SAPIENZA
UNIVERSITÀ DI ROMA

Dipartimento di Chimica e Tecnologie del Farmaco
XXIV Ciclo di Dottorato in Scienze Farmaceutiche

N-linked peptidoresorc[4]arene-based receptors designed for protein surface recognition.

NMR analysis of Italian virgin olive oils and structural analysis of polysaccharides.

Supervisor:
Prof. Bruno Botta

PhD Student:
Antonella Cerreto

Academic year 2011-2012

Alla mia famiglia

Contents

Overview

During the PhD course, the research activity focused on three main topics.

N-linked peptideresorcarenes. [Part I]
An important class of protein surface receptors based on the attachment of four cyclic peptides to a calix[4]arene scaffold was developed by Hamilton and targeted to the serine protease α-chymotrypsin (ChT). Moving from Hamilton's results, we performed preliminary molecular modelling studies suggesting the design of tetra-anionically functionalized receptors that might target the predominantly cationic region of ChT surrounding the active site. We have previously designed valyl-leucine and leucyl-valine peptidoresorc[4]arenes, inspired by the evidence that N- or C-linked peptides endow calixarenes with novel properties, such as recognition of protein surfaces and carbohydrates, formation of stable inclusion complexes, and self-assembly phenomena. By varying sequence, nature, and stereochemistry of the chains we prepared anionically functionalized N-linked peptidoresorc[4]arenes by hydrogenation of their precursor benzyl esters. From this family of receptors we have identified noncompetitive inhibitors of ChT, which function by binding to the surface of the enzyme in the neighbourhood of the active site cleft (K_i values ranging from 13 to 0.8 μM).
All the obtained results have been published in an article in JOC (see hereinafter).

NMR characterization of Italian virgin olive oils. [Part II]
Following the experimental protocol established in the Annalaura Segre NMR Laboratory (CMI, CNR in Montelibretti, Rome), based on ^{1}H-NMR and statistical multivariate analysis of the data (ANOVA, PCA and LDA), nearly 300 Italian virgin olive oil samples, produced in two harvesting years,

2009/10 and 2010/11, by certified chains from Lombardy, Tuscany, Sardinia, Puglia, Calabria, Sicily, Lazio, Molise and Liguria, have been studied.
The NMR results confirm that the chemical composition of olive oils, with regards to minor components, is influenced by geographical, ecological and genetic factors and depends also on the harvesting year.
We could also identify some oil components that remain stable in different harvesting years, that may represent seasonal-independent characteristics of the oil produced in each region.

NMR structural analysis of polysaccharides. [Part III]
A contemporary investigation, carried out at CNR, was focused on structural analysis of polysaccharides. Specificaly, we performed NMR characterization of a carboxymethyl derivative of scleroglucan, a polysaccharide produced by fungi of the *Sclerotium* genus, to identify the positions involved in the carboxymethylation; furthermore, a preliminary structural study has been performed on a set of polysaccharides obtained from the cultures of *Hericium erinaceus,* with the aim of identifying the configuration and the connections between the monosaccharide units.

The results of the research were object of the following communications:

FULL PAPERS

- D'Acquarica I., Cerreto A., Delle Monache G., Subrizi F., Boffi A., Tafi A., Forli S., and Botta B. *J. Org. Chem.*, **2011**, *76* (11), pp 4396–4407 *"N*-Linked Peptidoresorc[4]arene-Based Receptors as Noncompetitive Inhibitors for α-Chymotrypsin"

ORAL COMUNICATIONS

- **A. Cerreto**, "Peptidoresorcin[4]areni: synthesis and surface interaction studies", Convegno unità operative PROGETTO FIRB PIATTAFORME-RETI e PRIN "ANGIOCHEMISTRY", Certosa di Pontignano (Si), 22-23/01/2009;
- **A. Cerreto**, F. Subrizi, B. Botta, I. D'Acquarica: "Sintesi di recettori artificiali per la ricognizione superficiale di proteine", 10° S.A.Y.C.S., Pesaro 18-20/10/2010;
- **A. Cerreto**, F. Subrizi, B. Botta, I. D'Acquarica "*N-linked Peptidoresorc[4]arene-based Receptors for Protein Surface Recognition"*, 11th International Conference on Calixarenes (Calix 11), Tarragona (Spagna), 26-29/06/2011 (Short Oral Presentation)

- POSTER COMUNICATIONS
  - **A. Cerreto**, F. Subrizzi, B. Botta, I. D'Acquarica: "Designed resorc[4]arene-based receptors for protein surface recognition" XXXIV Corso Estivo in Sintesi Organica "Attilio Corbella", Gargnano (Bs) 22-26/6/2009
  - **A. Cerreto**, B. Botta, L. Mannina, R. Petroccione, A. Sobolev, D. Capitani: "Metodologia NMR nello studio degli alimenti: il caso dell'olio d'oliva", II Workshop "Applicazioni della Risonanza Magnetica nella Scienza degli Alimenti" 27-28/05/2010;
  - R. Petroccione, L. Mannina, **A. Cerreto**, M. Gobbino, A. P. Sobolev, D. Capitano, A. Scaramozzino, R. Filo della Torre: "Gli oli di oliva italiani "visti" dalla risonanza magnetica nucleare", II Workshop "Applicazioni della Risonanza Magnetica nella Scienza degli Alimenti" 27-28/05/2010;
  - **A. Cerreto**, B. Botta, L. Mannina, R. Petroccione, M. Gobbino, A. P. Sobolev, D. Capitani, A. Scaramozzino, R. Filo della Torre: "The

Italian olive oils as “seen” by the Nuclear Magnetic Resonance”, ESF-COST High-Level Research Conference “Natural Products Chemistry, Biology and Medicine III”, Acquafredda di Maratea (PO) 5-10/09/2010;

- **A. Cerreto**, F. Subrizi, B. Botta, I. D’Acquarica “*N-linked Peptidoresorc[4]arene-based Receptors for Protein Surface Recognition”*, 11th International Conference on Calixarenes (Calix 11), Tarragona (Spagna), 26-29/06/2011.

Part I

N-linked peptidoresorc[4]arenes receptors designed for protein surface recognition

The content of this chapter is published in the following paper: D'Acquarica I., Cerreto A., Delle Monache G., Subrizi F., Boffi A., Tafi A., Forli S., and Botta B. *J. Org. Chem.*, **2011**, *76* (11), pp 4396–4407 *"N*-Linked Peptidoresorc[4]arene-Based Receptors as Noncompetitive Inhibitors for α-Chymotrypsin"

1. Introduction

1.1 Supramolecular Chemistry

Jean-Marie Lehn, who won the Nobel Prize for his work in the area in 1987, defined supramolecular chemistry as the "chemistry of molecular assemblies and of intermolecular bonds".[1]
This kind of "chemistry beyond the molecule", considered in its simplest sense, regards the (non covalent) binding of a molecule (host) with another molecule (guest), producing a new entity, defined complex or supramolecule, with properties different from those of each single component.
Commonly the host is defined as the molecular entity possessing *convergent* binding sites (*e.g.* Lewis basic donor atoms, hydrogen bond donors *etc.*), while the guest is the entity possessing *divergent* binding sites (*e.g.* Lewis acidic metal cation or hydrogen bond acceptor halide anion). In turn a binding site is defined as a region of the host or guest capable of taking part in a non-covalent interaction.[2]
Generally the host appears as the largest component, such as an enzyme or synthetic cyclic compounds with a central hole or cavity, while the guest correspond to a small organic or inorganic ion, or ion pair, or a little molecule with biological activity such as hormone or neurotransmitter.
Moreover the host-guest aggregate could be identified using different term, with respect to the forces occurring between the components:[3]

- if electrostatic interactions are the principal forces (including ion–dipole, dipole–dipole, hydrogen bonding *etc.*) the term *complex* is used;
- if less specific, weaker, non-directional interaction are involved (i.e. hydrophobic, van der Waals or crystal close-packing effects) the most appropriate terms are *cavitate* and *clathrate*: the former is constituted by the aggregation of guest and cavitand, host possessing permanent intramolecular cavities available for guest binding as intrinsic molecular property, both in solutions and in solid state; the latter is obtained by the aggregation of guest and

clathrands, hosts possessing extramolecular recess that constitute a cavity only after aggregation in solid or crystalline state. (Figure **1.1**). The distinctions between these classes of supramolecular aggregates are blurred and often the word "complex" is used to cover all these phenomena.

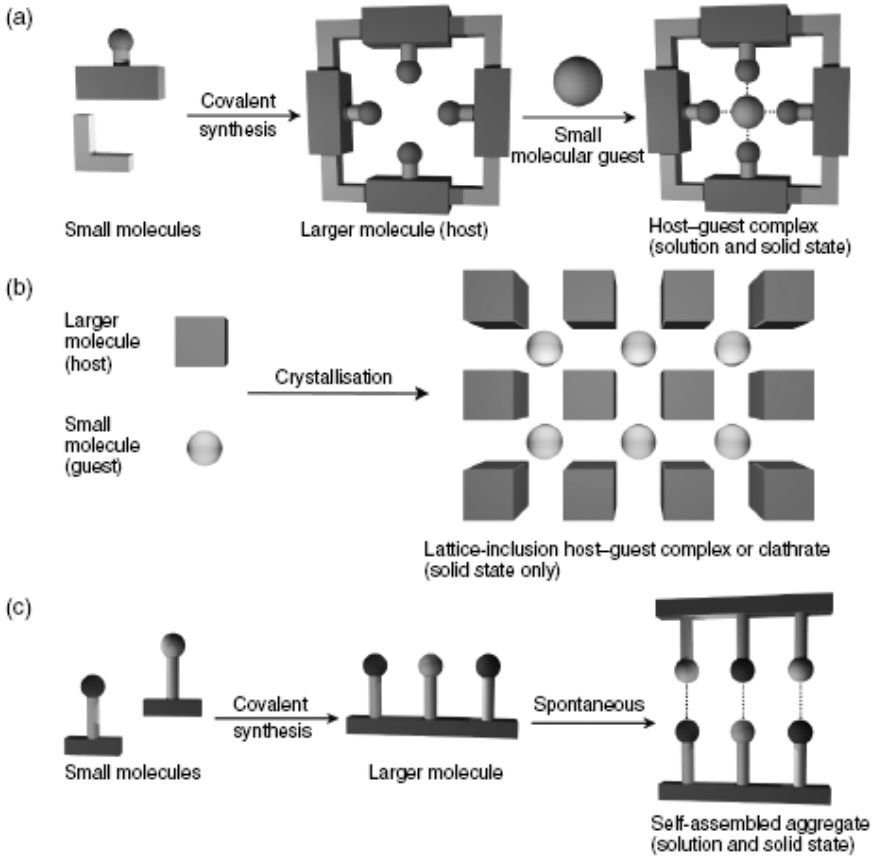


Figure 1.1: Schema illustrating the difference between a cavitate and a clathrate: (a) synthesis and conversion of a cavitand into a cavitate by inclusion of a guest into the cavity of the host molecule; (b) inclusion of guest molecules in cavities formed between the host molecules in the lattice resulting in conversion of a clathrand into a clathrate; (c) synthesis and self-assembly of a supramolecular aggregate that does not correspond to the classical host-guest description.

The host-guest chemistry is based on three fundamental concepts: [4]

1. the recognition by Paul Ehrlich, at the beginning of 20th century, that molecules do not act if they do not bind, "*Corpora non agunt nisi fixate*"; in this way Erlich introduced the concept of a biological receptor;.

2. the recognition in 1894 by Emil Fischer that binding must be selective, as part of the study of receptor–substrate binding by enzymes. He described this by a *lock and key* image of steric fit in which the guest has a geometric size or shape complementarity to the receptor or host. This concept laid the basis for *molecular recognition*, the discrimination by a host between a number of different guests;
3. the fact that selective binding must involve *attraction* or mutual affinity between host and guest. This is, in effect, a generalization of Alfred Werner's 1893 theory of coordination chemistry, in which metal ions are coordinated by a regular polyhedron of ligands binding by dative bonds.

1.2 Molecular Recognition and Host- Guest Chemistry

Associations between host (H) and guest (G) molecules are usually based on simultaneous non-covalent interactions between single binding sites, A (acceptor) and D (donor).[5] Exceptions are solvent-driven equilibria and enforced guest encapsulations within closed host cavities. The need for several binding sites is quite evident: non-covalent interactions are usually weak, and concerted interplay between many sites is the only way to achieve strong and specific complexation (recognition) of a guest molecule. The principle of multi-site complexation is very general in living systems, where it ensures the efficiency of many important biological functions, including DNA replication, enzyme-substrate recognition, and antigen-antibody interactions.
Furthermore, one can view multi-site complexation as a generalized chelate effect, which is of course well known from coordination chemistry. An important requirement for multi-site binding is complementarity between binding sites of host and guest molecules. In other words, complexation will be more efficient when the shapes and arrangements of binding sites in host and guest molecules fit each other. This is the general lock and key principle of Emil Fischer, who had already explained the remarkable specificity of enzyme catalysis a century ago.
In order to bind, a host must have binding sites that present the correct electronic character (polarity, hydrogen bond donor/acceptor ability, hardness or softness *etc.*) to complement those of the guest. Hydrogen bond donors must match acceptors, Lewis acids must match Lewis bases and so on. Furthermore, those binding sites must be spaced out on the host

in such a way as to make it possible for them to interact with the guest in the binding conformation of the host molecule. If a host fulfils these criteria, it is said to be *complementary*. The principle of complementarity has been summed up by Donald Cram:

”To complex, hosts must have binding sites which cooperatively contact and attract binding sites of guests without generating strong non-bonded repulsions”.[6]

An example of extremely efficient biological multi-site binding[7] of a low molecular weight guest (biotin) by a protein host (streptavidin) with an association constant K= 2.5 x 10^{13} M^{-1} is shown in figure **1.2** (**a**). A very large binding free energy of -76 kJ mol^{-1} results here from simultaneous action of more than 10 weak pairwise van der Waals, electrostatic, H-bonding and some lipophilic interactions. For comparison, a synthetic receptor for biotin, shown in figure **1.2** (**b**), uses only four hydrogen bonds for the guest recognition and therefore exhibits only K = 9.3 x 10^3 M^{-1} even in the less polar solvent $CDCl_3$, which favors hydrogen bonds.[8] In some cases, however, synthetic host compounds approach the affinity of biological receptors. For example, a protonated azacrown ether[9] (Figure **1.3**) can bind ATP (adenosine triphosphate) in water with an association constant K= 10^{11} M^{-1}; ethyl adenine can be bound by an artificial receptor in chloroform essentially by multiple hydrogen bonds with K= 4 x 10^5 M^{-1}.[10] Some cyclophanes[11] and calixarenes[12] bind choline and acetylcholine with association constants 10^4-10^5 M^{-1}; crown ethers and cryptands can bind alkali cations with stability constants and selectivities similar or even higher than natural ionophores.[13] Some synthetic siderophores have binding constants for Fe(III) ions around 10^{60} M^{-1}, generating severe problems to determine such values.

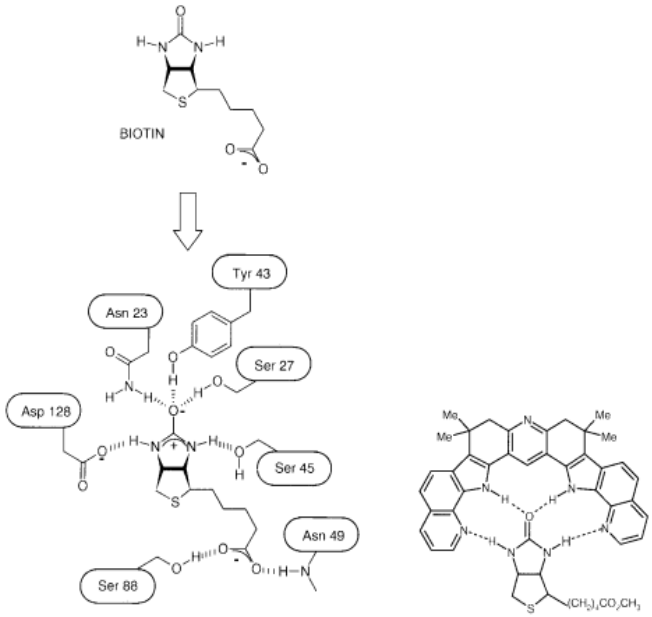


Figure 1.2: (a) Schematic illustration of biotin binding to streptavidin;[7] (b) Binding of biotin methyl ester to a synthetic host. [8]

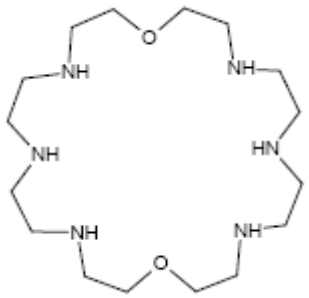


Figure 1.3. Azacrown ether able to complex ATP with an high stability constant.

Much of the emphasis in the construction of supramolecular host molecules concerns summative or even multiplicative interactions (Figure **1.4**). This means that we can construct a stable host–guest complex using non-covalent interactions (often weak), if we ensure that there are as many as possible of these interactions stabilizing the complex. The small amount of stabilization energy gained by each of these interactions when added to all the other small stabilizations from the other interactions (*summative*), results in a significant binding energy and hence complex stability.

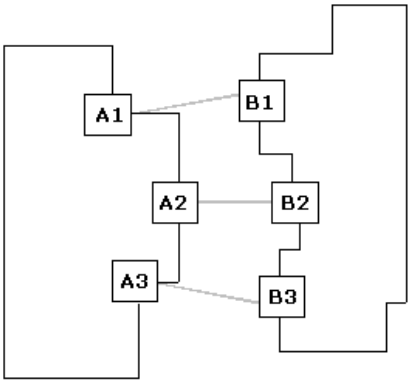


Figure 1.4: Schematic illustration of pairwise multisite host-guest interactions.

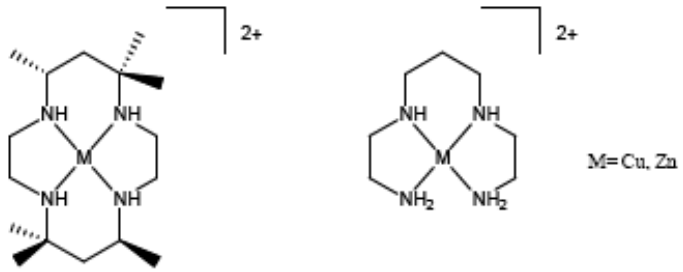
In some cases, the interaction of the whole system is synergically greater than the sum of the parts (multiplicative interaction). When two or more binding sites on a host cooperate in this way in binding a guest, the phenomenon is termed *cooperativity*. If the overall stability of the complex is greater than the sum of the energies of the interaction of the guest with binding groups A and B individually then the result is termed *positive cooperativity*. On the other hand, if unfavourable steric or electronic effects, arising from the linking of A and B together into one host, cause the overall binding free energy for the complex to be less than the sum of its parts, then the phenomenon is termed *negative cooperativity*. Binding site cooperativity in a supramolecular host-guest interaction is simply a generalisation of the *chelate effect* found in classical coordination chemistry. The chelate effect relates to the observation that metal complexes of bidentate ligands (such as 1,2-diaminoethane) are significantly more stable than closely related complexed that contain unidentate ligands (such as ammonia).

1.3 Preorganisation and Macrocyclic Effect

Many supramolecular host–guest complexes are even more stable than what would be expected from cooperative/chelate effects alone. The hosts in these species are usually macrocyclic (large ring) ligands that chelate their guests, again *via* a number of binding sites. Such compounds are additionally stabilised by what is traditionally termed the *macrocyclic effect*.[6] This effect

relates not only to the chelation of the guest by multiple binding sites, but also to the *organisation* of those binding sites in space prior to guest binding (*i.e. preorganisation*). Furthermore the enthalpic penalty associated with bringing donor atom and lone pairs into close proximity to one another (with consequent unfavourable repulsion and desolvation effects) has been "paid in advance" during the synthesis of the macrocycle. This makes macrocycles difficult to make but stronger complexing agents than analogous non-macrocyclic hosts (podands).

The macrocyclic effect was first elucidated by Cabbiness and Margerum in 1969 who studied the Cu(II) complexes **1.5.a** and **1.5.b**.[14] Both ions benefit from the stability associated with four chelating donor atoms. However, the macrocyclic complex **1.5.a** is about 10^4 times more stable than the acyclic analogue **1.5.b** as a consequence of the additional preorganisation of the macrocycle.



a) b)

Figure 1.5: Ion complexes **1.5.a** and **1.5.b** studied by Cabbiness and Margerum in 1969.

The stabilisation by macrocyclic preorganisation has both enthalpic and entropic contributions. The enthalpic term arises from the fact that macrocyclic hosts are frequently less strongly solvated than their acyclic analogues. This is because they simply present less solvent-accessible surface area. As a result there are fewer solvent–ligand bonds to break than in the extended acyclic case. Entropically, macrocycles are less conformationally flexible and so lose fewer degrees of freedom upon complexation.

In general, the relative importance of the entropic and enthalpic terms varies according to the system studied, although the enthalpy is frequently dominant. Bicyclic hosts such as *cryptands* are found to be even more stable than monocyclic *corands* for additional factors such as lone pairs repulsions. Historically this further additional stability is referred to as the *macrobicyclic effect* (Figure **1.6**) and simply represents the more rigid, preorganised nature of the macrobicycle.

