoxi ical risk assessment of other xenobiotics such as BTEX compounds, organochlorinateds compound and antibiotics represent examples of interest. The mono-aromatic volatile compounds like benzene, toluene, ethylbenzene, xylenes, collectively termed BTEX are emitted from various human and industrial activities and undergo variable partitioning and distribution between the solid and liquid phases of water, soil, and vegetation. While the emission of the BTEX has led to environmental concern for their contribution to global warming, caused by their chemical reactivity and their potential to reduce the ozone in the troposphere, the BTEX, as well as other volatile xenobiotics, also have harmful effects to human health even at lower concentrations. In fact, they can affect different organs such as nervous systems, liver, kidney, reproductive apparatus and cause asthma in children, particularly in the urban environment (Hinwood et al. 2007).

Whole cell biosensors responding to bioavailable benzene, toluene and xylene (BTEX) in soil have been constructed by inserting the *luxCDBAE* or *luc* genes in different plasmidial genes involved in the catabolic pathways of such compounds, using *Pseudomonas putida* strains (Burlage et al. 1994) or *Escherichia coli*. The insertion of different genes improve the sensitivity towards specific BTEX compounds because their degradation to pyruvate or acetaldehyde is controlled by several intermediates and enzymes. To date, there is no comprehensive comparison between the BTEX chemosensor responses and the models developed for assessing their chemical fate in the environment.

Normal? see p.54

In ref.--Burlage and Kuo 1994

Available whole cell biosensors for detection of chlorinated and organic polychlorinated pollutants are based on a *Pseudomonas fluorescens* 10586s, *Burkholderia sp. Rasc* strain or equipped with a complete *luxCDABE* cassette on the same plasmid pUCD607 (Palmer et al. 1998; Boyd et al. 2001), or

See p. 53

on *Pseudomonas fluorescens* 8866 and *Pseudomonas putida* (Weitz et al. 2001) with chromosomally integrated luxCDABE responding to soil pollution by different chlorobenzene derivatives. However, the latter two strains also responded to heavy metals such as Zn and Cu (Daunert et al. 2000). Whole cell biosensors for detection of polychlorinated biphenils (PCBs) have been obtained by fusions of luxCDBAE cassette with different genes of plasmidial operons of *Ralstonia eutropha* strains, coding for the degradative enzymes regulated by the presence of monochlorinated or polychlorinated biphenyls (Layton et al. 1998). Two gfp-based whole cell biosensors sensitive to 2,4-dichlorophenoxyacetic acid have been constructed by the fusion of gfp with tfdCI gene in *Ralstonia eutropha* (Füchslin et al. 2003) and by the fusion of gfp with orf0-bphA1 in *Pseudomonas fluorescens* (Boldt et al. 2004). A luc-based *Arthrobacter chlorophenolicus* strain responding to 4-chlorophenol in soil has been successfully used by Elvang et al. (2001) for monitoring soil remediation.

Another class of whole cell biosensors includes strains capable of detecting groups of substances functionally similar such as antibiotics (Bahl et al. 2004) and endocrine disruptors (Desbrow et al. 1998). Such bioreporters have been constructed inserting reporter genes in DNA coding for co-regulation mechanisms, as structurally similar compounds may activate common intracellular reactions. An early example of this type of whole cell biosensors can be considered the *Pseudomonas putida* TVA8 bioreporter, in which the chromosomal the tod genes were fused with a complete luxCDABE cassette, controlled by the Ptod promoter. Thus emitted bioluminescence in response to benzene, toluene and xylene isomers, and phenol (Applegate et al. 1998). These type of biosensors may be useful to study the fate of different xenobiotics in the rhizosphere, as they can be partially or totally degraded or transformed in different substances in soil. However, none of these bioreporters have been used for studying soils. In this class of whole cell biosensors, multiple stress-responsive bioreporters may also be included which are particularly suited to study the rhizosphere and soil conditions, as bacterial physiology under natural conditions (e.g. exposure to fluctuating temperature or osmolarity, variable moisture content, presence of radicals) are controlled by the expression of global regulatory metabolic genes (regulons). Stress-responsive bioreporters respond to a broad range of conditions and can be useful to characterize a soil environment in terms of general cytotoxicity. Due to the extended knowledge on such systems in enteric bacteria, *Escherichia coli* and *Salmonella typhimurium* have been used in early works, e.g. linking the lux reporter system to general stress promoters, such as the heat shock promoters dnaK, protein-damage sensitive grpE, oxidative-damage sensitive katG or membrane-damage sensitive fabA promoters (Bechor et al. 2002). More recently, lux constructs soil-borne *Pseudomonas* species have been genetically

Is this correct?

engineered for bioreporters of genotoxicity in environmental samples based on the SOS response involving DNA repair (Rabbow et al. 2002).

APPLICATIONS TO AREAS OF HEALTH AND DISEASE

The examples reported below show how an integrated use of biosensors and chemical extraction techniques is very useful for an ecotoxicological risk assessment of xenobiotics in soils. Being an ecotoxicological approach, the main focus is on ecological receptors in soils, primarily bacteria. It should be however taken into account that most soil xenobiotics considered here (e.g. chlorinated pollutants or BTEX compounds) also hold relevance for human health, and that soil is very often the primary route through which these compounds can enter the food chain and finally reach human receptors. For these reasons, we believe that an ecotoxicological risk assessment in soils is also very important to prevent xenobiotics-related human health issues.

Recent advances in both metabolic pathways engineering capacities, and detection technologies along with discovery of new mutants, driven primarily by studies in medicine, pharmacology and drug discovery, let us also hypothesize that more sensitive and specific whole-cell biosensors will be available for application to soil studies with relevance for human health too.

KEY TERMS

Is this correct?

- **Aging**: reduction of a xenobiotic's bioavailability in time.
- **Total xenobiotic fraction**: the total mass of xenobiotic that can be extracted from soil with the most-exhaustive methods available.
- **Bioavailable contaminant fraction**: the fraction of xenobiotic accessible for assimilation and possible toxicity to biological receptors. It is the relevant fraction from an ecotoxicological point of view.
- **Ecotoxicological risk assessment**: procedure to assess the risk posed by environmental pollutants to representative ecological receptors. As every risk assessment, is the resultant of a hazard and an exposure assessment.
- **Ecotoxicity**: the property of a chemical to cause harmful effects on ecosystems.
- **Contaminant**: a substance present in greater than natural concentration as a result of human activity.
- **Pollutant**: a substance that occurs in the environment at least in part as a result of man's activities, and which has a deleterious effects on living organisms (Moriarty 1983).

- **Xenobiotic**: a foreign chemical or material not produced in nature and not normally considered a constitutive component of a specific biological system (Rand and Petrocelli 1985).

KEY FACTS ABOUT THE USE OF BIOSENSORS FOR THE ECOTOXICOLOGICAL RISK ASSESSMENTS OF XENOBIOTICS IN SOILS

- Environmental contamination by xenobiotics can potentially result in adverse effects for both human and ecological receptors.
- Ecotoxicological risk assessment is a procedure where information about the ecotoxicity of a xenobiotic is linked to information about the degree of exposure of ecological receptors to the xenobiotic.
- In the soil environment, ecological receptors are exposed to a fraction of the total xenobiotics mass: the bioavailable fraction.
- Biosensors are useful tools for ecotoxicological risk assessment of xenobiotics: they can provide information on exposure, ecotoxicity or on both.
- A correct ecotoxicological risk assessment of xenobiotics in soils can be achieved by coupling chemical methods for the assessment of the bioavailable fraction of the xenobiotics with the application of specific biosensors.

SUMMARY POINTS

- Risk assessment is made up of four steps: hazard identification, exposure assessment dose-response assessment and risk characterization.
- Knowledge of the environmental fate of xenobiotics in soils is necessary to assess their ecotoxicological risk.
- A wide range of biosensors can be used for ecotoxicological risk assessment of xenobiotics. The main classes include whole cell biosensors, chemosensors and supramolecular sensors.
- Biotechnology allows construction of biosensors specific for many different compounds, including important soil xenobiotics that also hold also relevance for human health.
- A correct application of biosensors for ecotoxicological risk assessment of xenobiotics in soil is important not only for assessing and promoting soil quality, but also for preventing xenobiotic-related human health issue.

ABBREVIATIONS

BTEX	:	benzene, toluene, ethylbenzene, xylenes
EC_{50}	:	Effective Concentration (for 50% of population)
GFP	:	green fluorescent protein
HPCD	:	hydroxypropyl-β-cyclodextrin
NOEC	:	No Observable Effect Concentration
PAHs	:	Polycyclic Aromatic Hydrocarbons
PCBs	:	PolyChloroBiphenyls

REFERENCES

Amin-Hanjani S, A Meikle, L Glover, J Prosser and K Killham. 1993. Plasmid and chromosomally encoded luminescence marker systems for detection of *Pseudomonas fluorescens* in soil. Mol Ecol 2: 47–54.