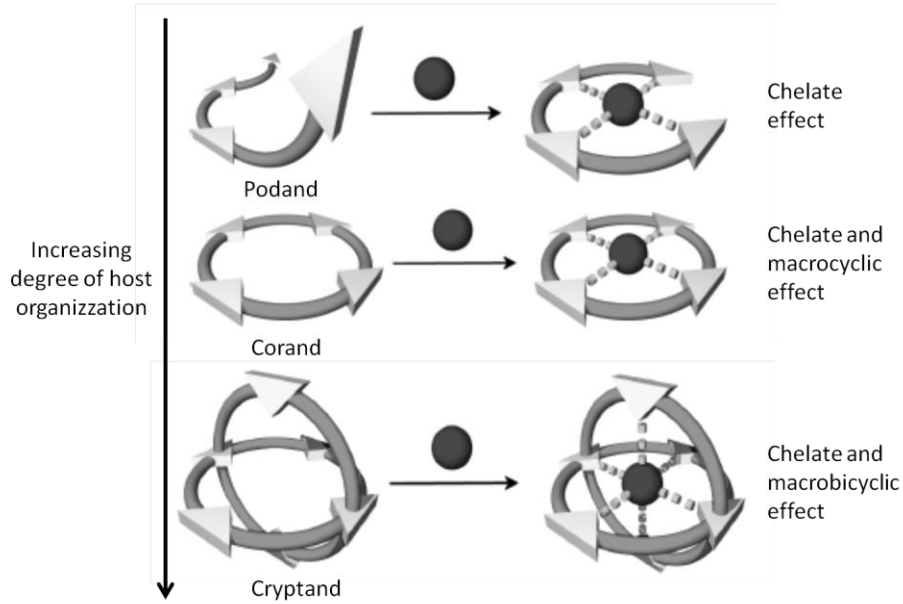
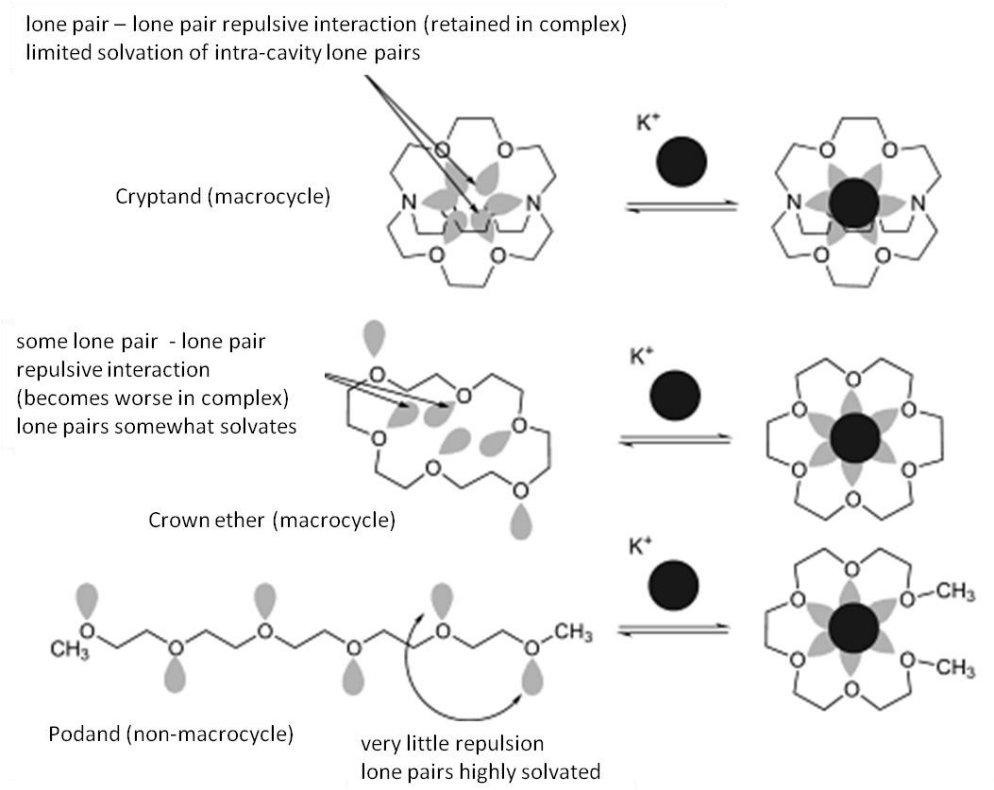


Figure 1.6: The chelate, macrocyclic and macrobicyclic effects.

The macrocyclic and macrobicyclic effects make an important contribution to hosts for alkali metal binding (Scheme **1.1**). The macrocyclic and macrobicyclic effects are simply manifestations of increasing *preorganisation*. We can say that if a host molecule does not undergo a significant conformational change upon guest binding, it is *preorganised*. Host preorganisation is a key concept because it represents a major (in some cases decisive) enhancement to the overall free energy of guest complexation.



Scheme 1.1: Comparison of preorganisation effects in K^+ binding by a macrobicycle, macrocycle and non-preorganised podand pentaethyleneglycol dimethyl ether.

Neglecting the effects of solvation, the host-guest binding process may be divided very loosely into two stages. First, there is an activation stage in which the host undergoes conformational readjustment; its binding sites must become complementary to the guest and at the same time unfavourable interactions between one binding site and another must be minimized. This is energetically unfavourable, and because the host must remain in this binding conformation throughout the lifetime of the host–guest complex, this energy is never paid back. In the second step, binding occurs; this is energetically favourable because of the enthalpically stabilising attraction between mutually complementary binding sites of host and guest.

The overall free energy of complexation represents the difference between the unfavourable reorganisation energy and the favourable binding energy. If the reorganisation energy is large, then the overall free energy is reduced,

destabilising the complex. If the host is preorganised, this rearrangement energy is small.
The corollary of preorganisation is in the guest binding kinetics. Rigidly preorganised hosts may have significant difficulty in passing through a complexation transition state and so tend to exhibit slow guest binding kinetics. Conformationally mobile hosts are able to adjust rapidly to changing conditions, and both complexation and decomplexation are rapid. Solvation enhances the effects of preorganisation since the solvation stabilisation of the unbound host is often greater than the case when it is wrapped around the guest, effectively presenting less surface area to the surrounding medium.
The preorganization[5] principle states that "the more highly hosts and guests are organized for binding and the lower is their solvation prior to their complexation, the more stable will be their complexes".[15]
To sum up: both molecules will be optimally preorganized if:
(a) all complementary binding sites geometrically match;
(b) in the complexed state they are in the same lowest free energy conformation as in the free state;
(c) polar binding sites need not to change solvation.
In this case all distortions will be negligible and the complexation will be energetically most favorable, including only the sum of intrinsic binding free energies. One also may expect the interactions of such preorganized molecules to be highly selective since guests of different structures, such as cations of different sizes, will need different optimal host conformations.

Generally it is not easy to create an universally valid rule to generate a complex, independent from the nature of host and guest, the situation and specific condition. Moreover there are only few highly efficient preorganized hosts. We can say that creating a rigid host molecule, with the right conformation for the complex formation could be the solution, implying a high number of interactive sites in the good position, with a minor entropy loss due to the formation of the complex. But usually the more rigidified is the conformation the more difficult is the synthesis.
Furthermore we can't forget that in the earlier phase of the complexation a little flexibility in the participants is required, to adapt each other and to get the best orientation for the weak non-covalent interaction to establish. Complexation with a rigid host can be kinetically slow, which is undesirable for many applications. Perfect biological protein receptors possess

considerable conformational mobility. Fischer's lock-and-key hypothesis originally viewed protein binding sites as rigid structures. Only later the conformational mobility of proteins was discovered, and the induced-fit hypothesis was proposed: structures of free and bound proteins often endure considerable conformational changes upon binding. (Figure **1.7**)

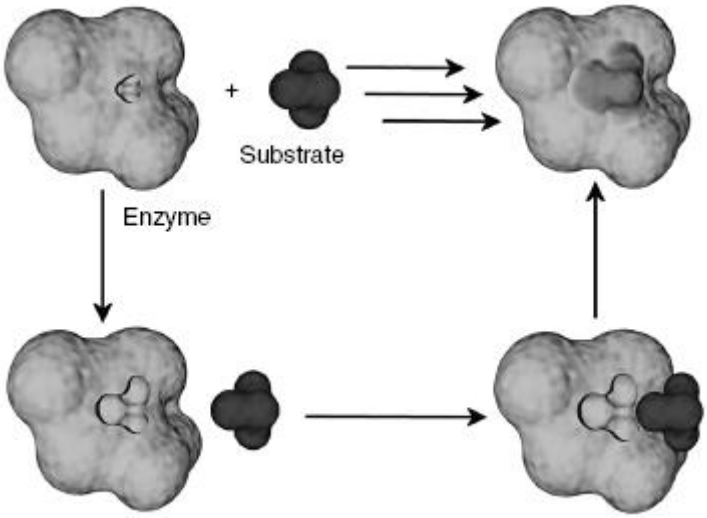


Figure 1.7: The induced-fit model of substrate binding. As the enzyme and substrate approach each other, the binding site of the enzyme changes shape, resulting in a more precise fit between host and guest.

Inspection of structural data for complexes of proteins with low molecular weight guests (such as sugar, biotin and anion binding proteins, enzymes) shows the primary importance of a very large number of pairwise interactions, more than 10 for the small biotin molecule, see figure **1.2**, while a secondary importance is given to a rigid preorganization.

1.4 Interactions Playing Role in Supramolecular Chemistry

The nature of the interactions involved in supramolecular chemistry is non-covalent, as elucidated many times in this chapter, and they can be divided into three categories: electrostatic, van der Waals and solvofobic interactions. The most important interactions are explained below.

- *Charge-charge interactions*: Ionic bonding is comparable in strength to covalent bonding (bond energy = 100–350 kJ mol^{-1}). A typical ionic solid is

sodium chloride, which has a cubic lattice in which each Na^+ cation is surrounded by six Cl^- anions; this simple ionic lattice does illustrate the way in which a Na^+ cation is able to organise six complementary donor atoms about itself in order to maximize non-covalent ion–ion interactions. A supramolecular example of charge-charge interactions is the interaction of a tetranionic cyclophan with an ammonium ion like acethylcoline (Figure **1.8**).

The charge-charge interactions follow the physical rules of classical electrostatic with a force depending on: the charges of the interested species, their distance and the dielectric constant of their medium. Because of the directionless and long range character, this interactions play a fundamental role in the early stages of the complex formation.

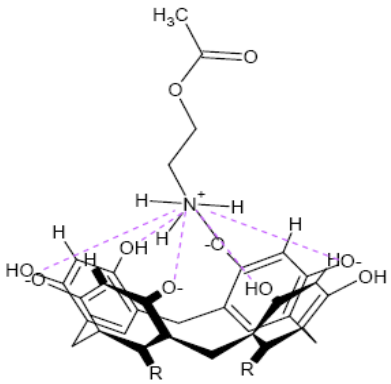


Figure 1.8: Charge-charge interactions between cyclophan and acethylcoline.

- *Hydrogen bond interactions*: these may be regarded as particular kind of dipole–dipole interactions in which a hydrogen atom attached to an electronegative atom such as O or N as the donor (D) is attracted by a similarly electronegative atom, often bearing a lone pair, as the acceptor (A).[16] Hydrogen bonds are commonly written D–H··A.

  Because of its relatively strong and highly directional nature, hydrogen bonding has been described as the "masterkey interaction in supramolecular chemistry".[17] In particular, hydrogen bonds are responsible for the overall shape of many proteins, recognition of substrates by numerous enzymes, and (along with π-π stacking interactions) for the double helix structure of DNA.

Hydrogen bonds come in a wide range of lengths, strengths and geometries and can be divided into three broad categories: strong, moderate and weak interactions.

A strong interaction is somewhat similar in character to a covalent bond, whereby the hydrogen atom is close to the centre-point of the donor and acceptor atoms. Strong hydrogen bonds are formed between a strong acid and a good hydrogen bond acceptor, for example in the $H_5O_2^+$ ion or in complexes of "proton sponge", which are practically linear with the hydrogen atom between the two electronegative atoms.

Moderate strength hydrogen bonds are formed between neutral donor and neutral acceptor groups *via* electron lone pairs, for example the self-association of carboxylic acids, or amide interactions in proteins. Moderate hydrogen bond interactions do not have a linear geometry but are slightly bent.

Weaker hydrogen bonds play a role in structure stabilization and can be significant when large numbers act in concert. They tend to be highly non-linear and involve unconventional donors and acceptors such as C–H groups, the π-systems of aromatic rings or alkynes or even transition metals and transition metal hydrides.

The types of geometries that can be adopted in a hydrogen bonding complex are summarized in figure **1.9**.

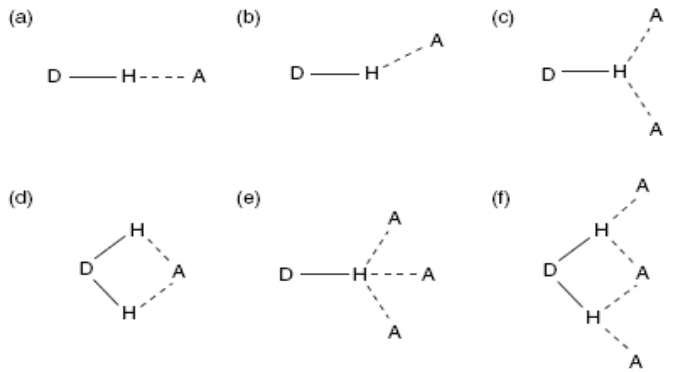


Figure 1.9: Various types of hydrogen bonding geometries; (a) linear, (b) bent, (c) donating bifurcated, (d) accepting bifurcated, (e) trifurcated, (f) three centre bifurcated.

These geometries are termed *primary hydrogen bond interactions*; this means that there is a direct interaction between the donor group and

the acceptor group. There are also *secondary interactions* between neighbouring groups that must be considered. The partial charges on adjacent atoms can either increase the binding strength by attraction between opposite charges or decrease the affinity due to repulsion between charges. Figure **1.10** shows two situations in which arrays of hydrogen bond donors and acceptors are in close proximity. An array of three donors (DDD) facing three acceptors (AAA) (Figure **1.10.a**) has only attractive interactions between adjacent groups and therefore the binding is enhanced. Mixed donor/acceptor arrays (ADA, DAD) suffer from repulsions by partial charges of the same sign being brought into close proximity by the primary interactions (Figure **1.10.b**).

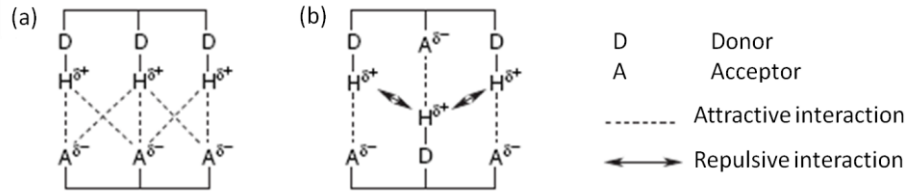


Figure 1.10: (a) Secondary interactions providing attractions between neighbouring groups between DDD and AAA arrays (primary interactions in bold) and (b) repulsions from mixed donor/acceptor arrays (ADA and DAD).

A real-life example of hydrogen bonding is the double helix of DNA. There are many hydrogen bond donors and acceptors holding base pairs together, as illustrated between the nucleobases cytosine (C) and guanine (G) in Figure **1.11**. The CG base pair has three primary interactions and also has both attractive and repulsive secondary interactions.

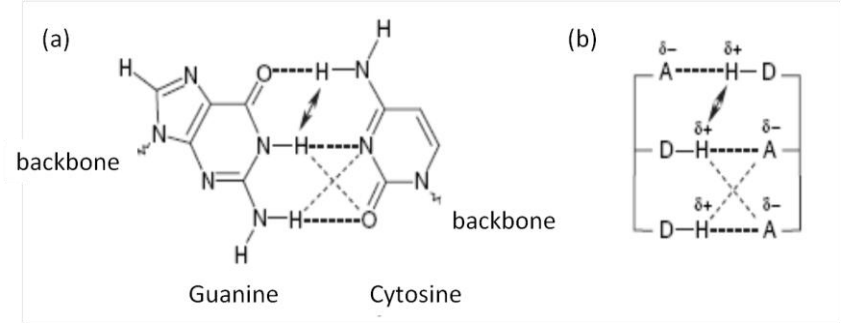


Figure 1.11: (a) Primary and secondary hydrogen bond interactions between guanine and cytosine base pairs in DNA and (b) a schematic representation.

- *Cation- π interactions*: Transition metal cations such as Fe^{2+}, Pt^{2+} *etc.* are well known to form complexes with olefinic and aromatic hydrocarbons such as ferrocene $[Fe(C_5H_5)_2]$ and Zeise's salt $[PtCl_3(C_2H_4)]^-$.[18] The bonding in such complexes is strong and could be considered non-covalent, since it is intimately linked with the partially occupied *d*-orbitals of the metals. Even species such as $Ag^{+}\cdots C_6H_6$ (Figure **1.12**) have a significant covalent component. The interaction of alkaline and alkaline earth metal cations with C=C double bonds is, however, a much more non-covalent "weak" interaction, and is suggested to play an important role in biological systems.

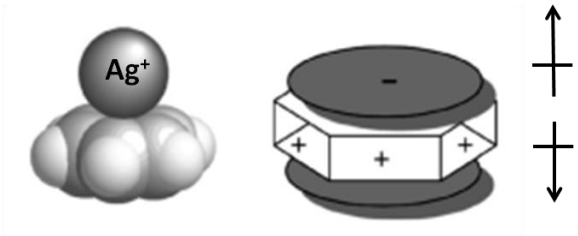


Figure 1.12: Cation- π interactions between a molecule of benzene and Ag^+.

Cation-π interactions are directly dependent on the distance; in fact the ion, because of its charge, has to induce polarization to the charge density of the near molecule. This explains why ammonium ion is not easily complexed by crown ethers and, instead, it interacts effectively with aromatic molecules: the electron density situated on the oxygen of the crown ethers is not sufficient to establish a purely electrostatic bond with a cation. On the contrary-benzene, for example, represents a good interaction site because of its permanent electrostatic potential distributed above and below the plane, which is able to determine attraction towards small cationic species. Moreover from analysis of potential area of aromatic molecules, it was noticed that the maximum of this kind of interaction could be reached when the electron rich specie is positioned in perpendicular direction to the ion. This lead to the synthesis of receptor having an indefinite number of aromatic ring in blocked conformation so that the most favourable geometry is reached.

- *π-π interactions*: Aromatic *π-π* interactions (sometimes called *π-π* stacking interactions) occur between aromatic rings, often in situations where one is relatively electron rich and one is electron poor.[19] There are two general types of *π*-interactions: face-to-face and edge-to-face, although a wide variety of intermediate geometries are known (Figure **1.13**)

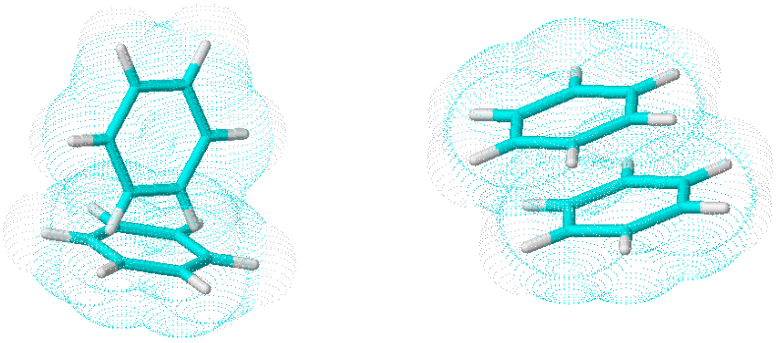


Figure 1.13: Edge-to-face and face-to-face interactions between two molecules of benzene.

Similar *π*-stacking interactions between the aryl rings of nucleobase pairs also help to stabilise the DNA double helix. Edge-to-face interactions may be regarded as weak forms of hydrogen bonds between the slightly electron deficient hydrogen atoms of one aromatic ring and the electron rich *π*-cloud of another ring.
Sanders and Hunter have proposed a simple model based on competing electrostatic and van der Waals influences, in order to explain the variety of geometries observed for *π–π* stacking interactions and to predict quantitatively the interaction energies. Their model is based on an overall attractive van der Waals interaction, which is proportional to the contact surface area of the two *π*-systems. This attractive interaction dominates the overall energy of the *π–π* interaction and may be regarded as an attraction between the negatively charged *π*-electron cloud of one molecule and the positively charged *σ*-framework of an adjacent molecule. The relative orientation of the two interacting molecules is determined by the electrostatic repulsions between the two negatively charged *π*-systems (Figure **1.14**). Sanders and Hunter

stress the importance of the interactions between individual pairs of atoms rather than molecules as a whole.[20]

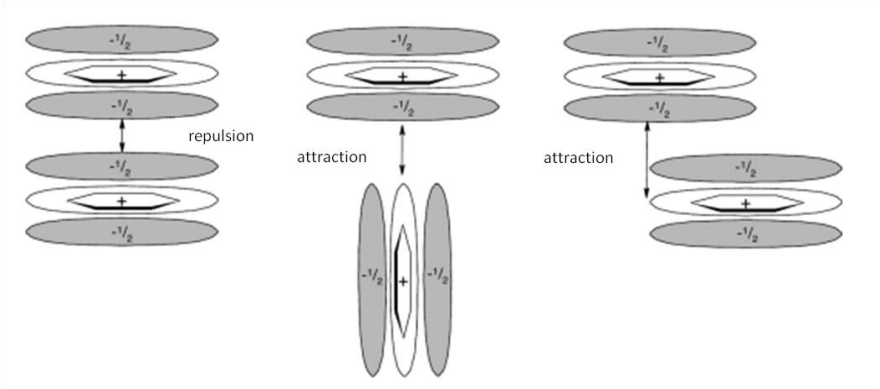


Figure 1.14: Interacting π- quadrupoles.