Andersen JB, C Sternberg, LK Poulsen, SP Bjorn, M Givskov and S Molin. 1998. New unstable variants of green fluorescent protein for studies of transient gene expression in bacteria. Appl Environ Microbiol 64: 2240–2246.

Applegate B, S Kehrmeyer and G Sayler. 1998. A chromosomally based *tod-luxCDABE* whole-cell reporter for benzene, toluene, ethybenzene, and xylene (BTEX) sensing. Appl Environ Microbiol 64: 2730–2735.

Bagni G, S Hernandez, M Mascini, E Sturchio, P Boccia and S Marconi. 2005. DNA biosensor for rapid detection of genotoxic compounds in soil samples. Sensors 5: 394–410.

Bahl MI, LH Hansen, TR Licht and SJ Sorensen. 2004. *In vivo* detection and quantification of tetracycline by use of a whole-cell biosensor in the rat intestine. Antimicrob Agents Ch 48: 1112–1117.

Bechor O, DR Smulski, TK Van Dyk, RA LaRossa and S Belkin. 2002. Recombinant microorganisms as environmental biosensors: pollutants detection by *Escherichia coli* bearing *fabA'':: lux* fusions. J Biotechnol 94: 125–132.

Besselink HT, C Schipper, H Klamer, P Leonards, H Verhaar, E Felzel, AJ Murk, J Thain, K Hosoe, G Schoeters, J Legler and B Brouwer. 2004. Intra- and interlaboratory calibration of the DR CALUX bioassay for the analysis of dioxins and dioxin-like chemicals in sediments. Environ Toxicol Chem 23: 2781–9.

Björklund E, S Bøwadt, L Mathiasson and SB Hawthorne. 1999. Determining PCB sorption/desorption behavior on sediments using selective supercritical fluid extraction. 1. Desorption from historically contaminated samples. Environ Sci Technol 33: 2193–2203.

Boldt TS, J Sørensen, U Karlson, S Molin and C Ramos. 2004. Combined use of different Gfp reporters for monitoring single cell activities of a genetically modified PCB degrader in the rhizosphere of alfalfa. FEMS Microbiol Ecol 48: 139–148.

Boyd EM, K Killham and AA Meharg. 2001. Toxicity of mono-, di- and tri-chlorophenols to lux marked terrestrial bacteria, *Burkholderia* species *Rasc c2* and *Pseudomonas fluorescens*. Chemosphere 43: 157–166.

Bulich A and D Isenberg. 1981. Use of the luminescent bacterial system for the rapid assessment of aquatic toxicity. ISA Trans (United States) 20.

Burlage RS and C Kuo. 1994. Living biosensors for the management and manipulation of microbial consortia. Annu Rev Microbiol 48: 291–309.

Ceresa A, E Bakker, B Hattendorf, D Günther and E Pretsch. 2001. Potentiometric polymeric membrane electrodes for measurement of environmental samples at trace levels: new requirements for selectivities and measuring protocols, and comparison with ICPMS. Anal Chem 73: 343–351.

Cuypers C, T Grotenhuis, J Joziasse and W Rulken. 2000. Rapid persulfate oxidation predicts PAH bioavailability in soils and sediments. Environ Sci Technol 34: 2057–2063.

Daunert S, G Barrett, JS Feliciano, RS Shetty, S Shrestha and W Smith-Spencer. 2000. Genetically engineered whole-cell sensing systems: coupling biological recognition with reporter genes. Chem. Rev. 100: 2705–2738.

Desbrow C, E Routledge, G Brighty, J Sumpter and M Waldock. 1998. Identification of estrogenic chemicals in STW effluent. 1. Chemical fractionation and *in vitro* biological screening. Environ. Sci. Technol. 32: 1549–1558.

Elväng AM, K Westerberg, C Jernberg and JK Jansson. 2001. Use of green fluorescent protein and luciferase biomarkers to monitor survival and activity of *Arthrobacter chlorophenolicus* A6 cells during degradation of 4 chlorophenol in soil. Environ Microbiol 3: 32–42.

Füchslin HP, I Rüegg, JR Van Der Meer and T Egli. 2003. Effect of integration of a GFP reporter gene on fitness of Ralstonia eutropha during growth with 2, 4 dichlorophenoxyacetic acid. Environ Microbiol 5: 878–887.

Ganjali MR, P Norouzi, M Rezapour, F Faridbod and MR Pourjavid. 2006. Supramolecular based membrane sensors. Sensors 6: 1018–1086.

Garrison PM, K Tullis, JM Aarts, A Brouwer, JP Giesy and MS Denison. 1996. Species-specific recombinant cell lines as bioassay systems for the detection of 2,3,7,8-tetrachlorodibenzo-p-dioxin-like chemicals. Fundam Appl Toxicol 30: 194–203.

Hi[illegible]AL, C Rodriguez, T Runnion, D Farrar, F Murray, A Horton, D Glass, V Sheppeard, [illegible]dwards and L Denison. 2007. Risk factors for increased BTEX exposure in four Australian cities. Chemosphere 66: 533–541.

Inouye S and FI Tsuji. 1994. Aequorea green fluorescent protein: expression of the gene and fluorescence characteristics of the recombinant protein. FEBS Lett 341: 277–280.

Kelsey JW, BD Kottler and M Alexander. 1997. Selective chemical extractants to predict bioavailability of soil-aged organic chemicals. Environ Sci Technol 31: 214–217.

King JM, PM Digrazia, B Applegate, R Burlage, J Sanseverino, P Dunbar, F Larimer and GS Sayler. 1990. Rapid, sensitive bioluminescent reporter technology for naphthalene exposure and biodegradation. Science 249: 778–781.

Kurosawa S, JW Park, H Aizawa, SI Wakida, H Tao and K Ishihara. 2006. Quartz crystal microbalance immunosensors for environmental monitoring. Biosens Bioelectron 22: 473–481.

Layton A, M Muccini, M Ghosh and G Sayler. 1998. Construction of a bioluminescent reporter strain to detect polychlorinated biphenyls. Appl Environ Microbiol 64: 5023–5026.

Moriarty F. 1983. Ecotoxicology. Academic Press.

National Research Council. 1983. Risk assessment in the Federal Government: managing the process National Academy Press, Washington DC, USA.

Newman MC and MA Unger. 2003. Fundamentals of ecotoxicology. CRC press.

Northcott GL and KC Jones. 2001. Partitioning, extractability, and formation of nonextractable PAH residues in soil. 2. Effects on compound dissolution behaviour. Environ Sci Technol 35: 1111–1117.

Palmer G, R McFadzean, K Killham, A Sinclair and GI Paton. 1998. Use of *lux*-based biosensors for rapid diagnosis of pollutants in arable soils. Chemosphere 36: 2683–2697.

Pancrazio J, J Whelan, D Borkholder, W Ma and D Stenger. 1999. Development and application of cell-based biosensors. Ann Biomed Eng 27: 697–711.

Prasher DC, VK Eckenrode, WW Ward, FG Prendergast and MJ Cormier. 1992. Primary structure of the *Aequorea victoria* green-fluorescent protein. Gene 111: 229–233.

Puglisi E, AJ Murk, HJ van den Berg and T Grotenhuis. 2007. Extraction and bioanalysis of the ecotoxicologically relevant fraction of contaminants in sediments. Environ Toxicol Chem 26: 2122–2128.

Puglisi E, G Fragoulis, AA Del Re, R Spaccini, A Piccolo, G Gigliotti, D Said-Pullicino and M Trevisan. 2008. Carbon deposition in soil rhizosphere following amendments with compost and its soluble fractions, as evaluated by combined soil-plant rhizobox and reporter gene systems. Chemosphere 73: 1292–1299.

Please insert place of publication

Rabbow E, P Rettberg, C Baumstark-Khan and G Horneck. 2002. SOS-LUX- and LAC-FLUORO-TEST for the quantification of genotoxic and/or cytotoxic effects of heavy metal salts* 1. Anal Chim Acta 456: 31–39.

Rand GM and SR Petrocelli. 1985. Fundamentals of aquatic toxicology: methods and applications. FMC Corp., Princeton, NJ, USA.

Sanderson JT, J Aarts, A Brouwer, KL Froese, MS Denison and JP Giesy. 1996. Comparison of Ah receptor-mediated luciferase and ethoxyresorufin-O-deethylase induction in H4IIE cells: implications for their use as bioanalytical tools for the detection of polyhalogenated aromatic hydrocarbons. Toxicol Appl Pharm 137: 316–325.

Semple KT, A Morriss and G Paton. 2003. Bioavailability of hydrophobic organic contaminants in soils: fundamental concepts and techniques for analysis. Eur J Soil Sci 54: 809–818.

Shingler V and T Moore. 1994. Sensing of aromatic compounds by the DmpR transcriptional activator of phenol-catabolizing *Pseudomonas* sp. strain CF600. J Bacteriol 176: 1555–1560.

Sticher P, MC Jaspers, K Stemmler, H Harms, AJ Zehnder and JR van der Meer. 1997. Development and characterization of a whole-cell bioluminescent sensor for bioavailable middle-chain alkanes in contaminated groundwater samples. Appl Environ Microbiol 63: 4053–60.

Truhaut R. 1977. Ecotoxicology: objectives, principles and perspectives. Ecotox Environ Safe 1: 151–173.

Walker C, S Hopkin, R Sibly and D Peakall. 2001. Principles of ecotoxicology. Taylor and Francis, London, UK.

Weitz HJ, JM Ritchie, DA Bailey, AM Horsburgh, K Killham and LA Glover. 2001. Construction of a modified mini Tn5 l*uxCDABE* transposon for the development of bacterial biosensors for ecotoxicity testing. FEMS Microbiol Lett 197: 159–165.

Journal of Environmental Monitoring

Cite this: *J. Environ. Monit.*, 2012, **14**, 1383

www.rsc.org/jem

PAPER

In situ phytoremediation of a soil historically contaminated by metals, hydrocarbons and polychlorobiphenyls

S. Doni,*[a] C. Macci,[a] E. Peruzzi,[a] M. Arenella,[b] B. Ceccanti[a] and G. Masciandaro[a]

Received 20th December 2011, Accepted 21st February 2012
DOI: 10.1039/c2em11016d

In the past several years, industrial and agricultural activities have led to serious environmental pollution, resulting in a large number of contaminated sites. As a result, much recent research activity has focused on the application of bioremediation technologies as an environmentally friendly and economically feasible means for decontamination of polluted soil. In this study horse manure and *Populus nigra* (var. *italica*) (HM + P treatment) have been used, at real scale level, as an approach for bioremediation of a soil historically contaminated by metals (Pb, Cr, Cd, Zn, Cu and Ni) and organic contaminants, such as polychlorobiphenyls and petroleum hydrocarbon. After one year, the HM + P phytotreatment was effective in the reclamation of the polluted soil from both organic and inorganic contaminants. A reduction of about 80% in total petroleum hydrocarbon (TPH), and 60% in polychlorobiphenyls (PCBs) and total metals was observed in the HM + P treatment. In contrast, in the horse manure (HM) treatment, used as control, a reduction of only about 30% of TPH was obtained. In order to assess both effectiveness and evolution of the remediation system to a biologically active soil ecosystem, together with the pollution parameters, the parameters describing the evolution of the soil functionality (enzymatic activities and protein SDS-PAGE pattern) were investigated. A stimulation of the metabolic soil processes (increase in dehydrogenase activity) was observed in the HM + P compared to the HM treatment. Finally, preliminary protein SDS-PAGE results have permitted the identification of proteins that have been recovered in the HM + P soil with respect to the HM; this may become a basic tool for improving the biogeochemical status of soil during the decontamination through the identification of microbial populations that are active in soil decontamination.

1. Introduction

Co-contamination of inorganic and organic compounds is a very common feature in nature. Urbanization and industrialization have led to dramatically high release of anthropogenic pollutants into the environment. Prevalent contaminants include hydrocarbons, such as polycyclic aromatic hydrocarbons (PAHs), halogenated hydrocarbons, pesticides, solvents, metals, and salts. Several recent research activities have focused on the application of phytoremediation as a sustainable reclamation strategy for bringing soil polluted by organic and inorganic contaminants into productive use.[1,2] A great number of results indicates that the technique might become a suitable alternative to physical and chemical approaches in decontamination of metal polluted sites or a final polishing solution for organic contamination.[3,4] Compared to existing physical and chemical methods of soil remediation, which often generate secondary waste, phytoremediation is, in fact, cost-effective and less disruptive for the environment. This technology can be employed in any geographical area that can support plant growth and it is more likely to be socially accepted due to its aesthetical aspect.

[a]*Institute of Ecosystem Study, National Research Council, Pisa, Italy. E-mail: serena.doni@ise.cnr.it; Tel: +390503152476; Fax: +390503152473*
[b]*Department of Soil Science and Plant Nutrition, University of Florence, Firenze, Italy*

Environmental impact

The bioremediation of a soil contaminated by both organic (TPH and PCBs) and inorganic pollutants has scarcely been investigated at field scale. Moreover, the synergic effect of organic matter, poplar and microorganisms in the reduction of contaminants has been studied. The innovative approach characterized in this study is the application of SDS-PAGE in order to support general biochemical parameters.

An additional benefit of phytoremediation is the improvement of the chemical soil quality through organic materials, nutrients and oxygen supply by plant and microbial metabolic processes. Plants also improve the physical soil quality providing groundcover and stabilizing the soil with their roots.[5,6]

Several studies indicate that poplar (*Populus* spp.) trees may be highly suitable for phytoremediation purposes.[7,8] The high biomass production, root depth, as well as the resistance to different types of environmental stress make them good candidates for uptake and removal of organic and inorganic chemicals from polluted soils.[9] Poplars are known to take up and detoxify several organic pollutants, such as chlorinated hydrocarbon solvents, the herbicide atrazine, and different explosives.[9,10] Laureysens *et al.*[11] showed poplar to grow relatively well on moderately contaminated soil.

In recent years some studies have demonstrated the efficacy of organic matter application in supporting phytoremediation; the organic matter addition increased microbial biodiversity and activity,[2] nutrients availability, cation exchange capacity, porosity and water-holding capacity. All these characteristics enhance the soil health and provide a medium satisfactory for plant growth.

To evaluate changes in soil quality, biological parameters were evaluated in addition to physical and chemical ones. Enzyme activities in soil have been associated with indicators of biogeochemical cycles, degradation of organic matter and soil remediation processes, so they can contribute to assess, together with other physical or chemical properties, the soil quality.[12] Authors such as Eldor,[13] Nielsen and Winding[14] and Dick[15] reported enzyme activities as good indicators because they are closely related to organic matter, physical characteristics, microbial activity and biomass in the soil. In particular, dehydrogenase is an intracellular enzyme involved in organic matter oxidation and it is related to microbial biomass.[15] It has been considered as a soil quality indicator because it is involved in electron transport systems of oxygen metabolism and requires an intracellular environment (viable cells) to express its activity.[16] In addition, β-glucosidase was used as a key indicator due to its importance in catalytic reactions on cellulose degradation; releasing glucose as a source of energy, its activity enhances the maintenance of metabolically active microbial biomass in soil. Phosphatase plays a fundamental role in phosphorus solubilization with the release of low molecular weight organic acids.

In the last years a particular interest is developed for a metaproteomic approach to evaluate soil functionality. Soil proteomics may lead to the discovery of sub-lethal microbial biomarkers responding to soil management and pollution, and may improve our understanding on the degradation of organic pollutants and organic matter, nutrient cycles, and molecular communication between plants and between plants and microbial communities.[17] Changes in environmental parameters may alter microbial activities. A lot of proteins present in soil show metabolic functions that could be used to relate specific microbial activities to a definite microorganism; the assignment of proteins to specific microbial functional groups can provide deeper information about the functional and phylogenetic microbial role in soil ecosystem. To date, this approach is still in the early stage but could reveal great potentiality. The optimization of protocols to analyze proteins derived from soil by mass spectrometry and the development of databases to recognize these macromolecules could permit, in future, the detection of specific health condition markers of soils, as already happens in other sectors. Giagnoni *et al.*[18] reported the problems related to the extraction and identification of soil microbial proteins derived from the inoculation of *Cupriavidus metallidurans* CH34, which has a known proteome, into sterile sand, kaolinite, montmorillonite and a mixture of sand, kaolinite, montmorillonite, goethite and humic acids.

In the present investigation, a biological approach made up of native plants (*Populus nigra*) and horse manure has been proposed at real-scale to bioremediate and functionally recover a soil historically contaminated by metals, hydrocarbons and polychlorobiphenyls. The study of soil directly in the polluted site under natural conditions in the area can lead to more representative results. This approach was proposed after the satisfactory results of a preliminary meso-scale experiment carried out on the same contaminated soil. The proteomic approach by SDS-PAGE is just a first step that aims to obtain significant information about soil quality from molecular biology techniques.