π-stacking interactions are of considerable interest and importance in the crystal structures of both organic and coordination compounds and have a marked influence on solution binding *via* the hydrophobic effect.

- *Charge transfer interactions*: these are short-range and site specific forces, arising from orbital overlap between the reacting species. They occur with an electron-transfer between high energy molecular orbitals (HOMO, high occupied molecular orbital) of electron rich species and empty low energy molecular orbitals (LUMO, low unoccupied molecular orbital) of electron deficient molecules.[21]
  A charge-transfer interaction[22] has two important effects on the complexation process:
  (1) it stabilizes (or destabilizes) complex formations, affecting the total interaction energy,
  (2) it results in the net partial transfer of charge (i.e., electrons) from one complexing molecule to the other.
  Those kinds of interactions are involved in the complex between a calixarene, which is the electron rich specie because of the presence of hydroxylic groups as substituents, and the fullerene, which is the electron deficient specie (Figure **1.15**).

Figure 1.15. Complex between fullerene C_{60} and a calix[4]arene.

- *Solvofobic interactions*: These play an important role in many cases, such as folding of proteins, protein-protein interactions and tensioactives interactions.
  The aggregation of two molecules in a polar medium through their apolar sites is caused by the balancing of two forces:
  a) the necessary energy to create a cavity in the solvent and so to win the cohesion between the molecules of solvent;
  b) the change of the solvating energy of the molecules participating to the process.
  From a thermodynamic point of view the process is favoured by entropic contribute. In fact, considering Frank and Evans' theory, the hydratation of a molecule which passes from a gaseous phase to solution is a process characterized by a negative entropy – and so an unfavourable process- because a shell of organized solvent molecules is formed around the molecules to solubilize. Such entropic disadvantage is not filled by the favourable enthalpy due to van der Waals interactions between solute and solvent. So aggregation of apolar particles follows and the molecules of solvent, which form the shell of solvation, are freed.
  Solvofobic interactions are directly proportional to the extension of the apolar area of the molecule, which represents the contact area of the complex. They are a good hint for the design of new host. In fact a new tendency is the synthesis of the so called "*deep*" cavitand,[23] i.e. a macrocyclic compound with deep hydrophobic cavity, which allows a better interaction with guest. As an example, calixarenes derivates are

able to complex fullerene C_{60}, utilizing the large hydrophobic area beside the charge transfer (Figure **1.16**).[24]

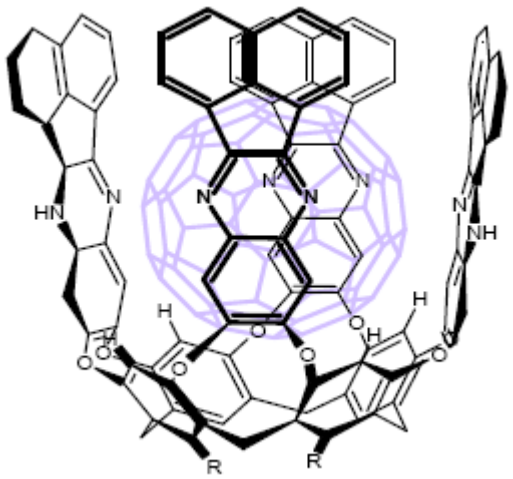


Figure 1.16. "*Deep*" cavitand interacting with fullerene C_{60}.

1.5 Supramolecular chemistry and the proteins

Enzyme structure may be divided into primary, secondary, tertiary and quaternary features.[25] The primary structure is the sequence of amino acid residues on the polypeptide chain and is determined by the way in which the enzyme is synthesized. The secondary structure relates to ordering of the chains into discrete units or segments such as α-helices and β-sheets, while tertiary structure is the way in which the secondary structural features arrange themselves to produce the full globular protein. This occurs *via* hydrogen bonding, stacking interactions and hydrophobic forces, and often involves the participation of water molecules buried deep within the enzyme, where they fill small cavities and act as a "glue" holding secondary features together. In many cases more than one protein strand is involved in a fully functioning enzyme system and the supramolecular association of more than one protein molecule is termed quaternary structure (*e.g.* haemoglobin is a tetramer of four globin units). The enzyme tertiary and quaternary structures are responsible for the organization of the binding site(s); they are also the most flexible components of the enzyme structure,

allowing the binding site to deform in response to guest binding. Hydrophobic effects, resulting in an overall globular shape, are important like the steric effects, which may be responsible for most of the organization; in addition, specific directional hydrogen bonds determine which, among a small number of globular conformations, is the one favoured.

The knowledge of the structure of an enzymatic protein is a fundamental step for understanding its catalytic mechanism. Once the catalytic process has been understood it may be possible to replicate the activity using an entirely synthetic molecule.[26] For example, given the range of zinc enzymes, it is not surprising that model systems have been developed to mimic the active site. The Reinaud group has successfully bond zinc in calyx[6]arenes with three pendent imidazole groups, as shown in figure **1.17** and X-ray structures show that zinc cation binds all three pendent groups while being drawn into the macrocyclic cavity.[27] Crystallographic evidence for propionitrile and heptylamine coordination is suggestive of the proposed structure of the carboxypeptidase.

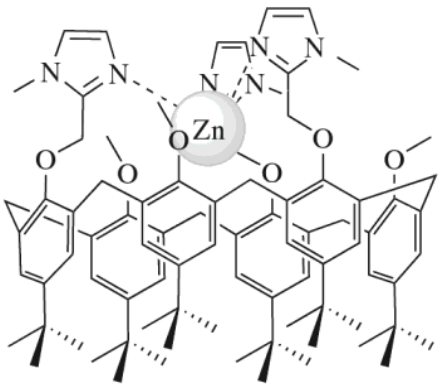


Figure 1.17: the calyx[6]arenes-zinc complex obtained by Reinaud.

Of course, besides the chance to simulate the enzyme activity with artificial compounds, the possibility of inhibiting or modulating this activity is an important aspect of supramolecular chemistry.

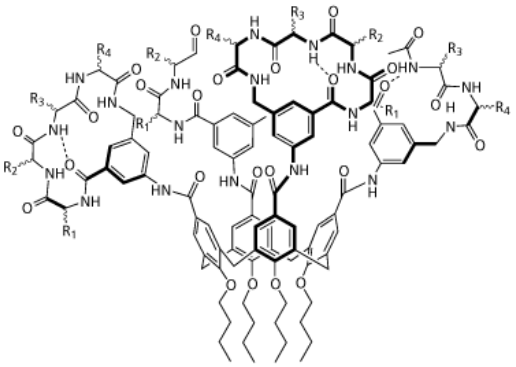
The classical approach generally is based on the production of small molecules to place within the active site or within the allosteric site of the enzyme. A different approach is based on the interaction between the synthetic inhibitor and the external surface of the enzyme, exposed to the

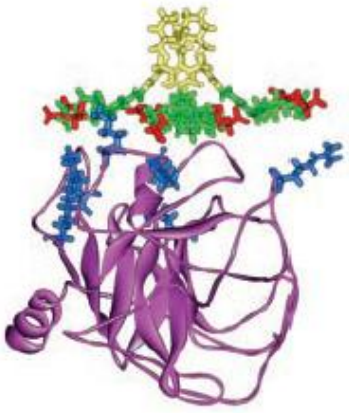
solvent molecules and generally involved in the so-called protein-protein interactions (PPI).

Protein-protein interactions[28] play a key role in many biological processes, and this makes them a fundamental target of novel therapeutics. The disruption of such interactions through synthetic receptors targeted at the protein surfaces[29] is an alternative mechanism to the active site inhibition. With the covalent modification of a protein, for example mediated by polymers, it is possible to modify its biological activity.[30] On the other hand, non-covalent interactions of proteomimetics, i.e., molecules that mimic the structure and function of extended regions of protein surfaces, offer the possibility of reversible binding and modulation of the protein function itself. Among these synthetic receptors, small molecules,[31] R-helix mimetics based on terphenyl or terephthalamide scaffolds,[32] multidentate ligands,[33] functionalized nanoparticle scaffolds,[34] and macrocyclic systems, such as porphyrins[35] and calixarenes[36], have been successfully used as protein inhibitors.

Figure **1.18** shows the structure of a class of synthetic receptors obtained by Hamilton, able to bind to the exterior surface of chymotrypsin, disrupting the interactions with others proteins,[36d] or cytchrome C, blocking its ability to interact with protein partners or simple reducing agents.[36c]

a)





b)

c)

Figure 1.18. General structure of a class of synthetic calixarenic receptors designed by Hamilton.(a) and the calculated complex structure with ChT (b) and cytochrome *c* (c)

1.6 Calixarenes as Artificial Receptors

The calixarenesare macrocyclic molecules containing phenolic rings bridged by methylene groups and are among the most ubiquitous host molecules in supramolecular chemistry.[37-42] The basic molecular scaffolds are, in general, simple to prepare in high yields from cheap starting compounds: they derive from the condensation of phenols and formaldehyde.
These compounds are characterized by an hydrophobic cavity which allows the interactions with neutral molecules,[43] but they can also be derivatized in the lower rim or in para positions on the aromatic nuclei (upper rim) with catalytic centers or other functional groups which favor interactions with ions (Figure **1.19**).[44]

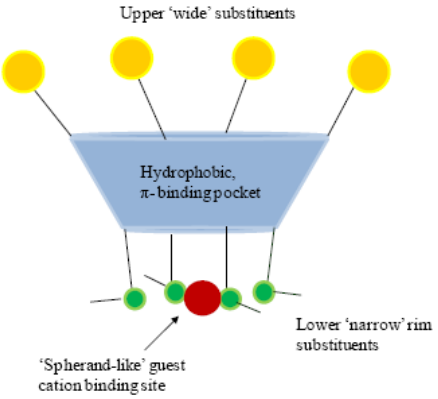


Figure 1.19. Anatomy of a calix[4]arene in the cone conformation.

The parent calixarenes are flexible during their high temperature synthesis, with rotation of the phenolic moieties about the bridging CH_2 groups possible, but the smallest members of the class "freeze out" upon cooling to ambient temperatures.[45] This is an important consideration when working with calix[4]arenes as they exist in four conformers that are hard to interconvert and become immobilized in a particular conformer if substituents are bound in the lower rim, even if re-heated to relatively high temperatures. Four principal conformers are observed at room temperature. If all four upper rim substituents are in the same orientation, then a cone conformer results, which is the average C_{4v} symmetry structure resulting from a fast equilibrium between two equivalent C_{2v} flattened cone conformers.[46] If one phenolic group is inverted with respect to the others, a partial cone conformer is found. Finally, two possibilities exist when two

phenol rings are inverted: 1,2 alternate and 1,3 alternate. All four conformers are illustrated in Figure **1.20**. Similar descriptions exist for larger calixarenes although these compounds are often conformationally dynamic and only frozen out in the solid state or in low temperature experiments.

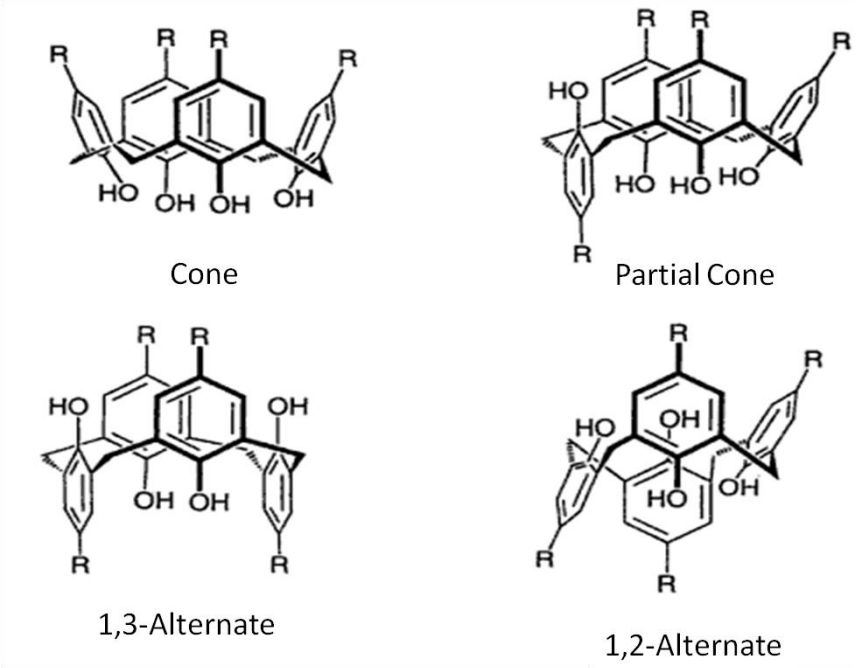
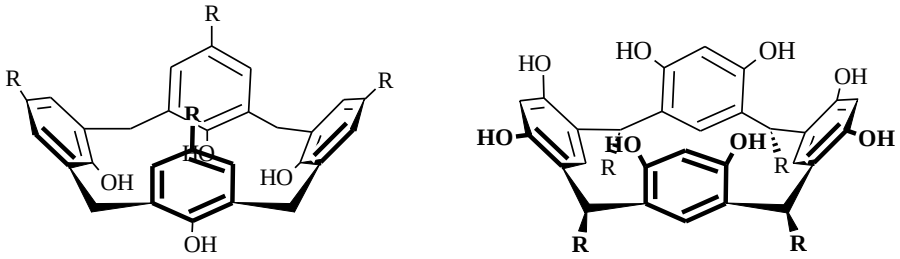


Figure 1.20. Calix[4]arene conformers: cone (top left), partial cone (top right), 1,2-alternate (bottom left), 1,3-alternate (bottom right)

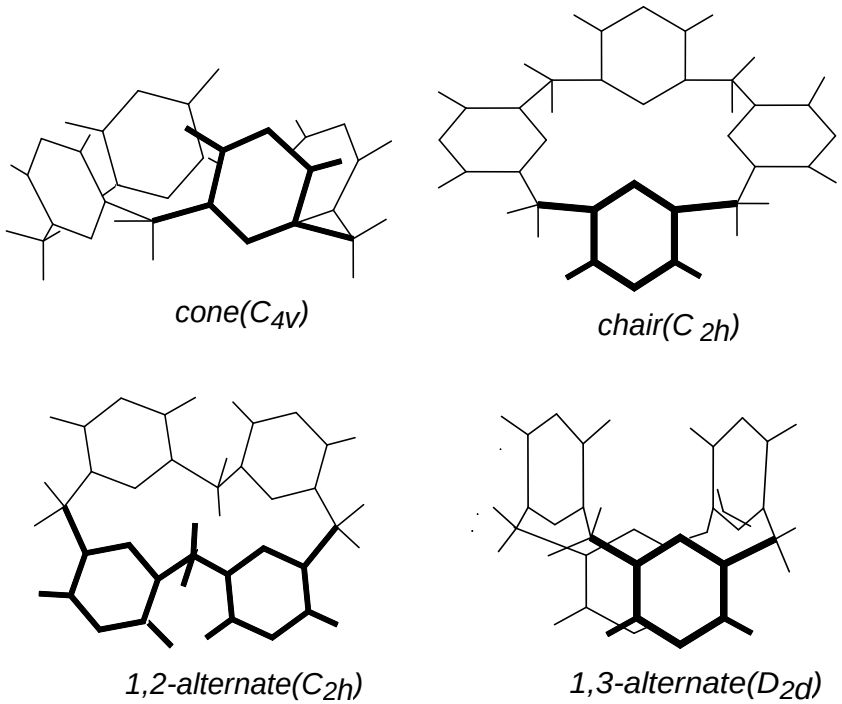
The *resorcarenes* (or *calixresorcarenes*) (Figure **1.21**) are closely related compounds. These are prepared in the similar way of calixarenes by condensation of resorcinol (3-hydroxyphenol) with aldehydes. In this case, acid-catalysed conditions are used and the preparation does not work with formaldehyde itself because of the consequent polymerisation reactions from the 2-position. However a wide range of other aldehydes are highly effective and commonly acetaldehyde (able to give a methyl ”feet” to the resorcarene bowl) or 2-phenylethanal (resulting in enhanced solubility of the product in organic solvents) are used.



calix[4]arene resorc[4]arene

Figure 1.21. Nuclei of calix[4]arene and resorc[4]arene in cone conformation.

Also resorc[4]arenes possess a bowl-shaped conformation in their most stable form. This bowl is wider and shallower than in the calixarene analogues as a consequence of the presence of the "upper rim" hydroxy substituents, which stabilizes the bowl by intramolecular hydrogen bonding. Resorcarenes bearing small substituent groups are relatively conformationally mobile, adopting partial cone, 1,2-alternate and 1,3-alternate conformations as well as the cone form. (Figure **1.22**)



cone(C_{4V}) *chair(C_{2h})*

1,2-alternate(C_{2h}) *1,3-alternate(D_{2d})*

Figure 1.22: principal conformations for the resorcin[4]arene

Using the cone resorcin[4]arene as central macrocyclic scaffold,[47a,b] functionalized on the lower rim by binding four constrained dipeptide chains, a new class of resorc[4]arenes has been obtained.
The conformational flexibility of the chains, combined with the multiplicity of the contact sites, both offered by the variable peptide sequence, make these compounds adapt to interact with the convex and solvent exposed surface of proteins.[48]
The loss of conformational entropy in the receptor, occurring during the formation of the complex with the protein, is limited by the structural preorganization offered by the same macrocyclic scaffold.
Another fundamental aspect of the recognition process is the insertion of charged hydrophilic carboxylate groups or lipophilic benzyl groups in the terminal amino acid, that could be involved in electrostatic or hydrophobic interaction, respectively, with the protein surface.[49]

On the basis of all these structural considerations, we propose our *N*-linked peptidoresorc[4]arenes as synthetic receptors for protein surface interaction, starting from the ex novo synthesis of the *N*-linked peptidoresorc[4]arene benzyl esters **10**, **11** and **16**, and, by reductive hydrolysis, the corresponding free carboxylic acids **12**, **13** and **17**.
Previous studies highlighted an effective interaction between **12** and human serum albumin (HSA), the most common protein in the serum. Based on this starting point, we explored the binding, and the consequent inhibitory activity, of this family of calix[4]arene receptors towards an enzyme commonly used in studies targeting surface inhibition mechanisms, i.e. α-chymotrypsin (ChT). [34,49]

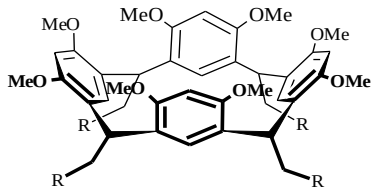
2. Results and discussion

2.1 Receptors Design and Synthesis

Moving from the idea developed by Hamilton, who created an important class of protein surface receptors based on the attachment of four cyclic peptides to a calix[4]arene scaffold and targeted them to the serine protease ChT,[36b] preliminary molecular modeling studies were performed, suggesting the design of tetra-anionically functionalized receptors that might target the predominantly cationic region of ChT surrounding the active site.

In previous works, it has been shown that resorc[4]arene octamethyl ethers functionalized at the feet with valine units[47c] exploited their capability of enantiodiscriminating amino acidic guests in the gas phase by mass spectrometry.

Moving by the evidence that *N*- or *C*-linked peptides calixarenes were featured by new properties, such as recognition of protein surfaces and carbohydrates, formation of stable inclusion complexes, and self-assembly phenomena, valyl-leucine and leucyl-valine peptidoresorc[4]arenes[47a,b] were synthesized, choosing the cone conformation of the resorc[4]arene core to get the all-cis preorganization of the peptide chains: in this way these flexible substituents were located on the same side of the scaffold. (Figure **2.1**)



R

1 = COOEt

2 = COOH

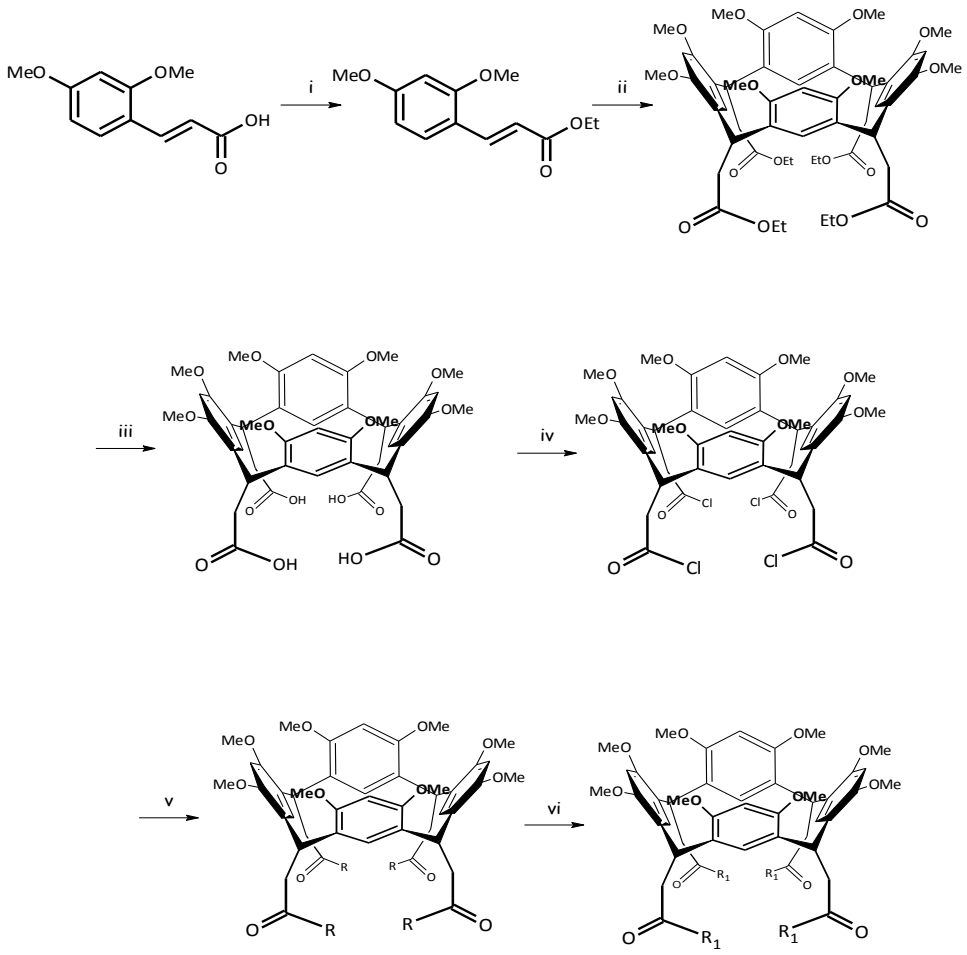
3 = COCl

	X	Stereochemistry
R = CO-NH-CH-CONH-CH-COOX		
(-)-10a	Bz	LL
(+)-10b	Bz	DD
(-)-12a	H	LL
(+)-12b	H	DD
R = CO-NH-CH-CONH-CH-COOX		
(-)-11a	Bz	LL
(+)-11b	Bz	DD
(-)-13a	H	LL
(+)-13b	H	DD
R = CO-NH-CH-CONH-CH-COOX (COOX)		
(-)-16a	Bz	LL
(+)-16b	Bz	DD
(-)-17a	H	LL
(+)-17b	H	DD

Figure 2.1. *N*-Linked peptidoresorc[4]arene benzyl esters (**10**, **11**, and **16**) and free acids (**12**, **13**, and **17**).

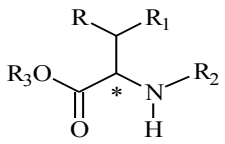
Among all the *N*-Linked peptidoresorc[4]arene we studied, the derivatives containing the valyl-leucine and leucyl-valine chains have been previously obtained in our laboratory by Dr. Fabiana Subrizi, and their synthesis was then repeated and optimized, while the valyl-aspartic acid derivatives have been synthesized entirely from the beginning.

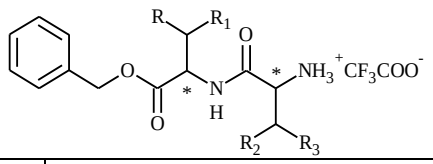
Following the scheme of a previous work,[47a] resorc[4]arene octamethyl ether tetraester **1**, obtained as previously described,[50] was hydrolyzed to tetracarboxylic acid **2** and subsequently treated with thionyl chloride in dry benzene, affording the corresponding acid chloride derivative **3** in quantitative yield. (Scheme **2.1**)



i = $EtOH/H_2SO_4$ reflux 4h
ii = BF_3-$Et_2O/CHCl_3$ reflux 40 min, purification
iii = NaOH 2N/EtOH reflux 4h
iv = $SOCl_2$/Bn dry reflux 4h
v = DIPEA/THF 40 min, NH_2-pep-OBz TFA salt, purification
vi = H_2/Pd 10% overnight

Scheme 2.1: Synthetic procedure for the synthesis of peptidoresorc[4]srenes benzyl ester and free carboxylic acid

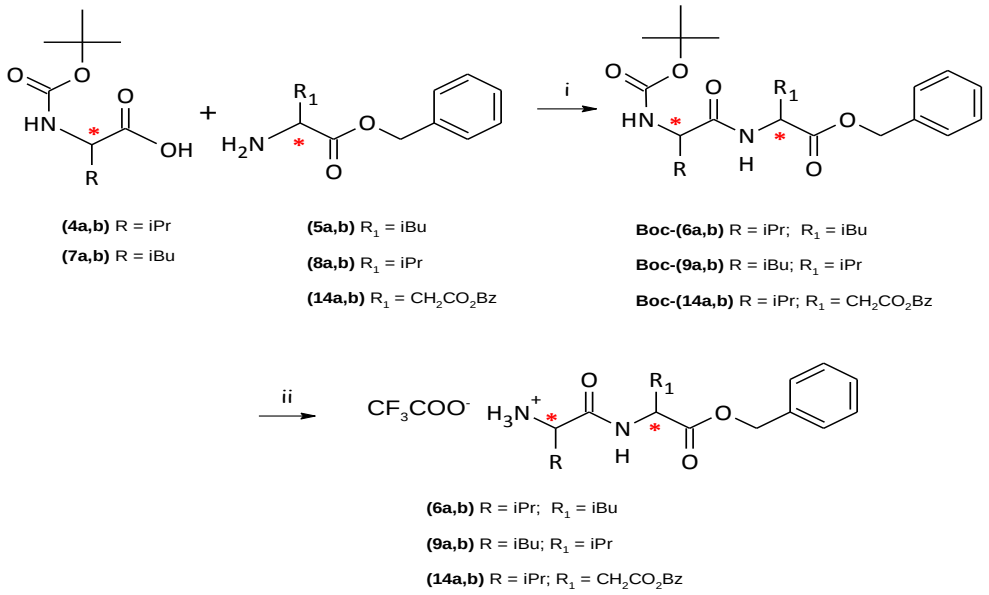




4a	R = R_1 = CH_3, R_2 = BOC, R_3 = H, * = L	**6a**	R = $(CH_3)_2CH$, R_1 = H, R_2 = R_3 = CH_3, ** = L,L
5a	R = $(CH_3)_2CH$, R_1 = H, R_2 = H· *p*-toluen-sulfonic acid, R_3 = Bz, * = L	**9a**	R = R_1 = CH_3, R_2 = $(CH_3)_2CH$, R_3 = H, ** = L,L
7a	R = $(CH_3)_2CH$, R_1 = H, R_2 = BOC, R_3 = H, * = L	**15a**	R= COOBz, R_1 = H, R_2=R_3 = CH_3, ** = L,L
8a	R = R_1 = CH_3, R_2 = H· *p*-toluensulfonic acid, R_3 = Bz, * = L	**6b**	R = $(CH_3)_2CH$, R_1 = H, R_2 = R_3 = CH_3, ** = D,D
14a	R= COOBz, R_1=H, R_2 = H· *p*-toluensulfonic acid, R_3 = Bz, * = L	**9b**	R = R_1 = CH_3, R_2 = $(CH_3)_2CH$, R_3 = H, ** = D,D
4b	R = R_1 = CH_3, R_2 = BOC, R_3 = H, * = D	**15b**	R= COOBz, R_1 = H, R_2=R_3 = CH_3, * = D,D
5b	R = $(CH_3)_2CH$, R_1 = H, R_2 = H· trifluoro acetic acid, R_3 = Bz, * = D		
7b	R = $(CH_3)_2CH$, R_1 = H, R_2 = BOC, R_3 = H, * = D		
8b	R = R_1 = CH_3, R_2 = H· trifluoro acetic acid, R_3 = Bz, * = D		
14b	R= COOBz, R_1=H, R_2 = H· trifluoro acetic acid, R_3 = Bz, * = D		

Figure 2.2. Building blocks of the dipeptide chains.

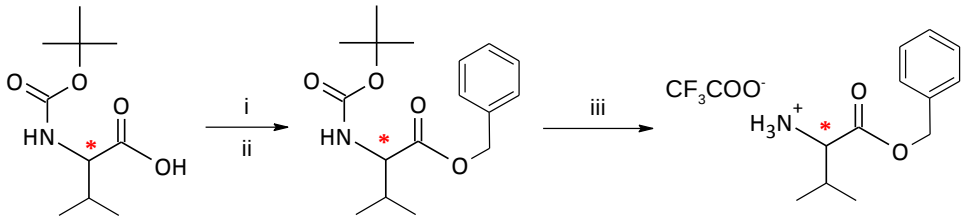
For the preparation of L-valyl-L-leucine benzyl ester (**6a**, Figure **2.2**), commercial *N*-Boc-L-valine **4a** was coupled with L-leucine benzyl ester p-toluensulfonate **5a** in the presence of HOBT and triethylamine to afford *N*-Boc-L-valyl-L-leucine benzyl ester, which after deprotection with TFA/DCM, gave the dipeptide **6a** as TFA salt. Similarly, L-leucyl-L-valine benzyl ester **9a**, as TFA salt, was prepared starting from commercial N-Boc-L-leucine **7a** and L-valine benzyl ester p-toluensulfonate **8a**. In both cases, the overall yield was about 85%. The same peptide coupling procedure was applied to the DD-dipeptides series: D-valyl-D-leucine benzyl ester **6b** and D-leucyl-D-valine benzyl ester **9b**, TFA salts, were obtained with comparable yields, starting from the pairs **4b/5b** and **7b/8b**, respectively. (Scheme **2.2**)



i = HOBT/TEA/DCC in DCM (purification)
ii = DCM/TFA

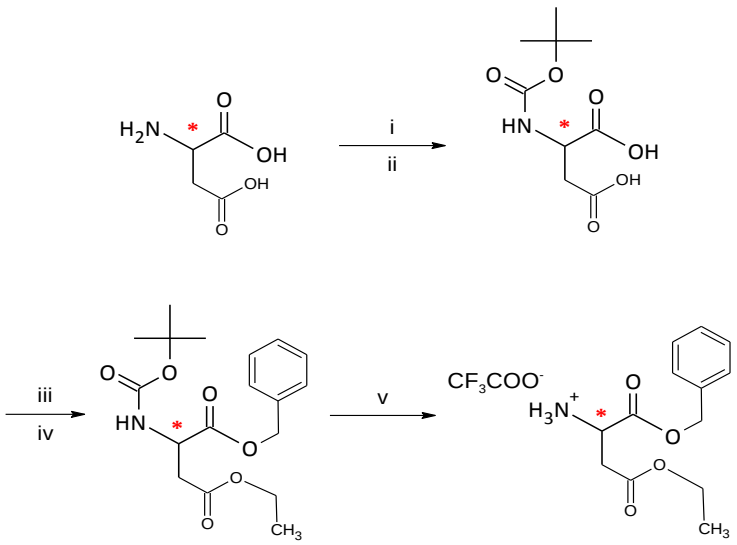
Scheme 2.2: Synthesis of the dipeptide chains benzyl esters as TFA salts

Notably, because **5b** and **8b** were not easily commercially available, they were ad hoc synthesized as TFA salts. (Scheme **2.3** and **2.4**)
The overall yield of both the dipeptides was about 80%.



* = D configuration
i = Cs_2CO_3/ DMF, 0 °C, 45-60 min; ii = BzBr, 0 °C 30 min; RT 12h
iii = TFA/DCM

Scheme 2.3: synthesis of **5b**



* = D configuration
i = $(Boc)_2O$, $NaOH_{(aq)}$ dioxane 0 °C; ii = $HCl_{(aq)}$
iii = Cs_2CO_3/ DMF, 0 °C, 45-60 min; iv = BzBr, 0 °C 30 min; RT 12h
v = TFA/DCM

Scheme 2.4: synthesis of **8b**

The four **6a**, **6b**, **9a**, and **9b** dipeptides were conjugated individually with macrocycle **3** as follows: DIPEA was added under nitrogen atmosphere to a solution of **3** in dry THF, and the reaction mixture was stirred at room temperature; after the addition of the proper dipeptide benzyl ester in THF, the mixture was refluxed for 1 h. Standard purification gave, in 65-75% yields, the four *N*-linked peptidoresorc[4]arenes benzyl esters **10a**, **10b**, **11a**, and **11b** (Figure **2.1**), which in turn, by hydrogenation with 10% Pd/C, afforded the free carboxylic acids **12a**, **12b**, **13a**, and **13b** (Figure **2.1**).
Previous studies performed with **12** and HSA pointed out an interaction between the receptor and the protein. According to this evidence, the hypothesis of an electrostatic interaction between the anionically charged resorc[4]arene derivatives and the positively charged surface of the α-chymotrypsin was formulated. The positive response of the experiments led us to examine four new compounds, in which the number of terminal functional groups was doubled, either for a lipophilic or electrostatic interaction with proteins. L-valyl-L-aspartic acid dibenzyl ester (**15a**) and D-valyl-D-aspartic acid dibenzyl ester (**15b**) were thus prepared by the usual

procedure, starting from the proper commercial *N*-Boc-valine (**4a** or **4b**) and the proper aspartic acid dibenzyl ester (commercial **14a** or ad hoc synthesized **14b**). The dibenzylated peptidoresorc[4]arenes **16a** and **16b**, obtained by the coupling with **3** of **15a** and **15b**, respectively, finally gave the corresponding free dicarboxylic acids **17a** and **17b** by Pd/C hydrogenation in quantitative yields. All the new compounds were structurally confirmed by NMR spectroscopy and electrospray ionization high-resolution mass spectrometry, as well as by the optical rotation values (see Table **2.1**).

Table 2.1. Physical Properties of Chiral N-Linked Peptidoresorc[4]arenes

Compd	peptide sequence	ESI-HRMS (m/z)	mp (°C)	$[\alpha]^{20}_D$ (c, solvent)
(-)-**10a**	L-Val-L-Leu-OBz	1043.53604 ($[M+2Na]^{2+}$)	220-221	-80.5 (1.0, $CHCl_3$)
(+)-**10b**	D-Val-D-Leu-OBz	1043.53635 ($[M+2Na]^{2+}$)	221-222	+79.4 (1.0, $CHCl_3$)
(-)-**11a**	L-Leu-L-Val-OBz	1043.53549 ($[M+2Na]^{2+}$)	248-249	-76.7 (0.3, $CHCl_3$)
(+)-**11b**	D-Leu-D-Val-OBz	1043.53656($[M+2Na]^{2+}$)	251-252	+75.9 (1.0, $CHCl_3$)
(-)-**12a**	L-Val-L-Leu-OH	566.62079 ($[M-4HþNa]^{3-}$)	212-213	-55.5 (1.0, MeOH)
(+)-**12b**	D-Val-D-Leu-OH	566.62022 ($[M-4HþNa]^{3-}$)	209-211	+52.5 (1.0, MeOH)
(-)-**13a**	L-Leu-L-Val-OH	566.62026 ($[M-4HþNa]^{3-}$)	135-136	-49.2 (0.6, MeOH)
(+)-**13b**	D-Leu-D-Val-OH	566.61975 ($[M-4HþNa]^{3-}$)	135-136	+46.3 (0.7, MeOH)
(-)-**16a**	L-Val-L-Asp-OBz	1228.01773 ($[M+2Na]^{2+}$)	218-219	-35.7 (0.5, MeOH)
(+)-**16b**	D-Val-D-Asp-OBz	1228.01416 ($[M+2Na]^{2+}$)	218-219	+35.9 (0.5, MeOH)
(-)-**17a**	L-Val-L-Asp-OH	843.33097 ($[M-2H]^{2-}$)	201-202[a]	-18.3 (0.1, MeOH)
(+)-**17b**	D-Val-D-Asp-OH	843.32839 ($[M-2H]^{2-}$)	203-204[a]	+17.7 (0.1, MeOH)

[a] With decomposition.

The discrepancy between the optical rotations of the DD and LL-forms of compounds **12** and **13** was demonstrated not to be dependent by an isomerization phenomenon: enantiomeric excesses (ee, %), checked by enantioselective HPLC, were in the 95-97% range.

2.2 NMR Analysis of *N*-Linked Peptidoresorc[4]arenes

In a previous paper,[47a] the symmetry of the starting cone conformer **1** was shown, with four symmetry planes (C_{4v}). These symmetry planes are no more present in the *N*-linked peptide derivatives **10**-**13**, **16**, and **17** (Figure **2.3**). The introduction of the chiral centers of the peptide chains reduces the symmetry of the derivatives: the planes R1/R3 and R2/R4 contain the chiral centers, whereas the two planes A/C and B/D separate two moieties that are no longer mirror images (the DD-configuration of a chain should

correspond the LL-configuration of the opposite chain). Therefore, C-1 and C-3 carbons, for instance, are no longer equivalent and give two distinct signals.

Because of the superimposability of the four chains, including the chiral centre, after a 90° rotation, we can still recognize a C_4 symmetry axis. So the C-1 carbon will be equivalent to C-7, C-13, and C-19 and C-3 will coincide with C-9, C-15, and C-21. Carbons (and protons) equidistant from the chiral centers, such as C-e/C-I (or H-e/H-i), show a single signal as also observed in the achiral resorc[4]arene **1** precursor, whereas all of the other nuclei cumulatively produce two signals. The distribution pattern of carbons and protons in Tables 2-4 is in agreement with these considerations.

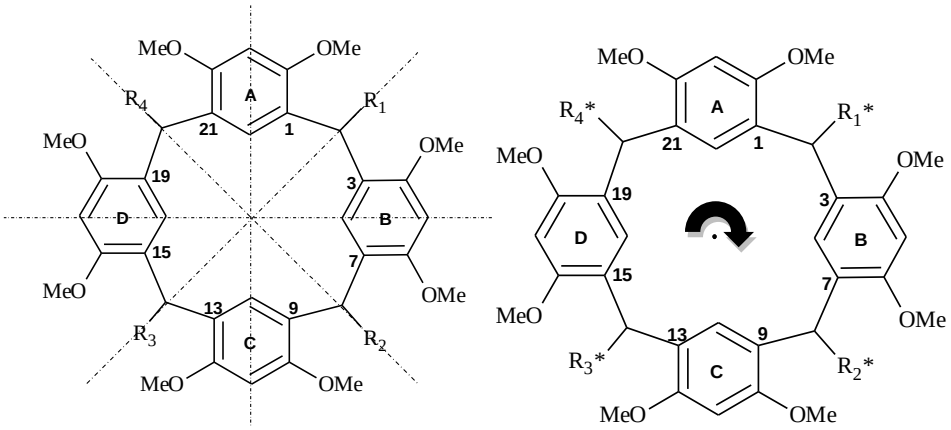


Figure 2.3. Symmetry in the achiral molecule precursor (left) and the final chiral product (right).

Table 2.2. ^{1}H NMR Signals* (ppm) of Peptidoresorc[4]arenes **10**–**13**[#a]

Proton	Resorc[4]arene **10**	Resorc[4]arene **11**	Resorc[4]arene **12**	Resorc[4]arene **13**
CH_i	6.80 *br* s	6.80 *br* s	6.78 *br* s	6.66 *br* s
CH_e	6.22 s	6.24 s	6.30 s	6.41 s
CH	4.86 t (7.0)	4.86 t (6.8)	4.88 *br* t (7.0)	4.89 *br* t (7.0)
CHN (leu)	4.48 *br* q (7.0)	4.48 *br* q (7.0)	4.45 *br* q (6.6)	4.43 *br* q (7.0)
CHN (val)	4.09 t (7.0)	4.15 t (8)	4.32 t (7.0)	4.28 br t (7.0)
OCH_3	3.69 *br* s 3.59 *br* s	3.70 *br* s 3.61 *br* s	3.66 *br* s 3.62 *br* s	3.66 *br* s 3.62 *br* s
CH_2(C=O)	2.89 dd (13.3, 7.0) 2.67 dd (13.3, 7.0)	2.89 dd (13.6, 6.8) 2.67 dd (13.6, 6.8)	2.81 *br* d (7.0)	2.84 dd (13, 6.7) 2.77 dd (13, 6.7)
CH (val)	1.85 m	1.86 m	2.07 m	2.06 m
CH (leu)	1.58 m	1.59 m	1.44 m	1.46 m
CH_2 (leu)	1.52 m	1.51 m	1.39 m	1.30 m
CH_3	0.85 d (5.5) 0.78 d (6.3) 0.63 d (6.3) 0.53 *br* s	0.85 d (6.0) 0.80 d (6.0) 0.78 d (6.5) 0.53 *br* s	0.89 d (6.0) 0.85 d (6.0) ×2 0.72 *br* s	0.88 d (6.5) 0.86 d (6.5) x2 0.80 d (6.5)
CO-NH	7.14 *br* d *not detected*	7.14 *br* d *not detected*	8.39 *br* d (8.0) 7.75 *br* d (8.0)	8.14 *br* d (8.0) 7.93 *br* d (8.0)

* 400 MHz, δ scale, $CDCl_3$ (for **10** and **11**) and CD_3OD (for **12** and **13**), *T* = 298 K; coupling
constants *J* (Hz) are given in parentheses.

[#] The reported spectra are exactly those of compounds (−)-**10a**, (−)-**11a**, (−)-**12a**, and (−)-**13a**. Since the two enantiomers gave superimposable NMR and MS spectra, we use the number without index to mean any member of the enantiomeric pair.

[a]Signals for the benzyl group: (**10**) 7.30 (5H, *br* s; C_6H_5), 5.13, 5.05 (1H each, d, *J* = 12.3 Hz; CH_2); (**11**) 7.30 (5H, *br* s; C_6H_5), 5.13, 5.05 (1H each, d, *J* = 12.4 Hz; CH_2).