2. Materials and methods

2.1 Experimental layout

The soil studied, historically contaminated by metals, hydrocarbons and polychlorobiphenyls, was located in an industrial area of about 0.5 ha (latitude 43°44′N and longitude 10°22′E) in San Giuliano Terme Municipality (Pisa, Italy).

The soil is a clay loam soil (USDA texture classification; 34.1% clay, 22.6% sand and 43.2% lime) and it is characterized by the presence of an uncontaminated clay basement at 1 m depth.

In October 2008 the contaminated soil was removed until the clay basement, the bulky wastes were removed manually and by sieving and, after mixing, the soil was replaced in the same field.

Immediately before bioremediation treatments (February 2009), five soil samples were randomly collected at 0–30 cm, homogenized and mixed together; the obtained soil sample (initial soil-*T*0) was analyzed in triplicate in order to obtain an initial soil characterization (Tables 2 and 3).

The treatments were the following: (i) *Populus nigra* var. *italica* plant and horse manure (HM + P) and (ii) horse manure (HM) used as control. The size of each of the two field plots (HM + P and HM treatments) was approximately 0.25 ha. *Populus nigra* plant species, which is very common in the Mediterranean area, was chosen for its rapid growth and the capability of developing in the specific substrate and local climatic conditions. *Populus nigra*, approximately 3 years old, was planted in February 2009 at a distance of 2 × 2 m and, immediately after planting (February 2009), horse manure was applied to the soil surface in both field plots (HM + P and HM) at the dose of 20 t ha^{-1} and incorporated into the soil by soft harrowing. According to a preliminary laboratory study, the horse manure addition is helpful in order to improve the chemical (permeability and nutrient content) and biological characteristics of the soil for a better plant adaptation and growth. The main chemical characteristics of horse manure are reported in Table 1.

Table 1 Major chemical properties of horse manure[a]

	Horse manure
pH	6.92 ± 0.103
TOC (%)	45.10 ± 1.18
TN (%)	1.44 ± 0.017
TP/mg P kg^{-1}	394 ± 24.5
WSC/mg C kg^{-1}	36 200 ± 429
N–NH_4^+/mg N kg^{-1}	118 ± 3.67
N–NO_3^-/mg N kg^{-1}	394 ± 12.2

[a] TOC, total organic carbon; TN, total nitrogen; TP, total phosphorus; WSC, water soluble carbon.

The climate in this area is predominantly Mediterranean with an average annual temperature of 15 °C and an average annual rainfall of 700 mm.

A drip irrigation system with frequency and time of irrigation changeable in relation to seasons and atmospheric conditions was carried out.

Soil samples were periodically collected from the experimental area at 0–30 cm depth (*T*1: March 2009; *T*2: July 2009; *T*3: February 2010). The soil samples in the HM + P treatment have been collected in an inter-row region corresponding to about 1 metre from the plant trunk. Three soil samples for each treatment were collected and separately analysed. Each one was composed of five subsamples collected in a defined 1 m^2 area. Soil samples were air dried, sieved (2 mm) and stored at room temperature prior to determining chemico-physical and biochemical properties. All the analyses were carried out in triplicate.

2.2 Chemical analysis

The particle size analysis was calculated by a pipette procedure.[19] pH was measured in a 1 : 10 w/v aqueous solution. Total organic carbon (TOC) and total nitrogen (TN) contents were determined by dry combustion in a RC-412 multiphase carbon and a FP-528 protein/nitrogen analyzer, respectively (LECO Corporation). Water soluble carbon (WSC) was extracted from the soil with distilled water in a 1 : 10 w/v aqueous solution by shaking at 60 °C for 1 h;[20] the C content of WSC was determined by dichromate oxidation.[21] N–NH_3 and N–NO_3^- were measured in an aqueous extract (1 : 10 w/v) with an ammonia-selective electrode (ORION 95-12) and DIONEX chromatograph, respectively. Total phosphorus (TP) and total metals were determined after acid digestion with nitric–perchloric acids by a colorimetric method[22] and atomic absorption spectrometry, respectively. The total petroleum hydrocarbons (TPHs) were determined by the gravimetric method according to ref. 23.

The extraction of polychlorobiphenyls (PCBs) was performed according to pressurized fluid extraction (PFE) EPA method 3545.[24] Subsequently, the extract was cleaned-up by gel permeation according to EPA method 3640.[25] Finally the PCBs concentration was determined by a gas chromatography/electron capture detector (GC/ECD). The PCBs analysis, PCBs being a recalcitrant portion of organic contaminants, was performed only at the beginning (*T*0 and *T*1) and the end (*T*3) of the experimentation. In fact, it has been estimated that the biodegradation half-life times of PCBs in sediments and soils vary from several years to decades.[26]

2.3 Biochemical analysis

To test β-glucosidase activity 0.05 M 4-nitro-phenyl-β-D-glucanopyranoside (PNG) was used as the substrate, while 0.115 M *p*-nitrophenyl phosphate (PNPP) was used as the substrate to measure the phosphatase activity. The *p*-nitrophenol (PNP) produced by both hydrolases was extracted and determined spectrophotometrically at 398 nm.[27]

Dehydrogenase activity was measured using 0.4% 2-*p*-iodophenyl-3-*p*-nitrophenyl-5-tetrazolium chloride (INT) as the substrate; iodonitrotetrazolium formazan (INTF) produced in the reduction of INT was measured by means of a spectrophotometer at 490 nm.[28]

2.4 Soil protein expression

The proteins were extracted from the soil with 0.5 M potassium sulfate pH 6.6 in a 1 : 3 w/v ratio using 100 g of soil sample. 10 mM EDTA was added as the metal–protease inhibitor, modifying the method reported by Masciandaro *et al.*[29] The extraction was carried out at 24 °C for 1 h under horizontal shaking at 100 oscillations min^{-1}. After the incubation the whole sample was centrifuged for 15 min at 12 400 × *g* and the supernatant was filtered through a 0.22 μm cellulose nitrate filter in order to remove microbial cells from the soil extract. Then the extract was diluted 4 times with bi-distilled water (Milli-Q) and dialysed overnight against distilled water using 3500 MW cellulose tubular membrane (Cellu Sep T1, Orange scientific). The dialysed extract was recovered and concentrated until a final volume of 1 ml by an Amicon PM-10 diaflomembrane (Millipore) under N atmosphere (1.5 bar).[30]

The crude protein extract was precipitated by the trichloroacetic acid (TCA)–Na deoxycholate detergent (DOC) method, which is for very low protein concentration.[31] The protein pellet was air-dried and then resuspended in 10 μl of 0.1 M NaOH.

In order to carry out the sodium-dodecylsulfate-polyacrylamide gel electrophoresis (SDS-PAGE) (4% stacking/12% resolving), the sample buffer (Laemmli)[32] 5× was added to the sample; the sample was heated at 100 °C for 3 min and centrifuged for 2 min. The run was carried out at 150 V until the tracking dye was near the bottom of the gel.

The gel was stained under gentle shaking overnight with colloidal Coomassie brilliant blue G250.

The stained gel was scanned by a Bio-Rad GS 800 densitometer. The software Quantity One (Biorad) was used for the construction of a phylogenetic tree based on the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) method.

2.5 Statistical analysis

The STATISTICA 6.0 software (StatSoft Inc., Tulsa, Oklahoma, USA) was used for the statistical analysis. All results are the means of three replicates ($n = 3$). All numerical parameters before statistical analysis were normalized and autoscaled: the result for each variable is a zero mean and a unit standard deviation.[33] Differences among the times (*T*0, *T*1, *T*2 and *T*3) and treatments (initial soil, horse manure and horse manure plus

plants treatments) were tested by analysis of variance (one way ANOVA). The means were compared by using least significant differences calculated at $P < 0.05$ (Fisher's test).

3. Results

3.1 Chemical parameters

The soil pH decreased during the time in both HM + P and HM treatments, reaching in the former a sub-acid value (6.6) at the end of the experimental period (Table 2). Total Organic Carbon (TOC) content showed an initial increase ($T1$) in both treatments with respect to the initial soil, probably due to the manure application. This parameter continued to increase over time in the HM + P treatment, while a significant decrease at $T2$ and $T3$ was observed in the horse manure (HM) treatment.

Referring to N and P concentrations no significant differences were observed between HM and HM + P treatments. In both treatments, in fact, total N content decreased during the time, while total P showed a slight increase (Table 2).