Table 2.3. ^{13}C NMR Signals* (ppm) of Peptidoresorc[4]arenes **10**–**13**[#a]

Proton	Resorc[4]arene **10**	Resorc[4]arene **11**	Resorc[4]arene **12**	Resorc[4]arene **13**
HN-C=O	172.80 172.47	173.22 171.31	173.55 171.44	173.76 173.33
O-C=O[a]	172.00	172.64	178.16	*not detected*
C_{Ar}-O	156.57 156.12	156.18 156.09	156.40 156.18	156.27 x 2
CH_i	127.76	127.45	125.86	126.86
C_{Ar}-C	123.54 123.32	123.96 124.44	124.05 123.83	123.95 123.58
CH_e	96.57	96.19	96.36	96.26
CHN (val)	58.86	57.59	59.39	58.75
OCH_3	56.37 55.83	56.45 55.93	55.01 54.93	55.34 54.82
CHN (leu)	51.16	52.19	51.12	51.57
CH (leu)	40.91	40.54	42.29	40.73
$\underline{CH_2}$(C=O)	40.68	40.04	40.44	40.11
CH	31.98	33.47	33.17	33.29
CH (val)	30.51	31.10	30.42	30.45
CH_2 (leu)	24.74	24.66	24.66	24.09
CH_3	22.58 22.12 19.11 18.19	22.81 22.54 18.97 17.86	22.30 21.30 18.56 17.81	22.16 21.16 18.33 17.14

* 75 MHz, δ scale, $CDCl_3$ (for **10** and **11**) and CD_3OD (for **12** and **13**), *T* = 298 K.

[#] The reported spectra are exactly those of compounds (−)-**10a**, (−)-**11a**, (−)-**12a**, and (−)-1**3a**. Since the two enantiomers gave superimposable NMR and MS spectra, we use the number without index to mean any member of the enantiomeric pair.

[a] Signals for the benzyl group: (**10**) 135.66, 128.51, 128.32, 128.23, 66.77; (**11**) 135.55, 128.63, 128.58, 128.44, 66.79.

Table 2.4. ^{1}H and ^{13}C NMR Signals* (ppm) of Peptidoresorc[4]arenes **16** and **17**[#]

	Peptidoresorc[4]arene 16		**Peptidoresorc[4]arene 17**	
Carbon	**^{13}C**	**^{1}H**	**^{13}C**	**^{1}H**
CO–NH	173.80 173.34	–	175.09 173.50	–
CO–O	170.35 × 2	–	174.14 174.06	–
C_{Ar}–O	156.58 156.10	–	157.99 157.61	–
C_6H_5	135.53 × 2 135.18 × 4 132.32 × 4 128.38 × 2	7.35–7.20 10H, m	–	–
CH_i	127.37	6.71 s	127.82	6.75 s
CAr–C	123.85 123.32	–	125.24 123.83	–
CHe	96.47	6.23 s	97.64	6.46 s
OCH_2Bz	67.34	5.04 d (12.0) 5.01 d (12)	–	–
OCH_2Bz	66.71	5.01 d (12.0) 4.97 d (12.0)	–	–
CHN (val)	58.83	4.20 *br* d (6.0)	60.18	4.18 d (7.3)
OCH_3	55.79	3.58 s 3.54 s	56.41	3.70 s 3.68 s
CHN (asp)	48.56	4.86 *br* t (3.8)	50.00	4.77 *br* t (3.5)
CH_2(CO)	40.51	2.82 dd (15.0, 7.0) 2.69 dd (15.0, 7.0)	41.81	2.97 dd (15.0, 7.3) 2.82 dd (15.0, 7.3)
CH_2 (asp)	36.04	2.86 d (3.8)	37.06	2.73 d (3.5)
CH	34.30	4.87 t (7.0)	34.61	4.98 t (7.3)
CH (val)	30.70	1.94 m	31.90	2.06 dq (7.3, 6.5)
CH_3	18.96 18.25	0.80 d (6.0) 0.68 d (6.0)	19.65 19.02	0.97 d (6.5) 0.81 d (6.5)

* 600 MHz (^{1}H) and 150 MHz (^{13}C), $CDCl_3$ (for **16**) and CD_3OD (for **17**), *T* = 298 K; coupling constants *J* (Hz) are given in parentheses.

[#] The reported spectra are exactly those of compounds (+)-**16b** and (+)-**17b**. Since the two enantiomers gave superimposable NMR and MS spectra, we use the number without index to mean any member of the enantiomeric pair.

2.3 Physical Characterization of *N*-Linked Peptidoresorc-[4]arenes

Physical properties of N-linked peptidoresorc[4]arenes are summarized in Table 1. In the ESI-HRMS characterization of peptidoresorc[4]arene benzyl esters **10**, **11**, and **16** (positive mode), a strong tendency was observed to form complexes with sodium (Na) and potassium (K) ion impurities, as already reported.[10a] The most intense peak in the spectrum was that for the double charged sodium adduct ion $[M + 2Na]^{2+}$, at m/z 1043.53604-1043.53656 for **10** and **11** (which are indeed regioisomers), and atm/z 1228.01773 for **16**. The corresponding monocharged species, $[M + Na]^{+}$, were also observed. The free acids **12** and **13**, which are indeed regioisomers, displayed in the mass spectra (negative mode) the most abundant ion corresponding to the fully deprotonated molecule as sodium adduct ($[M - 4H + Na]^{3-}$), at m/z 566.61975-566.62079, in agreement with their molecular formula. For free dicarboxylic acids **17a**,**b**, a diagnostic peak at $[M - 2H]^{2-}$ was found in the mass spectrum (negative mode), at m/z 843.32839-843.33097. Melting points were normally restricted in a 1.0 °C interval, but the free dicarboxylic acids (-)-**17a** and (+)-**17b** decomposed instead of melting. Finally, optical rotations for the two members of each enantiomeric pair were comparable in most cases (see Table **2.1**), excluding any significant racemization. Also for the enantiomeric pairs **12a**,**b** and **13a**,**b**, that present quite different values of optical rotation, the racemization was ruled out by enantioselective HPLC. (This part of the experimental work has been conducted by **Dr. Fabiana Subrizi** in collaboration with Prof. Francesco Gasparrini, Sapienza University of Rome.) Notably, the signs of optical rotation are correlated with the stereochemistry of peptidoresorc[4]arenes: the L,L-series always show negative values.

2.4 Initial Screening for Inhibition Activity

The inhibition activity of the synthesized *N*-linked peptidoresorc[4]arenes was evaluated by in vitro assays of bovine pancreatic α-chymotrypsin (α-ChT), a serine protease which was already shown to be inhibited by a specific family of calix[4]arene receptors,[36b] using a chromogenic substrate.

ChT hydrolyzes proteins selectively attacking the protein chain next to aromatic residues such as tyrosine (Tyr) and phenylalanine (Phe), but it has also affinity for other amino acids bearing apolar side chains (e.g., methionine, leucine).[51] This enzyme, which is characterized by the presence of positive charges on its surface, as suggested by the isoelectric point of 8.8, has been considered a suitable target for studying the interaction with receptors containing terminal carboxylic acids, in anionic form at the physiological pH.[34,49] At the same pH, some basic amino acids surrounding the active site are protonated (Figure **2.4**): electrostatic interactions with the positively charged patches of ChT are thus expected to lead to a consequent inhibition of the enzyme. Inhibition assays should thus provide a useful handle for studying the protein-receptor complex.

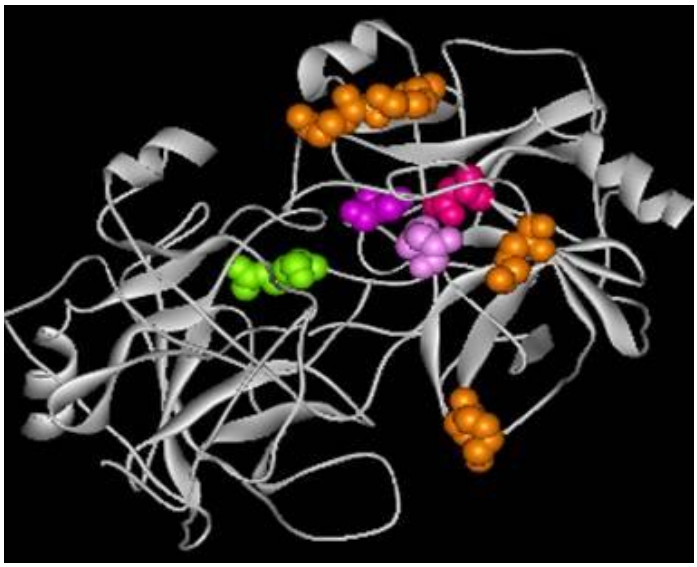


Figure 2.4: representation of α-chymotrypsin with highlighted in purple the catalitic triad and in orange and green some basic amino acids (Lys-36, Lys-90, Lys-175, and Arg-145) in the neighborhood of the active site cleft.

Preliminary studies, **performed in our laboratories by Dr. Fabiana Subrizi,** showed an effective inhibition of ChT by peptidoresorc[4]arenes free acids **12** and **13**. The inhibition was investigated by monitoring at 405 nm the rate of enzymatic hydrolysis of the chromogenic substrate N-succinyl-Ala-Ala-Pro-Phe-p-nitroanilide (Suc-AAPF-pNA) to p-nitroaniline, after the mixing

of protein and receptors. Control experiments were carried out under identical conditions without the addition of the resorc[4]arenes.

In order to optimize the procedure and confirm the reproducibility of the results, the experiments were repeated, according to the following protocol: the enzyme (4.0 x 10^{-8} M) was treated with solutions of **12a**, **12b**, **13a**, or **13b** (4.0x10^{-5} M) in 0.005 M sodium phosphate buffer for various preincubation periods followed by addition of Suc-AAPF-pNA in 2-methoxyethanol (cosolvent). The inhibitory effect is shown in Table **2.5**, and for this set of compounds it resulted to be time-independent. Receptors (-)-**12a** and (+)-**12b** were found to decrease the Suc-AAPF-pNA hydrolysis to 13 and 11%, respectively, as compared with those of the untreated enzyme, whereas (-)-**13a** and (+)-**13b** led to a smaller inhibition.

The greater inhibition by the pair **12a**/**12b** (with the valyl-leucine instead of leucyl-valine sequence) can be attributed to a preference of the enzyme for the cleavage of dipeptide chains containing leucine next to hydrophobic residues. The dipeptide configuration seems to be not so important, even though the DD-derivatives were slightly more active than the LL-enantiomers, which are the natural ligands of the enzyme.

Table 2.5. Residual Activity (%) of ChT (4.0x10^{-8} M) after Mixing with Peptidoresorc[4]arenes **12**, **13**, **16**, and **17** in Comparison with the Activity of the Native Enzyme

Comp.	Config.	Receptor concentration	Residual activity	Dependence on time of inhibitory activity
(-)-12a	LL	4 x 10^{-5} μM	13 %	Time-independent
(+)-12b	DD	4 x 10^{-6} μM	43 %	Time-independent
		4 x 10^{-5} μM	11 %	
		1 x 10^{-4} μM	0 %	
(-)-13a	LL	4 x 10^{-5} μM	25 %	Time-independent
(+)-13b	DD	4 x 10^{-5} μM	17 %	Time-independent
(-)-17a	LL	4 x 10^{-6} μM	28 %	Time-independent
		4 x 10^{-5} μM	5 %	
(+)-17b	DD	4 x 10^{-5} μM	71→100 %	Decreases by the time

(+)-10b	DD	4×10^{-6} µM	68 %	Time-independent
		4×10^{-5} µM	5 %	
(-)-16a	LL	4×10^{-7} µM	60 %	Time-independent
		4×10^{-6} µM	5 %	
(+)-16b	DD	4×10^{-6} µM	13 →6 %	Increases by the time

Peptidoresorc[4]arene (+)-12b.

To shed light on the mechanism of inhibition exerted by the resorc[4]arenic inhibitors toward ChT, we focused our attention on the most effective inhibitor (+)-**12b**. Dose-dependency experiments were performed by using the inhibitor at concentrations ranging from 4.0×10^{-6} to 1.0×10^{-4} M with ChT maintained at a fixed concentration (4.0×10^{-8} M) for a 90 min preincubation period.

The dose-responsive inhibition of ChT by (+)-**12b** at pH 7.4 in 0.005 M sodium phosphate buffer (Figure 4) showed that, in the presence of (+)-**12b** at a 4.0×10^{-6} M concentration, the activity of ChT decreased to 43%. Further increase in the concentration of (+)-**12b** (4.0×10^{-5} M) resulted in a decrease of the initial velocity to 11% (each value represents an average of three independent sets of experiments, SD < 3%).

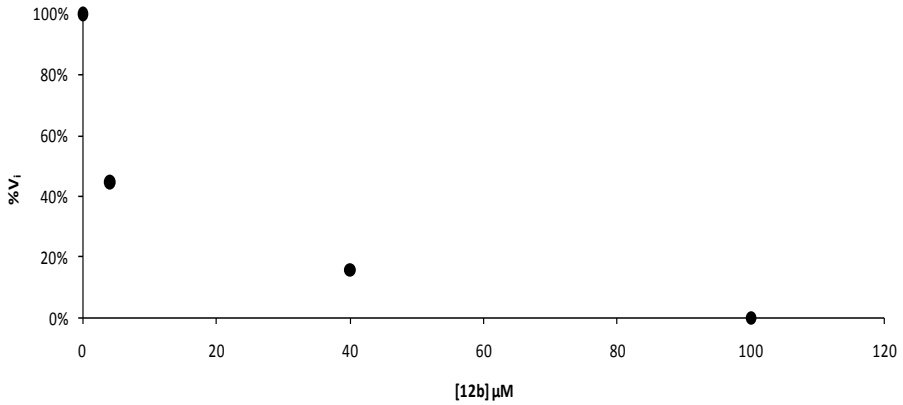


Figure 2.5. Dose-Responsive Inhibition Plot of ChT by (+)-**12b** at pH 7.4 in 0.005 M sodium phosphate buffer.

Nonlinear regression analysis to fit the Michaelis-Menten plot (Figure **2.6**) yielded the apparent maximum enzymatic reaction rate ($V_{max} = 7.3 \times 10^{-3}$ abs sec^{-1}) and the Michaelis-Menten constant (K_m = 96.9 ± 7.2 µM), whereas

Dixon plots[52a] (Figure **2.7**, bottom) were used to measure the inhibition constant (K_i = 12.4 ± 5.1 μM). The kinetic studies pointed out that receptor (+)-**12b** behaves as a noncompetitive inhibitor of ChT because it decreases the V_{max} values, with minor effects on the K_m constant.[52b] On these bases, it can be envisaged that (+)-**12b** does not compete directly with the substrate molecules, though it may affect the size and shape of the catalytic cleft.

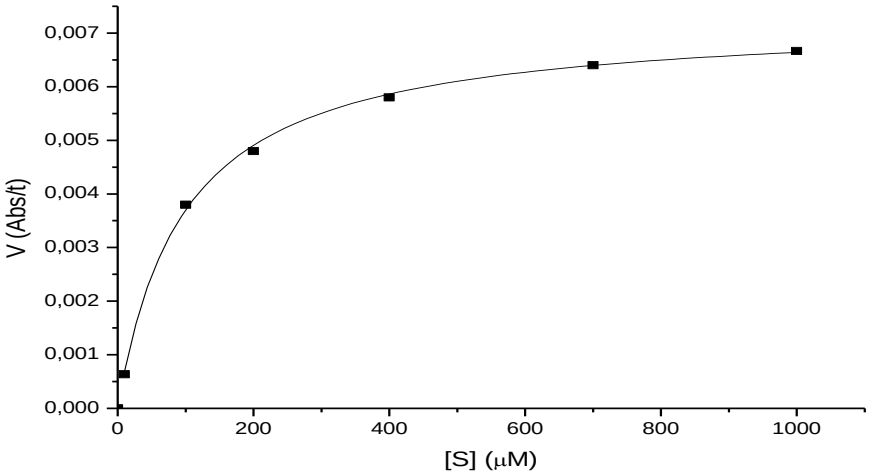


Figure 2.6: nonlinear regression analysis to fit the Michaelis-Menten data in the absence of inhibitor (Km = 96.9±7.2 μM).

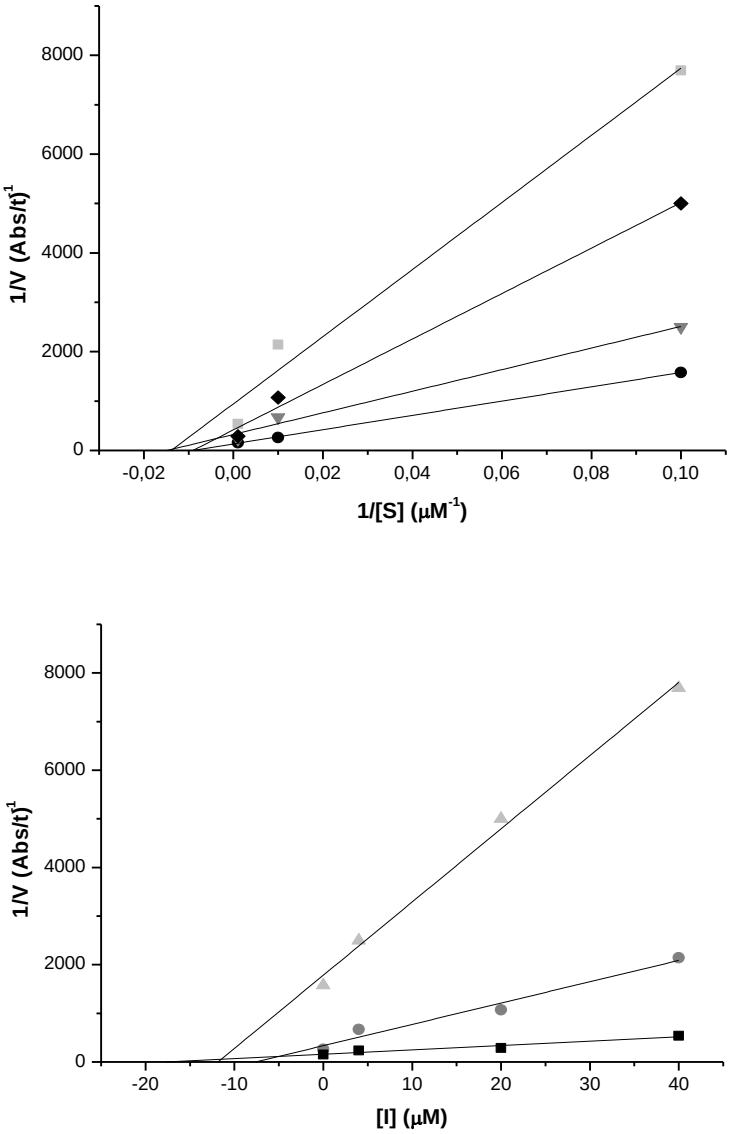


Figure 2.7: Noncompetitive inhibition of ChT (4.0 x 10^{-8} M) by receptor (+)-**12b**. **Top** Lineweaver-Burk plot at three fixed Suc-AAPF-pNA concentrations in the absence (●) and presence of 4 μM(▼), 20 μM(◆), and 40 μM(■) of (+)-**12b**. **Bottom**: Dixon plot at three fixed Suc-AAPF-pNA concentrations, namely (▲) 10, (●) 100, and (■) 1000 μM. K_i = 12.4 ± 5.1 μM

To determine some possible binding areas of (+)-**12b** on the surface of ChT, a computational docking experiment was carried out using the program AutoDock 4.0.[53] The computationsl analysis was performed in collaboration with Prof. Andrea Tafi, University of Siena. The calculations point to a complex of the type shown in Figure **2.8**, where the carboxylate groups in (+)-**12b** make electrostatic interactions with several basic residues (Lys-36, Lys-90, Lys-175, and Arg-145) surrounding the active site cleft of ChT. Some of these amino acids were already found to be involved in ChT complexes with calix- [4]arene receptors.[36b,d]

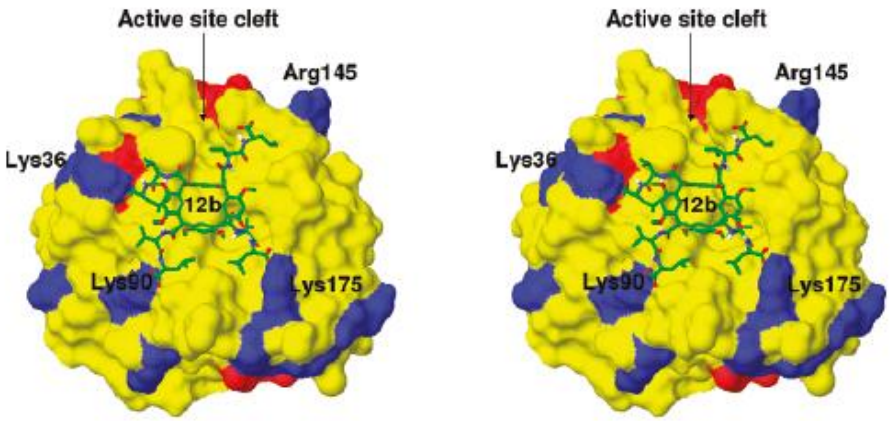


Figure 2.8. Calculated structure (crossed stereoview by AutoDock 4.0) for the interaction of (+)-**12b** with ChT. The protein structure is displayed as a Connolly surface, while the 3D structure of the docked ligand is explicitly shown colored by atom type. Uncharged amino acids are colored in yellow, negatively charged in red and positively charged in blue. Few relevant basic residues are tagged. The active site cleft entrance is marked by a black arrow.

Peptidoresorc[4]arenes (+)-10b.