The products of the degradation of C and N compounds are water soluble carbon, ammonia and nitrates. The $N–NH_3$ content was significantly lower in the treated than in the initial soil, even though a slight increase was observed during the experimental period in both HM + P and HM treatments (Table 2). Furthermore, the better the oxygenation the easier the start-up of the nitrification processes, with a consequent transformation of ammonia into nitrates. In both treatments significantly higher $N–NO_3$ values with respect to $T1$ were, in fact, observed at $T2$ and $T3$ sampling times, while an increase in the $N–NO_3$ concentration during the time was only shown by the HM + P treatment (Table 2).

Water soluble C (WSC) increased significantly in the HM + P treatment over time, while in the HM treatment it showed a positive variation with respect to the initial soil only at $T2$ sampling time (Table 2).

3.2 Biological parameters

At $T1$ sampling time, the dehydrogenase activity, which is usually used as an indicator of overall microbial activity, was higher in both treatments (HM and HM + P) than in the initial soil; this enzyme continued to increase in the HM + P treatment, while it decreased from $T1$ to $T3$ in the HM treatment (Table 2).

β-Glucosidase, an enzyme linked to the C cycle, and acid phosphatase, an enzyme representative of the P cycle, activities showed a clear increasing trend during the time in both the treated soils (Fig. 1).

3.3 Inorganic and organic contaminants

At the end of the experimentation a decrease in inorganic contaminants was observed (Table 3, Fig. 2) in the HM + P treatment already from the $T1$ sampling time. Ni and Cu quickly (at $T1$) reached concentrations lower than the law limits (limit values for urban use 120 mg kg^{-1}, D.Lgs.152/2006), while Cd values were under the legislation limits at $T2$ (limit values for urban use 2 mg Cd kg^{-1}; D.Lgs.152/2006). Concerning Pb and Zn, a significant decrease was observed from $T1$ to $T3$, even though the values remained slightly higher than law limits (limit values for urban use 100 mg Pb kg^{-1}, 150 mg Zn kg^{-1}; D.Lgs.152/2006). Instead, Cr concentration, already under the law limit value (urban use 150 mg kg^{-1} D.Lgs.152/2006) in the initial soil, decreased further during the HM + P treatment. The HM

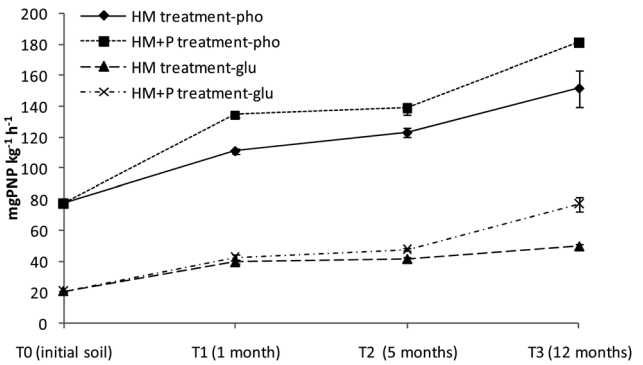


Fig. 1 Phosphatase (pho) and β-glucosidase (glu) activities in the initial soil at $T0$ and horse manure (HM) and horse manure and plant (*Populus nigra*) (HM + P) treatments at $T1$ (one month), $T2$ (five months) and $T3$ (12 months) sampling times.

Table 2 Chemical and biochemical parameters of the initial soil at $T0$ sampling time and horse manure (HM) and horse manure and plant (*Populus nigra*) (HM + P) treatments at $T1$ (one month), $T2$ (five months) and $T3$ (12 months) sampling times[a]

	Initial soil	HM			HM + P		
	$T0$	$T1$	$T2$	$T3$	$T1$	$T2$	$T3$
pH	7.69 ± 0.043	7.62 ± 0.007a(a)	7.49 ± 0.014b(a)	7.19 ± 0.064c(a)	7.81 ± 0.085a(a)	7.03 ± 0.014b(b)	6.63 ± 0.021c(b)
TOC (%)	1.29 ± 0.011	1.46 ± 0.016a(a)	1.10 ± 0.028b(b)	1.06 ± 0.011c(b)	1.40 ± 0.021b(a)	1.85 ± 0.081a(a)	1.80 ± 0.085a(a)
TN (%)	0.118 ± 0.002	0.121 ± 0.002a(a)	0.102 ± 0.002b(b)	0.076 ± 0.016c(b)	0.122 ± 0.005a(a)	0.119 ± 0.002a(a)	0.104 ± 0.011b(a)
TP/mg kg^{-1}	823 ± 17	560 ± 5.2b(a)	656 ± 48a(a)	698 ± 31a(a)	539 ± 4.3b(a)	639 ± 15.8a(a)	675 ± 22a(a)
WSC/mg kg^{-1}	189 ± 8.0	169 ± 8.2b(a)	207 ± 16a(b)	165 ± 18b(b)	139 ± 18b(a)	255 ± 7.6a(a)	248 ± 13a(a)
$N–NH_4^+$/mg kg^{-1}	4.97 ± 0.334	0.735 ± 0.07b(a)	0.702 ± 0.04b(a)	1.80 ± 0.20a(a)	0.470 ± 0.05b(b)	0.528 ± 0.06b(b)	1.62 ± 0.05a(a)
$N–NO_3^-$/mg kg^{-1}	33.1 ± 2.58	35.6 ± 5.1c(a)	59.8 ± 5.9a(a)	48.0 ± 0.4b(b)	38.8 ± 2.5c(a)	73.8 ± 6.3b(a)	191 ± 6.3a(a)
DHase activity/ mg INTF kg^{-1} h^{-1}	1.12 ± 0.008	1.20 ± 0.05a(a)	1.16 ± 0.03a(b)	0.970 ± 0.02b(b)	1.21 ± 0.03c(a)	1.52 ± 0.07b(a)	2.02 ± 0.08a(a)

[a] TOC, total organic carbon; TN, total nitrogen; TP, total phosphorus; WSC, water soluble carbon; DHase, dehydrogenase activity. The first letter indicates significant difference ($P < 0.05$) among the times ($T1$, $T2$ and $T3$) in each treatment. The letter in brackets indicates significant difference ($P < 0.05$) between the treatments at the same sampling time (HM and HM + P).

Table 3 Inorganic (metals) and organic (PCBs) contaminants in the initial soil at *T*0 sampling time and horse manure (HM) and horse manure and plant (*Populus nigra*) (HM + P) treatments at *T*1 (one month), *T*2 (five months) and *T*3 (12 months) sampling times[a]

	Initial soil	HM			HM + P		
	*T*0	*T*1	*T*2	*T*3	*T*1	*T*2	*T*3
Cr/mg kg^{-1}	40.3 ± 0.77	42.8 ± 2.0a(a)	36.4 ± 7.2a(a)	37.2 ± 4.5a(a)	35.8 ± 2.8a(a)	18.0 ± 2.3b(b)	15.1 ± 1.6b(b)
Cu/mg kg^{-1}	137 ± 4.24	146 ± 9.0a(a)	157 ± 13a(a)	156 ± 16a(a)	94.9 ± 14a(b)	69.0 ± 1.4b(b)	64.6 ± 8.0b(b)
Ni/mg kg^{-1}	131 ± 0.71	120 ± 8.9a(a)	97.2 ± 3.9b(a)	90.6 ± 8.1b(a)	89.8 ± 3.6a(b)	65.3 ± 3.9b(b)	62.2 ± 1.2b(b)
Cd/mg kg^{-1}	2.99 ± 0.14	3.14 ± 0.08a(a)	2.94 ± 0.07a(a)	2.75 ± 0.19a(a)	2.91 ± 0.10a(a)	1.50 ± 0.13b(b)	1.05 ± 0.10c(b)
Pb/mg kg^{-1}	554 ± 34	651 ± 73 ab(a)	764 ± 57a(a)	625 ± 38b(a)	516 ± 22a(b)	271 ± 21b(b)	211 ± 44b(b)
Zn/mg kg^{-1}	628 ± 11.9	590 ± 23a(a)	540 ± 55ab(a)	438 ± 54b(a)	474 ± 38a(b)	243 ± 16b(b)	220 ± 15b(b)
PCBs/mg PCB kg^{-1}	15.4 ± 1.23	15.9 ± 0.46a(a)		14.7 ± 1.03a(a)	15.1 ± 1.09a(a)		5.92 ± 0.32b(b)

[a] PCBs, polychlorobiphenyls. The first letter indicates significant difference ($P < 0.05$) among the times (*T*1, *T*2 and *T*3) in each treatment. The letter in brackets indicates significant difference ($P < 0.05$) between the treatments at the same sampling time (HM and HM + P).