To get information on some other possible interactions between peptidoresorc[4]arenes and ChT, the investigation was extended to the available benzyl ester (+)-**10b**, which was the synthetic precursor of the analyzed free acid. The inhibition experiments, carried out under the same conditions as for (+)-**12b** (see Table **2.5**), resulted in only 5% residual activity of the enzyme after incubation with (+)-**10b** at a concentration of 4.0×10^{-5} M, increasing to 68% with the macrocycle at 4.0×10^{-6} M (time-independent inhibition). A different type of interaction, actually hydrophobic, by peptidoresorc[4]arenes was thus shown to be possible for the ChT inhibition

(vide infra), since free acid (+)-**12b** was a much weaker inhibitor of ChT than its benzyl ester homochiral precursor (+)-**10b.**

Peptidoresorc[4]arenes 17a,b.

The receptors pair **17a**/**17b** features a change in the nature of the dipeptide chain (valylaspartic acid instead of valyl-leucine, as compared with **12a**/**12b**), with 2-fold increase of ionizable carboxylic groups and, as expected, an increased amount of electrostatic interactions.

However, only the LL-derivative (-)-**17a** exhibited enhanced inhibition as compared to the homochiral (-)-**12a** (see Table **2.5**) at the same concentration: in fact, only 5% of residual activity of ChT was found at a concentration of 4.0×10^{-5} M. The inhibition resulted to be time-independent and was still significant (28% of residual ChT activity) at a receptor concentration of 4×10^{-6} M. Surprisingly, in the presence of the DD-derivative (+)-**17b**, ChT exhibited a 71% of residual activity after 90 min.

The inhibitory activity of (+)-**17b** was shown to be time-dependent (complete recovery, i.e. 100% of residual activity of ChT, was observed after 180 min).

As in the case of (+)-**12b**, for comparative purposes we tested the inhibition activity of benzyl esters (-)-**16a** and (+)-**16b**, which were the synthetic precursors of receptors **17a**,**b**: in these cases, because the benzyl esters were slightly soluble in phosphate buffer, concentrations higher than 4×10^{-6} M were not allowed, and the cosolvent (2-methoxyethanol) was replaced by dimethyl sulfoxide (DMSO) for the preparation of stock solutions of receptors. The final concentration of DMSO (2.4×10^{-2} M) was indeed much lower than that needed for influencing the enzyme activity.[54] In the presence of receptor (+)-**16b**, the residual activity of ChT decreased by the time to 13% after 90 min to reach a minimum (6%) after 120 min. The time-dependent inhibition supports the slow formation of a network of weak hydrophobic interactions.

Moreover, the LL-enantiomer (-)-**16a** was shown to be the most potent inhibitor of ChT of all the family of peptidoresorc[4]arenes: 5% of residual activity of the enzyme was observed in the presence of (-)-**16a** at a concentration of 4.0×10^{-6} M, and 60% at 4.0×10^{-7} M. The dose-responsive inhibition plots of ChT by (-)-**16a** is reported in Figure **2.9**. Notably, the inhibition was immediate and time-independent. The fact that LL-configuration of dipeptides is coincident with the stereochemistry of the

natural substrates of the enzyme is not enough to explain the unexpected results for interactions based exclusively on hydrophobicity.

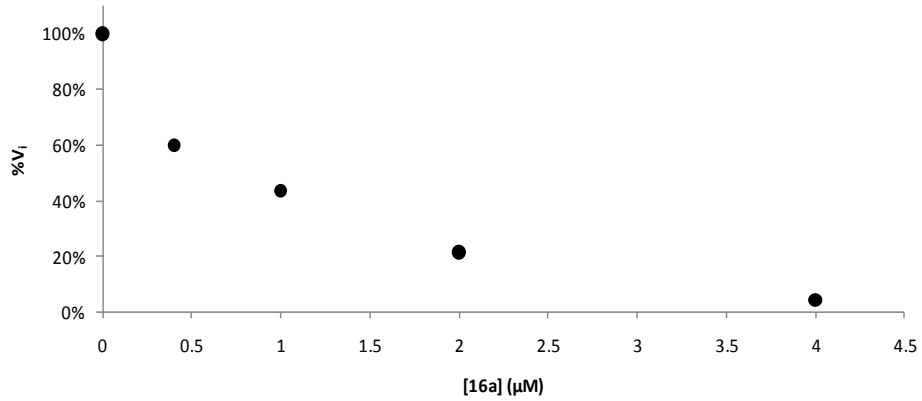


Figure 2.9. Dose-Responsive Inhibition Plot of ChT by (−)-**16a** at pH 7.4 in 0.005 M sodium phosphate buffer.

Peptidoresorc[4]arene (-)-16a.

To get more information on the unexpected strong inhibition of benzyl ester (-)-**16a** toward ChT, we performed the same set of experiments as for free acid (+)-**12b** (vide supra). Receptor (-)-**16a** was thus treated at concentrations ranging from 4.0 x 10^{-7} to 4.0 x 10^{-6} M with ChT maintained at a fixed concentration (4.0 x 10^{-8} M) for a 90 min preincubation period. Kinetic analysis, based on Lineweaver- Burk (Figure 9, left) and Dixon (Figure **2.10**, right) plots, pointed out that receptor (-)-**16a** is a noncompetitive inhibitor of ChT (K_i = 0.76 ± 0.14 μM) because it decreases the Vmax values without affecting the K_m constant.[52] However, receptor (+)-**12b**, featuring more complementary recognition surfaces to ChT than receptor (-)-**16a**, showed a significantly lower inhibition potency (K_i = 12.4 ± 5.1 μM). To explain such strong interaction between benzyl ester (-)-**16a** and ChT we may invoke the chemical similarity between the terminal benzyl groups and the benzyl side chain of Phe; in fact, examination of the specificity requirements for ChT reveals that a critical interaction for efficient binding of a substrate to this enzyme includes a van der Waals interaction of the benzyl side chain of Phe with a hydrophobic pocket adjacent to the active site of the enzyme.[34]

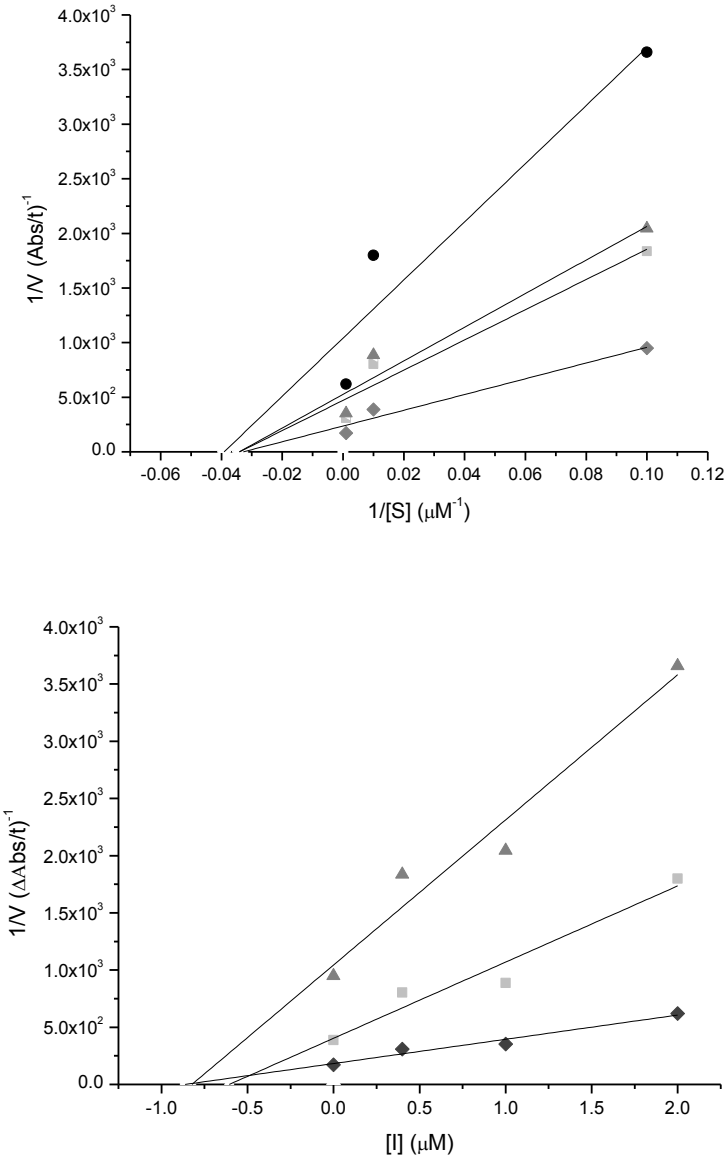


Figure 2.10. Noncompetitive inhibition of ChT (4.0 x 10^{-8} M) by receptor (-)-**16a**. Left: Lineweaver-Burk plot at three fixed Suc-AAPF-pNA concentrations in the absence (◆) and presence of 0.4 μM (■), 1 μM (▲), and 2 μM (●) of (-)-**16a**. Right: Dixon plot at three fixed Suc-AAPF-pNA concentrations, (▲) 10 μM, (■) 100 μM, and (◆) 1000 μM. Ki = 0.76 ± 0.14 μM.

Effect of Medium Ionic Strength on the Protein-Receptor Complexation.
Due to the noncovalent electrostatic nature of binding, the interaction between ChT and receptor (+)-**12b** was supposed to be reversible. Therefore, we investigated the possible influence of the increase of the ionic strength in the medium both on the formation of the complex not yet established, and on the stability of the same complex once it is formed.
For the effect on the complex formation, ChT (4.0×10^{-8} M) was incubated with inhibitor (+)-**12b** (4.0×10^{-5} M) in the presence of sodium chloride (NaCl) at five different concentrations ranging from 1.0×10^{-4} to 1.0×10^{-1} M, as already reported.[33e,34b] The samples were incubated at room temperature for 90 min, a time sufficient for an almost complete (∼90%) inhibition, in the absence of NaCl. The enzyme activity was monitored as above-described and compared with the activity of a control solution containing only ChT and NaCl under the same conditions (system named before mixing, BM). The normalized residual activities of each sample and the corresponding NaCl concentration are summarized in Table **2.6**.
To evaluate the effect of ionic strength on the stability of the complex ChT/(+)-**12b**, NaCl at the above-mentioned concentrations was added to the preformed ChT/(+)-**12b** complex, after 90 min incubation period. As control condition, a solution containing pure ChT was used. Both mixtures were analyzed after further 90 min of incubation (system named after mixing, AM). The results, i.e., the residual activities of the enzyme and the corresponding NaCl concentrations, are reported in Table **2.6**.
As expected, the residual activity of ChT increases with the increase of NaCl concentration (BM mode): at NaCl concentrations between 1.0×10^{-4} and 1.0×10^{-2} M (see Table **2.6**), the ionic strength is not strong enough to break completely the electrostatic attraction between enzyme and inhibitor, and the residual activity undergoes a small increase (23% and 36% of the control). By contrast, the ChT residual activity increases sharply in the presence of 5.0 x 10-2 M NaCl (95% of the control) and is almost completely restored (∼ 98%) for NaCl 1.0 x 10-1 M. These findings confirm that the binding of inhibitor (+)-**12b** to the enzyme is mostly based on electrostatic complementarities and, as such, is reversible and does not denature the native conformation of ChT.
When we performed the same set of experiments on receptor (-)-**17a**, which features a 2-fold number of ionizable carboxylic groups for the replacement of leucine with an aspartate residue, we observed that the formation of the complex ChT-receptor seemed to be mitigated only partially by the increase

of the ionic strength: actually, the residual activity of the enzyme is still 50% in the presence of NaCl 1.0 x 10^{-1} M (Table **2.6**, system BM).
This result may be attributed to the higher number of ionizable functional groups, and thus to an increased number of electrostatic interactions to be broken.
The AM mode experiments, aimed at investigating the stability of the complex ChT-receptor under the influence of the medium ionic strength, substantially confirmed the mainly electrostatic nature of the interaction between ChT and receptors (+)-**12b** and (-)-**17a** (the first complex is slightly more stable). Also in this case, the investigation on (+)-**12b** was extended to the available benzyl ester (+)-**10b**, which was its synthetic precursor: the enzyme restored only partially its activity, which was 51% in the presence of NaCl 1.0 x 10^{-1} M (Table **2.6**, system BM), denoting a low importance of ionic strength on the establishment of the lipophilic interaction between ChT and benzyl ester (+)-**10b**. As a confirmation, the stability of the complex was shown to be slightly influenced by the ionic strength as well, since, at the same NaCl concentration, ChT showed only 15% residual activity (AM mode).
With regard to receptor (+)-**16b**, which was supposed to slowly form a network of hydrophobic interactions with ChT (see Table **2.5**), the experiments in the presence of increasing concentrations of NaCl confirmed such hypothesis (see Table **2.6**). In fact, the reduced influence of the ionic strength on the formation (BM mode) and on the stability (AM mode) of the ChT-(+)-**16b** complex is highlighted by the fact that residual ChT activity was 32% and 40%, respectively, in the presence of NaCl 1.0 x 10^{-1} M .
Notably, the LL-enantiomer (-)-**16a** was confirmed to be the most effective inhibitor of ChT activity, since the inhibition was not significantly affected by the ionic strength (Table **2.6**), both in BM and in AM modes.

Table 2.6. Residual Activity (%) of ChT (4.0 X 10^{-8} M) in the Presence of NaCl before Mixing (System BM) and after Mixing (System AM) with Peptidoresorc[4]arenes (+)-**12b**, (-)-**17a**, (+)-**10b**, (+)-**16b**, and (-)-**16a**, in Comparison with the Activity of the Native Enzyme

		NaCl					
compd[a]	System	0 M	1.0 x 10^{-4} M	1.0 x 10^{-3} M	1.0 x 10^{-2} M	5.0 x 10^{-2} M	1.0 x 10^{-1} M
(+)-**12b**	BM	11	23	36	36	95	98
	AM	8	10	10	20	62	92
(-)-**17a**	BM	5	3	5	5	17	50
	AM	3	4	3	10	64	100
(+)-**10b**	BM	5				9	51
	AM	6				11	15
(+)-**16b**	BM	13	13	13	12	17	32
	AM	13	16	13	14	22	40
(-)-**16a**	BM	4	5	4	3	2	2
	AM	1	1	1	2	2	2

3. Conclusions

Our efforts into the design of novel N-linked peptidoresorc-[4]arenes with appropriately modified functional groups proved that they are able to interact with proteins, such as HSA, and acquire inhibition properties vs enzymes such as ChT. For compounds with terminal carboxylate functions (**12**, **13**, and **17**), the ChT inhibition is based essentially on electrostatic interaction; the bound resorcarene can be released from the enzyme surface by increasing the ionic strength, allowing the enzyme activity to be almost completely restored. Differently from **12**, **13**, and **17**, hydrophobic network can be suggested for receptors with terminal benzyl ester groups (**10** and **16**).

The inhibitors were not modified by ChT, as shown by TLC analysis ($CHCl_3$/MeOH mixtures) of mixtures of peptidoresorc[4]arenes and enzyme incubated for 3 h, after complex denaturation.

A computational docking experiment carried out on (+)-**12b** suggested a complex where the carboxylate groups make electrostatic interactions with several basic residues (Lys-36, Lys-90, Lys-175, and Arg-145) surrounding the active site cleft of ChT. In particular, receptor (+)-**12b** was shown to be a noncompetitive inhibitor of ChT (K_i = 12.4 (5.1 μM) because it decreases the V_{max} values without affecting the K_m. Within the same framework, the benzyl ester derivatives seem to be more effective: for example, free acid (+)-**12b** is a much weaker inhibitor of ChT than its benzyl ester homochiral precursor **(+)-10b**, as judged by the residual enzymatic activity after a 90 min incubation period. The same observation can be made for derivatives **17a**,**b** and their precursors **16a**,**b**. In the cases of receptors **12** and **13**, compounds with D,D-configuration were more active than the L,L-derivatives. However, for all the other cases, the most active compounds featured the L,L-configuration, which is the sterical arrangement of the natural ligand. This is mainly true for the **17a**/**17b** pair: the arrangement of the eight carboxylate groups, responsible for electrostatic interaction with the enzyme, seems to be significantly affected by the D,D-configuration of the dipeptide chains (71% of residual enzymatic activity vs 5%). This trend was observed on the benzyl ester homochiral precursor as well, where the LL-enantiomer (-)-**16a** was found to be the most effective noncompetitive inhibitor of ChT activity,

with a K_i = 0.76 ± 0.14 μM. In this case, inhibition was not significantly affected by the ionic strength, in both the BM and AM modes.

Focusing on the nature of the dipeptide chains, we noticed that the terminal amino acids of peptidoresor[4]arenes **12**, **13** and **17** could be placed in the following scale of inhibition activity: aspartic acid (receptor **17**) > leucine (receptor **12**) > valine (receptor **13**), somehow corresponding to the 2-fold number of interaction sites (for aspartic acid) and to the affinity of ChT for the amino acids (for leucine), as a substrate.[51] Similarly, to explain the strong hydrophobic interactions between ChT and peptidoresor[4]arenes benzyl esters **10** or **16**, we may invoke the chemical similarity between the terminal benzyl groups and the benzyl side chain of phenylalanine, one of the main substrates of ChT.[51] In fact, a van der Waals interaction of the Phe benzyl side chain with a hydrophobic pocket adjacent to the active site is one of the specificity requirements for ChT.

4. Experimental section

Synthesis of Resorc[4]arene Octamethyl Ethers 1-3. Compounds 1-3 were synthesized as previously described.[47a]

Preparation of *N*-Boc-D-aspartic Acid. NaOH 1 N (45 mL) and di-tert-butyl dicarbonate (2.5 g, 11.5 mmol) were added to an ice-cooled solution of D-aspartic acid in 1,4-dioxane (1.5 g, 11.3 mmol in 20 mL). The mixture was left at room temperature for 3 h, acidified to pH 1-2 with HCl 1 N, and extracted with ethyl acetate (5 x 80 mL). The pooled organic extract, after drying over Na_2SO_4, gave *N*-Boc-D-aspartic acid with a 63% yield.
$C_9H_{15}NO_6$; Mol. Wt.: 233.21. ^{1}H NMR ($CDCl_3$): δ 5.64 (d, 1H, NH), 4.65 (m, 1H, CHN), 3.16, 2.90 (dd, 1H each, CH_2), 1.47 (s, 9H, *t*-Bu). ^{13}C NMR ($CDCl_3$): δ 177.3 (C=O side chain), 174.9 (COOH), 156.0 (COO*t*-Bu), 79.5 ($C(CH_3)_3$), 52.0 (CHN), 39.1 (CH_2), 28.5 (3 x CH_3).

General Procedure for the Preparation of Benzyl Esters 5b, 8b, and 14b as TFA Salts. Cs_2CO_3 (32.6 g, 100 mmol) was added to an ice-cooled solution of the proper *N*-Boc-protected D-amino acid (100 mmol) in DMF (200 mL). The mixture was stirred at 0 °C for 45-60 min. Benzyl bromide (11.9 mL, 17.1 g, 100 mmol) was added to the reaction mixture, which was maintained under magnetic stirring at 0 °C for 30 min and allowed to reach room temperature overnight. The reaction mixture was poured into H_2O (800 mL) and extracted with n-hexane (4 x 250 mL). The combined extracts were washed with H_2O (200 mL) and then with saturated NaCl solution, finally dried over Na_2SO_4. Filtration, concentration, and drying under vacuum gave *N*-Boc-D-leucine benzyl ester (95%), *N*-Boc-D-valine benzyl ester (95%) and *N*-Boc-D-aspartic acid dibenzyl ester (86%) as colorless oils. Benzyl esters (1 mmol) were deprotected by treatment with TFA-DCM (1:1, 25 mL). The reaction mixture was stirred at 25 °C for 45 min and then evaporated to dryness; the addition of a suitable hexane/Et_2O mixture (50 mL) under vigorous stirring yielded white solids that were filtered, washed several times with hexane/Et_2O, and dried under vacuum to afford the pure TFA salts **5b**, **8b**, and **14b** in quantitative yields.

D-Leucine Benzyl Ester, TFA Salt (5b). $C_{15}H_{20}F_3NO_4$; Mol. Wt.: 335.32. ^{1}H NMR ($CDCl_3$): δ 8.19 (br s, 3H, NH_3^+), 7.37 (br s, 5H, C_6H_5), 5.21, 5.16 (d, 1H each, OCH_2), 4.03 (m, 1H, CHN), 1.71 (m, 3H, CH_2CH), 0.93 (2 × d, 6H, 2 × CH_3). ^{13}C NMR ($CDCl_3$): δ 171.6 (−COOBz), 141.2 (Ar−C), 129.0 (Ar−CH), 127.7 (Ar−CH), 127.2 (Ar−CH), 68.5 (CH_2−Ar), 50.9 (CHN), 43.3 (CH_2), 24.2 (CH), 22.2 (CH_3), 22.0 (CH_3). The signals due to the TFA molecule are not reported for simplicity.

D-Valine Benzyl Ester, TFA Salt (8b). $C_{14}H_{18}F_3NO_4$; Mol. Wt.: 321.29. ^{1}H NMR ($CDCl_3$): δ 8.34 (br s, 3H, NH_3^+), 7.41 (br s, 5H, C_6H_5), 5.36, 5.22 (d, 1H each, OCH_2), 4.15 (m, 1H, CHN), 2.43 (m, 1H, CH), 1.10 (d, 6H, 2 × CH_3). ^{13}C NMR ($CDCl_3$): δ 171.6 (−COOBz), 141.2 (Ar−C), 129.0 (Ar−CH), 127.7 (Ar−CH), 127.2 (Ar−CH), 67.2 (CH_2−Ar), 51.9 (CHN), 30.4 (CH), 18.7 (CH_3), 17.9 (CH_3). The signals due to the TFA molecule are not reported for simplicity.