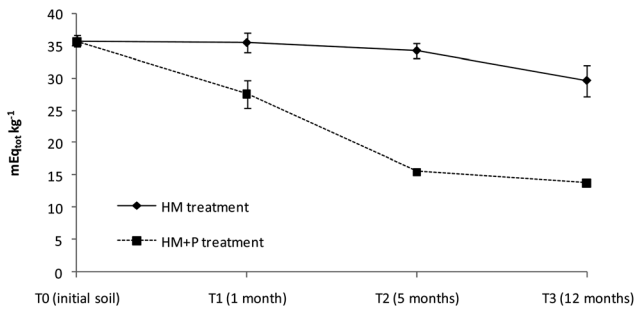


Fig. 2 Metals (Cr, Cu, Ni, Zn, Cd and Pb) in mEq$_{tot}$ kg^{-1} (the mg kg^{-1} of each metal was transformed to mEq kg^{-1} and then summed) in the initial soil at *T*0 and horse manure (HM) and horse manure and plant (*Populus nigra*) (HM + P) treatments at *T*1 (one month), *T*2 (five months) and *T*3 (12 months) sampling times.

treatment, as expected, generally did not show significant differences in the total metal concentration during the time (Table 3, Fig. 2).

Total petroleum hydrocarbons (TPHs) showed a significant decrease in the HM + P treatment over time, while in the HM treatment they were significantly lower than those in the initial soil only at the end of the experimentation (*T*3) (Fig. 3).

Regarding the polychlorobiphenyls (PCBs) a decrease of about 60% was observed between *T*1 and *T*3 sampling time in the HM + P treatment, while no significant differences were found between the HM treatment and initial soil (Table 3).

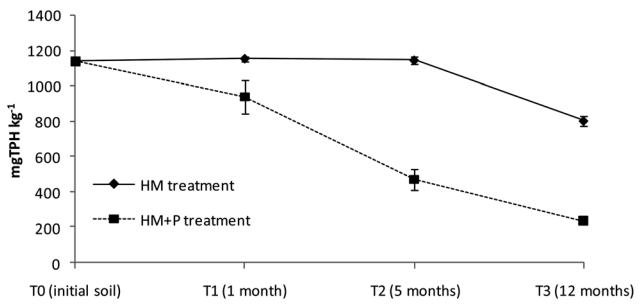


Fig. 3 Total petroleum hydrocarbon (TPH) in the initial soil at *T*0 and horse manure (HM) and horse manure and plant (*Populus nigra*) (HM + P) treatments at *T*1 (one month), *T*2 (five months) and *T*3 (12 months) sampling times.

3.4 Protein expression pattern

Protein electrophoresis (SDS-PAGE) has permitted the identification of different protein expression patterns in treated soils at *T*3 with respect to the initial soil (Fig. 4). The gel electrophoresis of the three extractable protein replicates showed the same expression patterns, confirming the repeatability of the extraction and purification procedure; in view of this just one replicate for each sample is reported in Fig. 4. From the SDS-PAGE it is possible to identify differential bands present in the treated soils at the end of the experimentation (*T*3) but not in the initial soil (*T*0). It appears clear that several bands are more intense in the

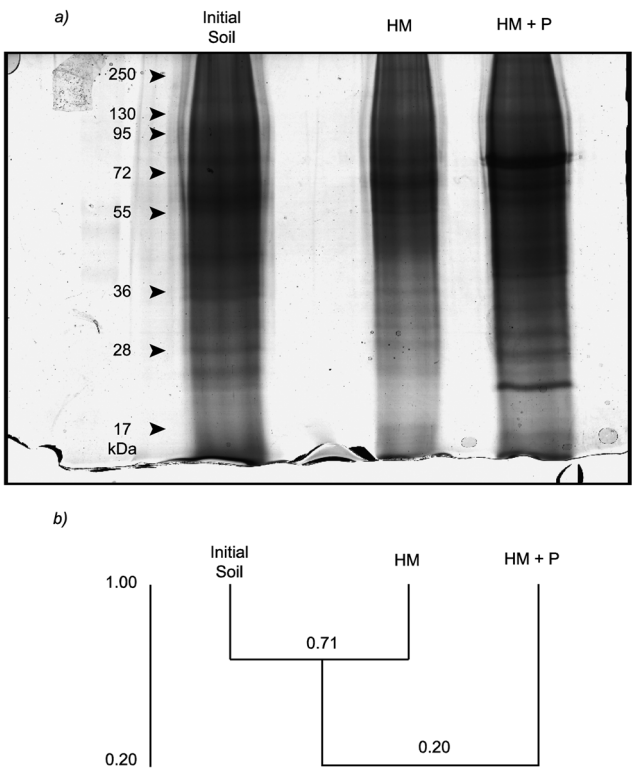


Fig. 4 (a) SDS-PAGE expression patterns of proteins detected in the initial soil at *T*0 (line 1), HM treatment at *T*3 (line 2) and HM + P treatment at *T*3 (line 3). (b) Cluster analysis of protein profiles performed among the initial soil (*T*0), and HM and HM + P treatments at *T*3.

planted soil (HM + P) with respect to the unplanted one (HM). A UPGMA dendrogram was performed in order to cluster together the treated and initial soils (Fig. 4). It is interesting to notice that the dice coefficient, used as a similarity measure, showed a value of 0.71 between the initial soil and HM treatment and just 0.20 between the initial soil and HM + P treatment.

4. Discussion

The initial soil, collected only at the beginning of the experimentation, was a degraded soil, with a low quality level and ability to support vegetation growth (scant coverage). As reported by other authors, the scant vegetation makes the soil susceptible to degradation.[34] In view of this, even though the chemical and biochemical evolution of this soil has not been monitored over time, only few positive changes in biochemical parameters could be expected in relation to the variation of environmental conditions.[35]

The horse manure one month after the application (*T*1 sampling time) increased the total organic C and N content of the soil and stimulated soil microbial activity. Beneficial effects of applying organic materials on soil chemical, physical and biological properties are reported in several experiments and they were attributed to a combination of factors, such as amelioration of soil porosity and water holding capacity, decrease in contaminants availability, increase in available organic carbon and other nutrients, and in microbial biomass from the organic amendments.[36–38]

The decrease over time in total organic C and N in the HM treatment is probably due to mineralization processes during horse manure decomposition, mainly composed of readily available substrates, while the decrease in N concentration observed in the HM + P treatment could be related to the N absorption by plants. In this treatment, the increase in C content suggested the contribution of plants to organic matter incorporation in soil. Plants can contribute in many ways to ameliorate the chemical characteristics of soil; plant roots, through the enhancement of soil aeration, can stimulate microbial activity, in particular nitrificants, causing an increase in the soil nitrate concentration.

Moreover, plants may also contribute to labile organic C input through root exudates and plant remains.[39] In fact, the water soluble C, which is a labile C fraction, was higher in the HM + P treatment at both *T*2 and *T*3 sampling time with respect to HM treatment. Therefore, this C fraction, easily degradable, acting as a substrate for soil microorganisms,[40] may stimulate the soil microbial functionality expressed by enzyme activities.

Dehydrogenase enzymes, together with other hydrolase enzymes, act on the basic processes of organic matter decomposition and have been used as biomarkers of the reactions which occur in contaminated soil.[41,42]

Dehydrogenase activity, which involves the biological oxidation of organic compounds, has been often used to evaluate soil microbial growth and activity.[28] As concerns the hydrolase enzymes, phosphatases are considered key enzymes in the phosphorus cycle in the soil.[43] Variations in this enzymatic activity, apart from indicating changes in the quantity and quality of soil phosphorated substrates, have been also used as good indicators of the soil biological status.[44,45] Moreover, β-glucosidase, an enzyme involved in the carbon cycle, is responsible for the last phase of microbial degradation of cellulose to glucose.[46] This enzyme hydrolyzes the non-reducing terminal groups of β-D-cellobiose releasing β-glucose, a very important energy source for microorganisms. Therefore, the β-glucosidase reflects the soil organic matter state and its activity is strictly bound to the presence of carbon compounds.

In our study, dehydrogenase activity increased over time in the HM + P treatment, confirming the contribution of organic substrates to the maintenance of the metabolic activity of soil microorganisms. Beneficial effects of plants on microbial stimulation through organic exudates at the root–soil interface were previously reported.[47,48] Dehydrogenase activity in the HM treatment decreased during the experimental time probably because the substrates were no longer available to sustain microbial biomass growth and activity.[44] In addition, as previously found in other studies,[49–51] dehydrogenase activity may also be inhibited by metals contamination. In view of this, the maintenance of high quantities of metals in the HM treatment could also be the reason for the lower dehydrogenase activity.