D-Aspartic Acid Dibenzyl Ester, TFA Salt (14b). $C_{20}H_{20}F_3NO_6$; Mol. Wt.: 427.37. ^{1}H-NMR ($CDCl_3$): δ 8.79 (br s, 3H, NH_3^+), 7.36 (br s, 10H, 2 × C_6H_5), 5.22, 5.16 (d, 1H each, OCH_2), 5.11 (br s, 2H, OCH_2), 4.56 (m, 1H, CHN), 3.26 (m, 2H, CH_2). ^{13}C NMR ($CDCl_3$): δ 173.1 (−COOBz side chain), 171.6 (−COOBz), 141.2 (Ar−C), 129.0 (Ar−CH), 127.7 (Ar−CH), 127.2 (Ar−CH), 68.5 (CH_2−Ar), 67.2 (CH_2−Ar), 50.2 (CHN), 40.0 (CH_2). The signals due to the TFA molecule are not reported for simplicity.

General Procedure for the Preparation of Dipeptide Benzyl Esters 6a,b, 9a,b, and 15a,b as TFA Salts. The proper amino acid benzyl ester as TFA or p-toluensulfonate salt (5.5 mmol), HOBT (0.75 g, 5.5 mmol), Et_3N (0.96 mL, 7.1 mmol), and the proper *N*-Bocprotected amino acid (5.5 mmol) were dissolved in DCM (250 mL). DCC (1.14 g, 5.5 mmol) was added to the solution cooled in an ice bath. After 4 h at room temperature, the solvent was removed under vacuum, and the residue was purified by silica gel column chromatography (hexane-ethyl acetate, 7.5:2.5) to yield the N-Boc-protected dipeptide benzyl esters as white powders (yields 80-85%). Dipeptides (0.9 mmol) were deprotected by treatment with TFA-DCM (1:1, 25 mL) as described previously to afford the pure TFA salts **6a**,**b**, **9a**,**b**, and **15a**,**b** in quantitative yields.

L-Valyl-L-Leucine Benzyl Ester, TFA Salt (6a). $C_{20}H_{29}F_3N_2O_5$; Mol. Wt.: 434.45. Mp: 143−145 °C. ^{1}H NMR ($CDCl_3$): δ 7.81 (br d, 1H, NH), 7.32 (br s, 5H, C_6H_5), 5.18, 5.09 (d, 1H each, OCH_2), 4.43 (m, 1H, CHN; leucine), 3.88 (br d, 1H, CHN valine), 2.11 (m, 1H, CH valine), 1.60 (m, 2H, CH_2 leucine), 1.52 (m, 1H, CH leucine), 0.97, 0.94, 0.86, 0.85 (d, 3H each, 4 × CH_3). ^{13}C NMR ($CDCl_3$): δ 171.8 (−COOBz), 168.9 (−CONH−), 135.2 (Ar−C), 128.6 (Ar−CH), 128.5 (Ar−CH), 128.4 (Ar−CH), 67.2 (CH_2−Ar), 58.6 (CHN, valine), 51.9 (CHN, leucine), 40.1 (CH_2, leucine), 30.4 (CH, valine), 24.6 (CH, leucine), 22.4 (CH_3), 21.6 (CH_3), 17.9 (CH_3), 17.8 (CH_3). The signals due to the TFA molecule are not reported for simplicity. ESI-MS (pos.): *m/z* found 320.5 ([*M* + H]$^+$), $C_{18}H_{29}N_2O_3$ (free base) requires 321.2. Anal. Calcd for $C_{20}H_{29}F_3N_2O_5$: C, 55.29; H, 6.73; F, 13.12; N, 6.45; O, 18.41. Found: C 55.18; H 6.71; N 6.43. $[\alpha]^{20}_D$ −18.0 (*c* 1.0, $CHCl_3$).

D-Valyl-D-Leucine Benzyl Ester, TFA Salt (6b). ^{1}H NMR, ^{13}C NMR, elemental analysis, and ESI-MS data are coincident with those reported for **6a**. $[\alpha]^{20}_D$ +17.8 (*c* 1.0, $CHCl_3$).

L-Leucyl-L-Valine Benzyl Ester, TFA Salt (9a). $C_{20}H_{29}F_3N_2O_5$; Mol. Wt.: 434.45. Mp: 160−162 °C. 1H NMR ($CDCl_3$): δ 7.51 (br d, 1H, NH), 7.34 (br s, 5H, C_6H_5), 7.27 (br d, 1H, NH), 5.18, 5.11 (d, 1H each, OCH_2), 4.38 (br s, 1H, CHN leucine), 4.14 (br s, 1H, CHN valine), 2.18 (m, 1H, CH valine), 1.64 (m, 2H, CH_2 leucine), 1.52 (m, 1H, CH leucine), 0.91 (d, 6H, 2 × CH3), 0.87 (d, 6H, 2 × CH_3). 13C NMR ($CDCl_3$): δ 171.0 (−COOBz), 168.9 (−CONH−), 135.2 (Ar−C), 128.6 (Ar−CH), 128.6 (Ar−CH), 128.5 (Ar−CH), 67.2 (CH_2−Ar), 58.7 (CHN, valine), 52.2 (CHN leucine), 40.7 (CH_2 leucine), 30.5 (CH valine), 24.2 (CH leucine), 22.2 (CH_3), 22.0 (CH_3), 18.7 (CH_3), 17.9 (CH_3). The signals due to the TFA molecule are not reported for simplicity. ESI-MS (pos.): *m/z* found 320.8 ([*M* + H]$^+$), $C_{18}H_{29}N_2O_3$ (free base) requires 321.2. Anal. Calcd for $C_{20}H_{29}F_3N_2O_5$: C, 55.29; H, 6.73; F, 13.12; N, 6.45; O, 18.41. Found: C 55.40; H 6.74; N 6.46. $[\alpha]^{20}_D$ +1.7 (*c* 1.0, $CHCl_3$).

D-Leucyl-D-Valine Benzyl Ester, TFA Salt (9b). ^{1}H NMR, ^{13}C NMR, elemental analysis, and ESI-MS data are coincident with those reported for **9a**. $[\alpha]^{20}_D$ −1.7 (*c* 1.0, $CHCl_3$).

L-Valyl-L-Aspartic Acid Benzyl Ester, TFA Salt (15a). $C_{25}H_{29}F_3N_2O_7$; Mol. Wt.: 526.50217. Mp: 141−142 °C. ^{1}H NMR (MeOH-*d*4): δ 7.33 (br s,

10H, 2 × C_6H_5), 5.12, 5.03 (d, 2H each, 2 × OCH_2), 4.95 (q, 1H, CHN; aspartic acid), 3.67 (t, 1H, CHN; valine), 2.96 (m, 2H, CH_2CO), 2.15 (m, 1H, CH; valine), 0.97, 0.87 (br s, 3H each, 2 × CH_3). ^{13}C NMR (MeOH-d_4): δ 170.2 (−COOBz), 170.0 (−COOBz), 168.1 (−CONH−), 135.8 (Ar−C), 135.4 (Ar−C), 128.2 (Ar−CH), 128.1 (Ar−CH), 128.0 (Ar−CH), 67.2 (CH_2−Ar), 66.5 (CH_2−Ar), 58.2 (CHN valine), 49.0 (CHN aspartic acid), 35.4 (CH_2 aspartic acid), 30.2 (CH valine), 17.4 (CH_3), 16.3 (CH_3). The signals due to the TFA molecule are not reported for simplicity. ESI-MS (pos.): *m/z* found 412.9 ([*M* + H]+), $C_{23}H_{29}N_2O_5$ (free base) requires 413.2. Anal. Calcd for $C_{25}H_{29}F_3N_2O_7$: C, 57.03; H, 5.55; F, 10.83; N, 5.32; O, 21.27. Found: C 56.91; H 5.54; N 5.31. $[\alpha]^{20}_D$ +6.0 (*c* 1.0, MeOH).

D-Valyl-D-Aspartic Acid Benzyl Ester, TFA Salt (15b). ^{1}H NMR, ^{13}C NMR, elemental analysis, and ESI-MS data are coincident with those reported for **15a**. $[\alpha]^{20}_D$ −6.3 (*c* 1.0, MeOH).

2,8,14,20-Tetrakis(L-valyl-L-leucinamido)resorc[4]arene Benzyl Ester (10a). DIPEA (0.42 mL, 2.5 mmol) was added under nitrogen to a dry THF solution of tetracarboxylic acid chloride **3** (150 mg, 0.15 mmol in 45 mL). The mixture was stirred for 40 min at room temperature, and a dry THF solution of **6a** (322 mg, 0.9 mmol in 30 mL) was added dropwise in a 1 h period. The reaction mixture was held at reflux under nitrogen for 1 h. Evaporation of the solvent and purification by column chromatography ($CHCl_3$/MeOH mixtures) gave peptidoresorc[4]arene (-)-**10a** as a pale yellow powder (70% overall yield). Mp: 220-221 °C. ^{1}H and ^{13}C NMR signals are reported in Tables **2.2** and 3. ESI-HRMS (pos): m/z found 1043.53604 ([M + 2Na]$^{2+}$), $C_{116}H_{152}O_{24}N_8Na_2$ requires 1043.53520 (monoisotopic mass). $[\alpha]^{20}_D$: -80.5 (c 1.0, $CHCl_3$).

2,8,14,20-Tetrakis(D-valyl-D-leucinamido)resorc[4]arene Benzyl Ester (10b). (+)-**10b** was obtained as described for (-)-**10a**, starting from tetracarboxylic acid chloride **3** and dipeptide benzyl ester **6b**. Pale yellow powder. 65% overall yield. ^{1}H NMR, ^{13}C NMR, and ESI-HRMS data are coincident with those reported for (-)-**10a**. $[\alpha]^{20}_D$: +79.4 (c 1.0, $CHCl_3$).

2,8,14,20-Tetrakis(L-leucyl-L-valinamido)resorc[4]arene Benzyl Ester (11a). (-)-**11a** was obtained as described for (-)-**10a**, starting from tetracarboxylic acid chloride **3** and dipeptide benzyl ester **9a**. Pale yellow

powder. 75% overall yield. Mp: 248-249 °C. ^{1}H and ^{13}C NMR signals are reported in Tables **2.2** and 3. ESI-HRMS (pos): m/z found 1043.53549 ([M + 2Na]$^{2+}$), $C_{116}H_{152}O_{24}N_8Na_2$ requires 1043.53520 (monoisotopic mass). $[\alpha]^{20}_{D}$: -76.7 (c 0.3, $CHCl_3$).

2,8,14,20-Tetrakis(D-leucyl-D-valinamido)resorc[4]arene Benzyl Ester (11b). (+)-**11b** was obtained as described for (-)-**10a**, starting from tetracarboxylic acid chloride **3** and dipeptide benzyl ester 9b. Pale yellow. 65% overall yield. ^{1}H NMR, ^{13}C NMR, and ESI-HRMS data are coincident with those reported for (-)-11a. $[\alpha]^{20}_{D}$: +75.9 (c 1.0, $CHCl_3$).

2,8,14,20-Tetrakis(L-valyl-L-aspartic acid amido)resorc[4]arene Benzyl Ester (16a). (-)-**16a** was obtained as described for (-)-**10a**, starting from tetracarboxylic acid chloride **3** and dipeptide benzyl ester **15a**. Pale yellow powder. 54% overall yield. Mp: 218-219 °C. ^{1}H and ^{13}C NMR signals are reported in Table **2.4**. ESI-HRMS (pos): m/z found 1228.01773 ([M + 2Na]$^{2+}$), $C_{136}H_{152}O_{32}N_8Na_2$ requires 1228.01647 (monoisotopic mass). $[\alpha]^{20}_{D}$: -35.7 (c 0.5, MeOH).

2,8,14,20-Tetrakis(D-valyl-D-aspartic acid amido)resorc[4]arene Benzyl Ester (16b). (+)-**16b** was obtained as described for (-)-**10a**, starting from tetracarboxylic acid chloride **3** and dipeptide benzyl ester **15b**. Pale yellow powder. 56% overall yield. ^{1}H NMR, ^{13}C NMR, and ESI-HRMS data are coincident with those reported for (-)-**16a**. $[\alpha]^{20}_{D}$: +35.9 (c 0.5, MeOH).

General Procedure for Pd/C Hydrogenation of Resorc[4]arene Benzyl Esters to Resorc[4]arene Free Acids 12a,b, 13a,b, and 17a,b. After two vacuum/nitrogen cycles to replace air inside the reaction tube, a mixture of the proper resorc[4]arene benzyl ester (0.05 mmol) and 10% Pd/C (70 mg) in absolute EtOH (4 mL) and dry THF (4 mL) was vigorously stirred at room temperature under 1 atm of hydrogen for 24 h. The reaction mixture was filtered using a membrane filter (Millipore, Millex-LH, 0.45 μm), and the filtrate was concentrated to provide the product, which was dissolved twice in ethylacetate to remove residual EtOH. The ethyl acetate solution was finally dried under high vacuum at 25 °C, to give resorc[4]arene free acids **12a**,**b**, **13a**,**b**, and **17a**,**b** in quantitative yields.

2,8,14,20-Tetrakis(L-valyl-L-leucinamido)resorc[4]arenecarboxylic Acid (12a). (-)-**12a** was obtained by Pd/C hydrogenation of (-)-**10a**. Mp: 212-213 °C. ^{1}H and ^{13}C NMR signals are reported in Tables **2.2** and **2.3**. ESI-HRMS (neg): m/z found 566.62079 ([M - 4H + Na]$^{3-}$), $C_{88}H_{124}O_{24}N_8Na$ requires 566.62142 (monoisotopic mass). $[\alpha]^{20}_{D}$: -55.5 (c 1.0, MeOH). Enantiomeric excess (ee, %) of (-)-**12a**, checked by enantioselective HPLC, was in the 95-97% range.

2,8,14,20-Tetrakis(D-valyl-D-leucinamido)resorc[4]arenecarboxylic Acid (12b). (+)-**12b** was obtained by Pd/C hydrogenation of (+)-**10b**. ^{1}H NMR, ^{13}C NMR, and ESI-HRMS data are coincident with those reported for (-)-**12a**. $[\alpha]^{20}_{D}$: +52.5 (c 1.0, MeOH).

2,8,14,20-Tetrakis(L-leucyl-L-valinamido)resorc[4]arenecarboxylic Acid (13a). (-)-**13a** was obtained by Pd/C hydrogenation of (-)-**11a**. Mp: 135-136 °C. ^{1}H and ^{13}C NMR signals are reported in Tables **2.2** and **2.3**. ESI-HRMS (neg): m/z found 566.62026 ([M - 4H + Na]$^{3-}$), $C_{88}H_{124}O_{24}N_8Na$ requires 566.62142 (monoisotopic mass). $[\alpha]^{20}_{D}$: -49.2 (c 0.6, MeOH). Enantiomeric excess (ee, %) of (-)-**13a**, checked by enantioselective HPLC, was in the 95-97% range.

2,8,14,20-Tetrakis(D-leucyl-D-valinamido)resorc[4]arenecarboxylic Acid (13b). (+)-**13b** was obtained by Pd/C hydrogenation of (+)-**11b**. ^{1}H NMR, ^{13}C NMR, and ESI-HRMS data are coincident with those reported for (-)-**13a**. $[\alpha]^{20}_{D}$: +46.3 (c 0.7, MeOH).

2,8,14,20-Tetrakis(L-valyl-L-aspartic acid amido) resorc-[4]arenecarboxylic Acid (17a). (-)-**17a** was obtained by Pd/C hydrogenation of (-)-**16a**. Mp: 201-202 °C dec. ^{1}H and ^{13}C NMR signals are reported in Table **2.4**. ESI-HRMS (neg): m/z found 843.33097 ([M - 2H]$^{2-}$), $C_{80}H_{102}O_{32}N_8$ requires 843.33056 (monoisotopic mass). $[\alpha]^{20}_{D}$: -18.3 (c 0.1, MeOH).

2,8,14,20-Tetrakis(D-valyl-D-aspartic acid amido) resorc-[4]arenecarboxylic Acid (17b). (+)-**17b** was obtained by Pd/C hydrogenation of (+)-**16b**. ^{1}H NMR, ^{13}C NMR, and ESI-HRMS data are coincident with those reported for (-)-**17a**. $[\alpha]^{20}_{D}$: +17.7 (c 0.1, MeOH).

General Conditions for Enzymatic Activity Assay. All experiments were performed in 0.005 M sodium phosphate buffer (pH = 7.4). All solutions were kept at 25 °C during incubation periods and kinetic analysis. The desired concentrations of the different receptors were achieved by adding a proper quantity of stock solutions in 2-methoxyethanol (for **10**, **12** and **13**) or DMSO (for **16** and **17**) to 1 mL of the phosphate buffer solution of ChT (4.0×10^{-8} M). ChT and different concentrations of the receptors were kept at 25 °C for a fixed preincubation period, and then 20 µL of a stock solution of Suc-AAPFpNA in 2-methoxyethanol were added to reach a final substrate concentration of 1×10^{-4} M. Enzymatic hydrolysis was followed by monitoring p-nitroaniline formation at 405 nm for 1 min. The assays were performed in triplicate, and the averages are reported in the Tables and Figures (SD < 3%). The activity of the native ChT (control) was taken to be 100%, and from this value the residual enzymatic activity (%) was calculated.
Kinetic assay. ChT activity was measured using Suc-AAPF-pNA as substrate in the absence or presence of (+)-**12b** or (-)-**16a** inhibitors, under identical conditions as activity assay. The reaction mixture contained three fixed Suc-AAPF-pNAconcentrations (10 µM, 100 µM, and 1000 µM) with inhibitors concentrations equal to 0, 4, 20, and 40 µM for (+)-**12b** and 0, 0.4, 1, and 2 µM for (-)-**16a**.

Statistical Analysis. All assays were run in triplicate. Graphs were plotted using Microcal Origin software, version 6.0. Values of the slopes and intercepts were obtained by the linear regression analysis using the same software. K_i is given as the mean ± SD of three values calculated by using the Dixon plots at three different substrate concentrations. Values of V_{max} and K_m were recovered by nonlinear regression analysis based on the Michaelis-Menten equation ($R^2 = 0.99872$).

Docking Analysis. The input conformation for receptor (+)-**12b** was generated by a conformational search carried out with the Macro Model/Maestro software package (Schrödinger)[55] equipped with Amber* force field.[56] The search was performed on the rotatable bonds of the four dipeptide side chains of the molecule, which were moved in 3D space with the aim of locating a reliable local minimum energy geometry.[57]
In this work, the default settings of Monte Carlo Multiple Minimum search protocol were left unchanged. Analysis of possible binding areas of (+)-**12b**

on the surface of ChT was then carried out by using the program AutoDock 4.0.[53]

5. References

1. J.-M. Lehn, Supramolecular chemistry – scope and perspectives. Molecules, supermolecules, and molecular devices (Nobel lecture), *Angew. Chem. Int. Ed. Engl*. **1988**, 27, 90.

2. Kyba, E. P.; Helgeson, R. C.; Madan, K.; Gokel, G. W.; Tarnowski, T. L.; Moore, S. S.; Cram, D. J., "Host-guest complexation .1. Concept and illustration", *J. Am. Chem. Soc.* **1977**, 99, 2564–2571.

3. J. W. Steed and J. L. Atwood, *Supramolecular Chemistry, 2nd edition*;, John Wiley & Sons, Ltd (**2009**).

4. Behr, J. P., *The Lock and Key Principle. The State of the Art –100 Years on*, John Wiley & Sons, Inc.: New York, (**1994**).

5. Schneider H.-J., Yatmirsky A., *Principles and Methods in Supramolecular Chemistry*, Wiley, Chichester (**2000**)

6. Cram, D. J., *Angew. Chem., Int. Ed. Engl.* **1986**, 25, 1039–1134

7. Weber, P.C.; Wendoloski, J.J.; Pantoliano, M.W.; Salemme, F.R. *J. Am. Chem. Soc.* **1992**, 114, 3197;

8. Hegde, V.; Hung, C.-Y.; Madhukar, P.; Cunningham, R.; Höpfner, T.; Thummel, R.P. *J. Am. Chem. Soc.* **1993**, 115, 872.

9. Hosseini, M.W.; Lehn, J.-M. *Helv*. *Chim. Acta* **1987**, 70, 1312;

10. Lonergan, D.G.; Deslongchamps, G.; Tomás, S. Tetrahedron Lett. **1998**, 39, 7861.

11. Dougherty, D.A. *Science* **1996**, 271, 163.

12. (a) Lehn, J.-M.; Meric, R.; Vigneron, J.-P.; Cesario, M.; Guilhem, J.; Pascard, C.; Asfari, Z.; Vicens, J. *Supramol. Chem*. **1995**, 5, 97; (b) Schneider, H.-J.; Güttes, D.; Schneider, U. *J. Am. Chem. Soc.* **1988**, 110, 6449.

13. *Cation Binding by Macrocycles,* Inoue, Y.; Gokel, G.W. (Eds.), Marcel Dekker, New York (**1990**).

14. Cabbiness, D. K.; Margerum, D. W. *J. Am. Chem. Soc.* **1969**, 91, 6540–6541.

15. Cram, D.J. *Angew. Chem., Int. Ed. Engl* **1988**, 27, 1009.

16. Jeffrey, G. A., *An Introduction to Hydrogen Bonding*, Oxford University Press: Oxford (**1997**).

17. Lehn, J.-M., *Supramolecular Chemistry*. 1 ed.; VCH: Weinheim, (**1995**).

18. Ma, J. C.; Dougherty, D., *Chem. Rev.* **1997**, 97, 1303–1324.

19. Hunter, C. A.; Lawson, K. R.; Perkins, J.; Urch, C. J., *J. Chem. Soc., Perkin Trans. 2* **2001**, 651–669.

20. Kim, E.-I.; Paliwal, S.; Wilcox, C. S. *J. Am. Chem. Soc.* **1998**, 120, 11192.

21. (a) Flurry R. L. *J. Phys. Chem.* **1965**, 69; (b) Flurry R. L. *J. Phys. Chem.* **1969**, 73.