Shen *et al.*[52] investigated the interactions of polycyclic aromatic hydrocarbons (PAHs) and heavy metals (Cd, Zn and Pb) with soil enzymes (urease and dehydrogenase). These authors found that dehydrogenase was more sensitive to the combined inorganic and organic pollution than urease enzyme. Similarly, Maliszewska-Kordybach and Smreczak[51] demonstrated that dehydrogenase activity is most sensitive to the combined effects of pollutants (metals and PAH). Shen *et al.*[52] reported that urease and dehydrogenase could be suitable indicators of both inorganic and organic contamination (metals and PAH), particularly at the early stages of pollution.[53]

In addition, different metals affect soil enzymes in different ways. Renella *et al.*[54] found that Cd inhibited alkaline phosphatase but did not affect acid phosphatase and β-glucosidase. In contrast, Balyaeva *et al.*[55] found that Pb decreased the activity of acid phosphatase significantly. Wyszkowska *et al.*[56] found that the metal sensitivities of enzymes followed the order: dehydrogenase > urease > alkaline phosphatase > acid phosphatase.

Several studies on phosphatase and β-glucosidase activities demonstrated the decrease in these enzyme activities with increasing metal concentrations,[57,58] while the addition of fresh organic matter to soil decreased the inhibitory effect of metals, indicating reduced toxicity.[59,60] In the work by Tejada *et al.*[61] a positive role of organic substances in neutralizing the adverse effect of Ni on soil enzymatic activities was reported.

Our study at real scale level, in agreement with the above mentioned laboratory experimentation, showed an increase of β-glucosidase and acid phosphatase activities over time in both treatments (HM and HM + P) and with respect to the initial soil. It is well known that fertilizer application to a bare soil results in an increase in enzyme activities.[62]

These biochemical parameters, showing higher values in the HM + P treatment than HM treatment, were surely positively influenced by the presence of plants. Planted soils have frequently been reported to have higher rates of microbial activity than unplanted soil due to the presence of additional surfaces for microbial colonization and organic compounds released by the plant roots as also suggested by Delorme *et al.*[63] These authors reported also that the plants are able to increase soil enzyme activities also under metal stress.

Some *in situ* experimentations on chelant-enhanced phytoextraction of Pb and Cd from soil proved the efficiency of poplars in heavy metal decontamination.[64,65]

In this study the decrease in metal concentration in the HM + P treatment confirmed the effectiveness of the *Populus nigra* treatment in soil detoxification.

Wang and Jia[66] reported a study on adsorption and remediation of heavy metals by poplar and larch in contaminated soil. In their study, metal concentrations in tissues of plants grown in Cd, Cu, and Zn treated soil were higher than controls. Moreover, the concentrations in poplar were higher than in larch, confirming the suitability of the poplar for phytoremediation in heavy metal-polluted soil.

Poplars are also known to take up and detoxify several organic pollutants, such as chlorinated hydrocarbon solvents, herbicide atrazine, and different explosives.[9]

In our study, the soil very likely had an indigenous microbial total petroleum hydrocarbon degradative capacity. This was enhanced and stimulated by the addition of the horse manure and *Populus nigra* plantation. It might be hypothesized that the indigenous bacteria of the soil revived their activity in the presence of a new carbon source supplied with the horse manure. In the HM + P treatment the particularly high reduction in TPH can be related to the root-associated microbial degradation activity.

The first reports on the potential of plants in bioremediation of PCBs were published in the late 1970s and early 1980s.[67,68] To our knowledge, few results about field scale experiments with poplar phytoremediation are available. Kacalkova and Tlusto[69] reported results about the effectiveness of poplar in uptaking PCBs from soil with a level of contamination of 1.530 mg kg^{-1}. Moreover, Donnelly *et al.*[70] suggested that some compounds of plant origin, such as coumarins and flavonoids, could be used as co-substrates to the PCB-degrading aerobic bacteria, thus making the biodegradation of PCBs with the presence of plants more effective. As concerns the effect of organic matter addition to PCBs contaminated soil, a strong adsorption onto the organic substance was observed and attributed to their physico-chemical properties, thus making them less bioavailable and more recalcitrant to biodegradation.[71,72] In agreement with this finding, in our study, a degradation of this recalcitrant portion of organic contaminants was observed only in the presence of plants.

SDS-PAGE permitted the comparison of the protein expression patterns among the initial soil and treated soils and between treatments. The differences in the protein expression between the HM + P treatment and the initial soil (dice coefficient 0.20) can be explained by the increase in microbial activity due to plant-roots associated microorganisms. Proteins in soil originate from plants, animals and microorganisms inhabiting this ecosystem either through active excretion or passive release. Plant roots continuously secrete compounds in the rhizosphere such as ions, metabolites and proteins. Godlewski and Adamczyk[73] reported the secretion of proteases by roots of various plant species. Microbial community composition and structure are different across a broad range of soil, habitat, vegetation, and management conditions.

In the planted soil, the increase in the enzyme activities could consequently have modified the banding profile, with a higher intensity of some bands in the HM + P treatment with respect to those found in the initial soil. It is well known that between plants and microorganisms a deep molecular signaling is present, but to date the difficulties in extraction and especially identification of soil proteins prevent the possibility of analysing them and using the obtained information to apply a more sustainable soil management during the soil decontamination.

5. Conclusions

The decrease during the time of both inorganic (metals) and organic (TPH and PCBs) contaminants in the HM + P treatment indicated the effectiveness of the phytoremediation system.

In this treatment, the increase of biological parameters (dehydrogenase, β-glucosidase and phosphatase activities) over time indicated the activation of microbial metabolism favored by the organic matter application and the plant root–microorganism interaction.

Moreover, the stimulation of hydrolase enzymes also in the horse manure (HM) treatment indicated the positive role of organic substances in neutralizing the adverse effect of metals on soil enzymatic activities.

Metaproteomic analysis permitted the purification of a greater number of proteins (SDS-PAGE bands) in planted (HM + P) with respect to unplanted (HM treatment) and initial soil, thus indicating the improvement of the HM + P soil functional status; this methodology may become a basic tool for the detection of the biogeochemical evolution of soil during decontamination. In addition, the combination of mass spectrometry analysis and environmental protein database enrichment could, in the future, permit the identification of microbial populations that are active in soil decontamination.

Acknowledgements

The authors gratefully acknowledge the technical assistance of Fernando Di Giovanni and Manuele Scatena for the samples collection and preparation.

References

1 S. Susarla, V. F. Medina and S. C. McCutcheon, *Ecol. Eng.*, 2002, **18**, 647–658.
2 A. Perez de Mora, J. J. Ortega-Calvo, F. Cabrera and E. Madejon, *Appl. Soil Ecol.*, 2005, **28**, 125–137.
3 H. I. Atagana, *Water, Air, Soil Pollut.*, 2011, **215**, 261–271.
4 Y. Sun, Q. Zhou, Y. Xu, L. Wang and X. Liang, *J. Environ. Chem. Ecotoxicol.*, 2011, **3**, 45–51.
5 G. J. Thoma, T. B. Lam and D. C. Wolf, *Int. J. Phytorem.*, 2003, **5**, 125–136.
6 P. M. White, D. C. Wolf, G. J. Thoma and C. M. Reynolds, *Water, Air, Soil Pollut.*, 2006, **169**, 207–220.
7 E. W. Aitchison, S. L. Kelley, P. J. J. Alvarez and J. L. Schnoor, *Water Environ. Res.*, 2000, **72**, 313–321.
8 J. Liu and J. L. Schnoor, *Chemosphere*, 2008, **73**, 1608–1616.
9 T. Komives and G. Gullner, in *Plant Environment Interactions*, ed. R. E. Wilkinson, Marcel Dekker, New York, NY, USA, 2nd edn, 2000, pp. 437–452.
10 J. G. Burken and J. L. Schnoor, *Environ. Sci. Technol.*, 1998, **32**, 3379–3385.
11 I. Laureysens, J. Bogaert, R. Blust and R. Ceulemans, *For. Ecol. Manage.*, 2004, **187**, 295–309.
12 A. Gelsomino, L. Badalucco, R. Ambrosoli, C. Crecchio, E. Puglisi and S. Meli, *Soil Biol. Biochem.*, 2006, **38**, 2069–2080.
13 P. Eldor, in *Tercera*, ed. P. Eldor, Academic Press, Chennai, India, 2007.

14 M. Nielsen and A. Winding, *Microorganisms as Indicators of Soil Health*, National Environmental Research Institute, Denmark, 2001.
15 R. Dick, in *Methods for Assessing Soil Quality*, ed. J. Doran and A. Jones, Soil Science Society of America, Madison, Wisconsin, 1996, pp. 121–156.
16 E. Kandeler and R. Dick, in *Biodiversity in Agricultural Production Systems*, ed. G. Benckiser and S. Schnell, Taylor & Francis Group, Boca Raton, FL, 2007, pp. 263–279.
17 P. Nannipieri, in *Nucleic Acids and Proteins in Soil*, ed. P. Nannipieri and K. Smalla, Springer-Verlag, Heidelberg, 2006, pp. 75–94.
18 L. Giagnoni, F. Magherini, L. Landi, S. Taghavi, A. Modesti, L. Bini, P. Nannipieri, D. Van derlelie and G. Renella, *Eur. J. Soil Sci.*, 2011, **62**, 74–81.
19 S. J. Indorante, L. R. Follmer, R. D. Hammer and P. G. Koenig, *Soil Sci. Soc. Am. J.*, 1990, **54**, 560–563.
20 C. Garcia, M. T. Hernandez and F. Costa, *Soil Sci. Plant Nutr.*, 1990, **37**, 399–408.
21 J. C. Yeomans and J. M. Bremner, *Commun. Soil Sci. Plant Anal.*, 1988, **19**, 1467–1476.
22 J. Murphy and J. P. Riley, *Anal. Chim. Acta*, 1962, **27**, 31–36.
23 B. Ceccanti, C. Garcia, G. Masciandaro, C. Macci and S. Doni, *Water, Air, Soil Pollut.*, 2006, **177**, 383–397.
24 EPA, *Method 3545, in Pressurized Fluid Extraction, Test Methods for Evaluating Solid Waste*, Update III, EPA SW-846, US GPO, Washington, DC, 3rd edn, 1995.
25 EPA, *Method 3640A, in Gel Permeation Cleanup*, EPA SW-846, US GPO, Washington, DC, 1994.
26 S. Sinkkonen and J. Paasivirta, *Chemosphere*, 2000, **40**, 943–949.
27 G. Masciandaro and B. Ceccanti, *Soil Tillage Res.*, 1999, **51**, 129–137.
28 G. Masciandaro, B. Ceccanti, V. Ronchi and C. Bauer, *Biol. Fertil. Soils*, 2000, **32**, 479–483.
29 G. Masciandaro, C. Macci, S. Doni, B. Maserti, L. A. Calvo-Bado, B. Ceccanti and E. Wellington, *Soil Biol. Biochem.*, 2008, **40**, 2156–2161.
30 B. Ceccanti, M. Bonmati-Pont and P. Nannipieri, *Biol. Fertil. Soils*, 1989, **7**, 202–206.
31 HUJI, in *The Protein Purification Facility, Protein Precipitation Protocols*, The Wolfson Centre for Applied Structural Biology, The Hebrew University of Jerusalem, 2006, http://wolfson.huji.ac.il.
32 U. K. Laemmli, *Nature*, 1970, **227**, 680–685.
33 M. J. Latorre, R. Peña, C. Pita, A. Botana, S. García and C. Herrero, *Food Chem.*, 1999, **66**, 263–268.
34 F. Bastida, J. L. Moreno, T. Hernández and C. García, *Sci. Total Environ.*, 2007, **378**, 183–186.
35 M. Ros, M. T. Hernandez and C. Garcia, *Soil Biol. Biochem.*, 2003, **35**, 463–469.
36 A. Perez de Mora, E. Madejón, P. Burgos and F. Cabrera, *Sci. Total Environ.*, 2006, **363**, 28–37.
37 S. Marinari, G. Masciandaro, B. Ceccanti and S. Grego, *Bioresour. Technol.*, 2000, **72**, 9–17.
38 A. K. Bandick and R. P. Dick, *Soil Biol. Biochem.*, 1999, **3**, 1471–1479.
39 C. Garcia, T. Hernandez and A. Roldan, *J. Environ. Qual.*, 1997, **26**, 285–291.
40 B. D. Cook and D. L. Allan, *Soil Biol. Biochem.*, 1992, **24**, 585–594.
41 H. Bentham, J. A. Harris, P. Birch and K. C. Short, *Biol. Fertil. Soils*, 1992, **29**, 711–718.
42 W. A. Dick and M. A. Tabatabai, in *Soil Microbial Ecology: Application in Agricultural and Environmental Management*, ed. F. Blaine, Marcel Dekker, New York, 1993, pp. 95–127.
43 W. A. Dick and M. A. Tabatabai, *Soil Sci.*, 1984, **137**, 7–15.
44 J. A. Pascual, T. Hernandez, C. Garcia and M. Ayuso, *Bioresour. Technol.*, 1998, **64**, 131–138.
45 J. A. Pascual, J. L. Moreno, T. Hernandez and C. Garcia, *Bioresour. Technol.*, 2002, **82**, 73–78.
46 K. Alef and P. Nannipieri, in *Methods in Applied Soil Microbiology and Biochemistry*, Academic Press, London, 1995.
47 E. I. Newman, in *Ecological Interactions in Soil: Plants, Microbes and Animals*, ed. A. H. Fitter D A, D. J. Read and M. B. Usher, Blackwell Scientific Publications, Oxford, UK, 1985, pp. 107–121.
48 C. Macci, S. Doni, E. Peruzzi, G. Masciandaro, C. Mennone and B. Ceccanti, *J. Environ. Manage.*, 2012, **95**, S215–S222.
49 O. Mikanova, *J. Geochem. Explor.*, 2006, **88**, 220–223.
50 C. R. Zheng, C. Tu and H. M. Chen, *Pedosphere*, 1999, **9**, 251–258.
51 B. Maliszewska-Kordybach and B. Smreczak, *Environ. Int.*, 2003, **28**, 719–728.
52 G. Shen, Y. Lu, Q. Zhou and J. Hang, *Chemosphere*, 2005, **61**, 1175–1182.
53 E. Baath, *Water, Air, Soil Pollut.*, 1989, **47**, 335–379.
54 G. Renella, M. Mench, L. Landi and P. Nannipieri, *Soil Biol. Biochem.*, 2005, **37**, 133–139.
55 O. N. Balyaeva, R. J. Haynes and O. A. Birukova, *Biol. Fertil. Soils*, 2005, **41**, 85–94.
56 J. Wyszkowska, J. Kurcharski and W. Lajszner, *Pol. J. Environ. Stud.*, 2006, **15**, 927–934.
57 A. Karaca, D. C. Naseby and J. M. Lynch, *Biol. Fertil. Soils*, 2002, **35**, 428–434.
58 R. Kizilkaya, *Ecol. Eng.*, 2004, **22**, 141–151.
59 D. Chaudhuri, S. Tripathy, H. Veeresh, M. A. Powell and B. R. Hart, *Environ. Geol.*, 2003, **45**, 115–123.
60 J. L. Moreno, C. García and T. Hernández, *Eur. J. Soil Sci.*, 2003, **54**, 377–386.
61 M. Tejada, J. L. Moreno, M. T. Hernandez and C. Garcia, *Eur. J. Soil Biol.*, 2008, **44**, 129–140.
62 A. de Varennes, C. Cunha-Queda and G. Qu, *Water, Air, Soil Pollut.*, 2010, **208**, 91–100.
63 T. A. Delorme, J. V. Gagliardi, J. S. Angle and R. L. Chaney, *Can. J. Microbiol.*, 2001, **47**, 773–776.
64 M. Komarek, P. Tlustos, J. Szakova and V. Chrastny, *Environ. Pollut.*, 2008, **151**, 27–38.
65 B. H. Robinson, T. M. Mills, D. Petit, L. E. Fung., S. R. Green and B. E. Clothier, *Plant Soil*, 2000, **227**, 301–306.
66 X. Wang and Y. Jia, *Environ. Sci. Pollut. Res.*, 2010, **17**, 1331–1338.
67 J. B. Weber and E. Mrozek, *Bull. Environ. Contam. Toxicol.*, 1979, **23**, 412–417.
68 T. R. Schwartz and R. G. Lehmann, *Bull. Environ. Contam. Toxicol.*, 1982, **28**, 723–727.
69 L. Kacalkova and P. Tlusto, *Cent. Eur. J. Biol.*, 2011, **6**, 223–235.
70 P. K. Donnelly, J. S. Fletcher and R. S. Hedge, *Chemosphere*, 1994, **28**, 981–988.
71 C. T. Chiou, in *Reactions and Movement of Organic Chemicals in Soil*, ed. B. L. Sawhney and K. Brown, Special Publication 22, SSSA and ASA, Madison, WI, 1989, pp. 1–29.
72 T. Chekol, *Biorem. J.*, 2005, **16**, 117–139.
73 M. Godlewski and B. Adamczyk, *J. Plant Biol.*, 2007, **45**, 657–664.