22. Kitaura, K.; Morokuma, K. *Int. J. Quantum Chem.* **1976**, 10, 325-340.

23. Rudkevich D. M.; Rebek J. Jr. *Eur. J. Org. Chem.* **1999**, 1991-2005.

24. Tucci F. C.; Rudkevich D. M.; Rebek J. Jr. *J. Org. Chem.* **1999**, 64, 4555-4559.

25. Kim, D. H., "Supramolecular aspects of enzymes", in *Comprehensive Supramolecular Chemistry*, J.L. Atwood, J.E.D. Davies, D.D. MacNicol and F. Vögtle (eds), Pergamon: Oxford, **1996**, vol. 4, 503–526.

26. Cragg, Peter J., *Supramolecular chemistry: from biological inspiration to biomedical applications,* Springer, New York (**2010**)

27. Sénéque, O.; Ranger, M.N.; Giorgi, M.; Reinaud, O. *J. Am. Chem. Soc.* **2000**, 122, 6183-6189

28. (a) Jones, S.; Thornton, J. M. *Proc. Natl. Acad. Sci.* U.S.A. **1996**, 93, 13–20. (b) Guo, Z.; Zhou, D.; Schultz, P. G. *Science* **2000**, 288, 2042–2045. (c) Berg, T. *Angew. Chem., Int. Ed.* **2003**, 42, 2462–2481. (d) Arkin, M. R.; Wells, J. A. *Nat. Rev. Drug Discovery* **2004**, 3, 301–317.

29. (a) Fletcher, S.; Hamilton, A. D. *Curr. Opin. Chem. Biol.* **2005**, 9, 632–638. (b) Fletcher, S.; Hamilton, A. D. *J. R. Soc., Interface* **2006**, 3, 215–233.

30. Hoffmann, A. S. J. *Biomed. Mater. Res.* **2000**, 52, 577–586.

31. (a) Zutshi, R.; Brichner, M.; Chmiclewsky, J. *Curr. Opin. Chem. Biol.* **1998**, 2, 62–66. (b) Peczuh, M. W.; Hamilton, A. D. *Chem. Rev.* **2000**, 100, 2479–2494. (c) Yin, H.; Hamilton, A. D. *Angew. Chem., Int. Ed.* **2005**, 44, 4130–4163. (d) Spring, K. I.; Hamilton, A. D. *Bioorg. Med. Chem. Lett.* 2**005**, 15, 3908–3911. (e) Fletcher, S.; Hamilton, A. D. *Curr. Top. Med. Chem.* **2007**, 7, 922–927.

32. (a) Orner, B. P.; Ernst, J. M.; Hamilton, A. D. *J. Am. Chem. Soc.* **2001**, 123, 5382–5383. (b) Okhanda, J.; Lockman, J. W.; Kothare, M. A.; Qian, Y.; Blaskovich, M. A.; Sebti, S. M.; Hamilton, A. D. *J. Med. Chem.* **2002**, 45, 177–178. (c) Yin, H; Hamilton, A. D. *Bioorg. Med. Chem. Lett.* **2004**, 14, 1375–1379. (d) Yin, H.; Lee, G.-I.; Sedey, K. A.; Rodriguez, J. M.; Wang, H.-G.; Sebti, S. M.; Hamilton, A. D. *J. Am. Chem. Soc.* **2005**, 127, 5463–5468. (e) Yin, H.; Lee, G. I.; Sedey, K.; Kutzki, O.; Park, H. S.; Orner, B. P.; Ernst, J. M.; Wang, G. H.; Sebti, S. M.; Hamilton, A. D. *J. Am. Chem. Soc.* **2005**, 127, 10191–10196. (f) Saraogi, I.; Hamilton, A. D. *Biochem. Soc. Trans.* **2008**, 36, 1414–1417.

33. (a) Ghosh, I.; Chmielewsky, *J. Chem. Biol.* **1998**, 5, 439–445. (b) Schramm, H. J.; de Rosny, B.; Reboud-Ravaux, M.; Buttner, J.; Dich, A.; Schramm, W. *Biol. Chem.* **1999**, 380, 593–596. (c) Strong, L. E.; Kiessling, L. L. *J. Am. Chem. Soc.* **1999**, 121, 6193–6196. (d) Kiessling, L. L.; Gestwicki, J. E.; Strong, L. E. *Curr. Opin. Chem. Biol.* **2000**, 4, 696–703. (e) Sandanaraj, B. S.;

Vutukuri, D. R.; Simard, J. M.; Klaikherd, A.; Hong, R.; Rotello, V. M.; Thayumanavan, S. *J. Am. Chem. Soc.* **2005**, 127, 10693–10698.

34. (a) Fischer, N. O.; McIntosh, C. M.; Simard, J. M.; Rotello, V. M. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, 99, 5018–5023. (b) Hong, R.; Fischer, N. O.; Verma, A.; Goodman, C. M.; Emrick, T.; Rotello, V. M. *J. Am. Chem. Soc.* **2004**, 126, 739–743. (c) Hong, R.; Emrick, T.; Rotello, V. M. *J. Am. Chem. Soc.* **2004**, 126, 13572–13573. (d) You, C.-C.; De, M.; Han, G.; Rotello, V. M. *J. Am. Chem. Soc.* **2005**, 127, 12873–12881. (e) Fillon, Y.; Verma, A.; Ghosh, P.; Ermenwein, D.; Rotello, V. M.; Chmiclewsky, J. *J. Am. Chem. Soc*. **2007**, 129, 6676–6677.

35. (a) Hayashi, T.; Hitomi, Y.; Ogoshi, H. *J. Am. Chem. Soc.* **1998**, 120, 4910–4915. (b) Jain, R. K.; Hamilton, A. D. *Org. Lett.* **2000**, 2, 1721–1723. (c) Aya, T.; Hamilton, A. D. *Bioorg. Med. Chem. Lett.* **2003**, 13, 2651–2654.

36. (a) Lin, Q.; Park, H. S.; Hamuro, Y.; Lee, C. S.; Hamilton, A. D. *Biopolymers* **1998**, 47, 285–297. (b) Park, H. S.; Lin, Q.; Hamilton, A. D. *J. Am. Chem. Soc.* **1999**, 121, 8–13. (c) Lin, Q.; Hamilton, A. D. *C. R. Chim.* **2002**, 5, 441–450. (d) Park, H. S.; Lin, Q.; Hamilton, A. D. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, 99, 5105–5109. (e) Mecca, T.; Consoli, G. M. L.; Geraci, C.; Cunsolo, F. *Bioorg. Med. Chem. Lett*. **2004**, 14, 5057–5062. (f) Woscholski, R.; Hailes, H. C.; Numbere, M. G. *PCT Int. Appl.* WO 2005023744 A2 20050317, 2005. (g) Francese, S.; Cozzolino, A.; Caputo, I.; Esposito, C.; Martino, M.; Gaeta, C.; Troisi, F.; Neri, P. *Tetrahedron Lett.* **2005**, 46, 1611–1615. (h) Mecca, T.; Consoli, G. M. L.; Geraci, C.; La Spina, R.; Cunsolo, F. *Org. Biomol. Chem.* **2006**, 4, 3763–3768.

37. Gutsche, C. D. In Calixarenes Revisited; Stoddart, J. F., Ed.; The Royal Society of Chemistry: Cambridge, (**1998**).

38. In Calixarenes in Action; Mandolini, L.; Ungaro, R., Eds.; Imperial College Press: London, (**2000**).

39. In Calixarenes 2001; Asfari, Z., Böhmer, V., Harrowfield, J., Vicens, J., Eds.; Kluwer Academic Publishers: Dordrecht, The Netherlands, (**2001**).

40. *Supramolecular Chemistry- Fundamentals and Application (Advanced Textbook),* Katsuhiko, A., Toyoki, K., Springer- Verlag (**2006**).

41. Casnati, A.; Sansone, F.; Ungaro, R. *Acc. Chem. Res.*, **2003**, 36, 246-254.

42. Ludwig, R., *Microchim. Acta*, **2005**, 152, 1-19.

43. Arduini, A.; Pochini, A.; Secchi, A.; Ugozzoli, F. "Recognition of Neutral Molecules". In Calixarenes 2001; Asfari, Z., Böhmer, V., Harrowfield, J., Vicens, J., Eds.; Kluwer Academic Publishers: Dordrecht, The Netherlands (**2001**); Chapter 25.

44. Dalla Cort, A.; Mandolini, L. "Calixarenes As Hosts for Quats". In Calixarenes in Action; Mandolini, L., Ungaro, R., Eds.; Imperial College Press: London (**2000**); Chapter 5.

45. Steed J. W. and Atwood J. L., *Supramolecular Chemistry, 2nd edition*; John Wiley & Sons, Ltd (**2009**).

46. Abis, L.; Dalcanale, E.; Du Vosel, A.; Spera, S. *J. Org. Chem.,* **1988**, 53, 5475.

47. (a) Botta, B.; D'Acquarica, I.; Delle Monache, G.; Subissati, D.; Uccello-Barretta, G.; Mastrini, M.; Nazzi, S.; Speranza, M. *J. Org. Chem.* **2007**, 72, 9283–9290. (b) Botta, B.; Fraschetti, C.; D'Acquarica, I.; Speranza, M.; Novara, F. R.; Mattay, J.; Matthias, C. L. *J. Phys. Chem. A* **2009**, 113, 14625–14629. (c) Botta, B.; Delle Monache, G.; Salvatore, P.; Gasparrini, F.; Villani, C.; Botta, M.; Corelli, F.; Tafi, A.; Gacs-Baitz, E.; Santini, A.; Carvalho, C. F.; Misiti, D. *J. Org. Chem.* **1997**, 62, 932–938.

48. Hamuro, Y.; Crego Calama, M.; Park, H. S.; Hamilton, A. D. *Angew. Chem., Int. Ed.* **1997**, 36, 2680–2683.

49. Verma, A.; Simard, J.; Rotello, V. M. Langmuir **2004**, 20, 4178–4181.

50. Botta, B.; Di Giovanni, M. C.; Delle Monache, G.; De Rosa, M. C.; Gàcs-Baitz, E.; Botta, M.; Corelli, F.; Tafi, A.; Santini, A.; Benedetti, E.; Pedone, C.; Misiti, D. *J. Org. Chem.* **1994**, 59, 1532–1541.

51. (a) Schellenberger, V.; Braune, K.; Hofmann, H.-J.; Jakubke, H.-D. *Eur. J. Biochem.* **1991**, 199, 623–636. (b) Hedstrom, L.; Perona, J. J.; Rutted, W. *J. Biochemistry* **1994**, 33, 8757–8763. (c) Perona, J. J.; Craik, C. S. J. *Biol. Chem.* **1997**, 272, 29987–29990.

52. (a) Dixon, M. Biochem. J. 1953, 55, 170–171. (b) Ahmad, V. U.; Lodhi, M. A.; Abbas,M. A.; Choudhary, M. I. *Fitoterapia* **2008**, 79, 505–508.
53. Huey, R.; Morris, G. M.; Olson, A. J.; Goodsell, D. S. *J. Comput. Chem.* **2007**, 28, 1145–1152.

54. Lozano, P.; De Diego, T.; Iborra, J. *L. Biotechnol. Prog.* **1996**, 12, 488–493.

55. MacroModel v. 9.5 (Amber*): Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T.; Still, W. C. *J. Comput. Chem.* **1990**, 11, 440–467.

56. Botta, B.; Caporuscio, F.; Subissati, D.; Tafi, A.; Botta, M.; Filippi, A.; Speranza, M. *Angew. Chem., Int. Ed.* **2006**, 45, 2717–2720.

57. Botta, B.; D'Acquarica, I.; Nevola, L.; Sacco, F.; Valbuena Lopez, Z.; Zappia, G.; Fraschetti, C.; Speranza, M.; Tafi, A.; Caporuscio, F.; Letzel, M.; Mattay, J. *Eur. J. Org. Chem.* **2007**, 5995–6002.

Part II

NMR characterization of Italian virgin olive oils

1. Introduction

1.1 Definitions

Olive oil can be defined as a vegetable fat obtained by extraction from olives, that at room temperature looks like a viscous liquid.
Olive oil composition depends on the action and the interaction of different factors:

- biological factors (genetic): cultivar, grafting;
- agronomic factors: type of farm, cultural practices (irrigation, pruning, fertilizing, harvesting, etc..);
- environmental factors: latitude, altitude, soil, slope, climate;
- technological and process factors: washing method, milling, extraction, storage, etc..;
- anthropological factors: structure of the land tenure, local customs and traditions, adulteration, etc.
- temporal factors: harvesting year.

The definition of the olive oil is often associated to the quality concept. It is important to specify that there is not only one concept of quality, but many concepts according to different situations or different users. For istance, there are the commercial quality, the agronomical quality, the legal quality, the health quality, the culinary quality, and so on. Moreover each user can give his opinion according to his preferences or needs, adding a subjectivity element to the quality definition.
On these basis we cannot say if an olive oil is good or not, but we can say if an olive oil meets certain standards. Moreover, we can say that we deal with a high quality oil, no matter what kind of quality it is, when all the phases of the production were “well” performed.

Another fundamental concept is the genuineness. This term often indicates the authenticity, referring to a food without substances foreign to the nature of the product, that at the same time meets all the qualifications that characterize the product.

A fundamental target is the development of more effective techniques for the detection of adulteration resulting both from industrial processes that attempt to reduce the defects of an olive oil (deacidification, deodorization, etc.) and from the addition of substances (chlorophyll, etc,) that can hide these defects. In this contest we must remember that the adulteration is a crime punished by the law.
The genuineness is different from the quality, because the former is only a part of the latter.

1.2 Olive oil chemical composition

Olive oil is a complex matrix, being its chemical composition affected by many factors.
In general, a saponifiable fraction, corresponding to the 98-99% of the product, and an unsaponifiable fraction, consisting of the remaining 1-2% are distinguished. The former contains all the substances that are degraded by treatment with concentrated alkali.

Saponifiable fraction
This fraction consists mainly of triglycerides, i.e. molecules obtained by the esterification of glycerol and fatty acids (figure **1.1**). The biosynthesis of these molecules in the fruit tissue, differently from the seed, follows the rule "1,3-random 2-restricted distribution", i.e. the secondary hydroxyl is esterified preferentially with unsaturated fatty acids, while the positions 1 and 3 are esterifyed with saturated fatty acids. [1]

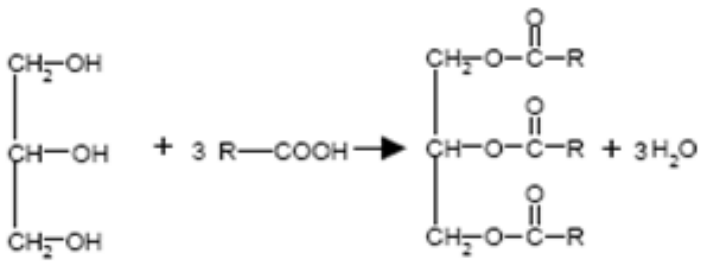


Figure 1.1*:* synthesis of triglycerides

There are also some monoglycerides (0.1-0.2%) and diglycerides (2-3%). The last group exists in two forms, as 1,2 and 1,3 substituted, depending on the position of the fatty chains on the glycerol molecule; moreover the 1,3 diglycerides derive from the 1,2 diglycerides transposition, while the 1,2 diglycerides are generated by the incomplete biosynthesis of the triglycerides. Consequently the 1,2 diglycerides are more abundant in the fresh olive oil, and, along time, they isomerise to 1,3 diglycerides, that become more abundant in the old olive oils. The 1,2 diglycerides/1,3diglycerides ratio is higher or equal to 2 in young and well conserved olive oils, it becomes 1.6-1.7 in one year old good quality olive oils and goes to 1.0-1.2 in very old olive oils or in those subjected to thermal stress or chemical treatment.[2]

The fatty chains linked to the glycerol can be saturated, without double bond, monounsaturated, with one double bond, or polyunsaturated, with more than one double bond.

The fatty acids composition of the olive oil is reported in table 1.1. The oleic acid is the most abundant fatty acid (56-84%, depending on the climate and the cultivar). This monounsaturated fatty acid has been correlated to the health properties of the olive oil more strictly than the other fats. It is easily metabolised by the cells, differing from the other saturated fatty acids, so it has a lower pro-atherogenic effect, and compared with the polyunsaturated fatty acids,it reduces the atherogenesis risk by limiting the oxidation of the circulating lipoprotein.[3,4] Moreover it reduces the plasma level of the LDL-cholesterol complex and increases those of the HDL-cholesterol.[5]

A good quality index consists in a concentration of the oleic acid lower than 73%, while the content of the linoleic acid must not reach 10%. The fatty acids content in the olive oil presents a polyunsaturated: monounsaturated: saturated ratio of 0.5:5:1 and this ratio gives a stability to the oxidation higher than a ratio of 5: 2: 1.

Table 1.1: Fatty acids composition of the olive oil.

Fatty acids composition of the olive oil		
Compound		%
Myristic acid	$C_{14:0}$	0-0.1
Palmitic acid	$C_{16:0}$	7.0-20.0
Palmitoleic acid	$C_{16:1}$	0.3-3.5
Eptadecanoic acid	$C_{17:0}$	0-0.4
Eptadecenoic acid	$C_{17:1}$	0-0.4
Stearic acid	$C_{18:0}$	1.4-4.0
Oleic acid	$C_{18:1}$	56.0-84-0
Linoleic acid	$C_{18:2}$	3.0-21.0
Linolenic acid	$C_{18:3}$	0.2-1.5
Arachidonic acid	$C_{20:0}$	0.1-0.7
Eicosenoic acid	$C_{20:1}$	0.1-0.5
Behenic acid	$C_{22:0}$	0-0.3
Lignoceric acid	$C_{24:0}$	0-0.4

Unsaponifiable fraction

This fraction contains different classes of compounds, at very low concentrations, that confer to the olive oil the organoleptic properties: the aroma (fruity, hint of apple, artichoke, almond, tomato), the typical flavour (bitter, spicy), and the biologic, healthy properties, due to the antioxidant and preservative activity of polyphenols and tocopherols.

Some of these substances (alcohols, sterols, hydrocarbons) can be used as frauds markers.

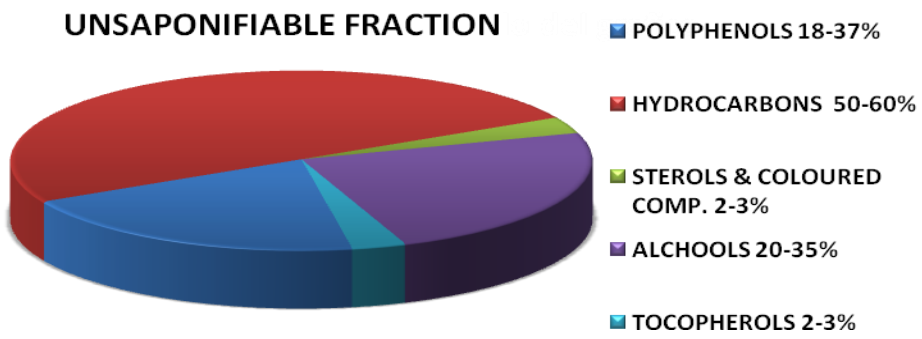


Figure 1.2: olive oil unsaponifiable fraction composition

Hydrocarbons: this class includes terpenoids, such as squalene, that is the most abundant (80%) and it is an intermediate of the steroids biosynthesis (a cholesterol precursor); the Polycyclic Aromatic Hydrocarbons (PAH), that can be found in traces and are considered as environmental poisoning, due to the proximity of the olive trees to factory sites or highways.

Alcohols: they are very volatile and contribute to determine the flavour of the olive oil; they are chemically unstable, so the olive oil tends to lose the aroma, by getting older. There are three most abundant kinds of alcohols:

- Triterpenic: these alcohols are very interesting because their composition is related to the botanic origin of the vegetable species. Some of them, such as the cycloartenol and the methylenecycloartenol, are present in all the vegetable oils, even if they have different concentrations; some others are detectable only in some particular species, such as butyrospermol in the tea.
- Biterpenic: eritrodiol and uvaol are the most abundant alcohols; they are detectable especially in the sansa oil, because they derive from the peel wax hydrolyzed during the extraction; so they are considered as a reference for the fraud discovering.
- Aliphatic: these alcohols include ethanol, pentacol, hexanol. As in the biterpenic case, they have high concentrations in the oils extracted from sansa.

Polyphenols: virgin olive oils are the only vegetable fats containing relevant concentrations of phenols;[6] they are antioxidant compounds, that protect the olive oil from the oxidation (rancidity); they are oxidized instead of the fatty acids, so they're concentration decreases with time and can give an indication about the age of the oil and its shelf life. Their concentration

depends on the cultivar and the harvesting period. In fact, they are more abundant in the green fruits and decrease with the fruit ripening. In nature they can be found esterified or etherified as glycosides; during the fruit ripening and the olive oil production, the release of the enzymes and the acidic pH of the medium allow the hydrolysis reaction and bring to the formation of the free phenols. The principal phenol in the olive oil is the oleuropeina, that hydrolyzes to elenolic acid and hydroxytyrosol. (Figure **1.3**)

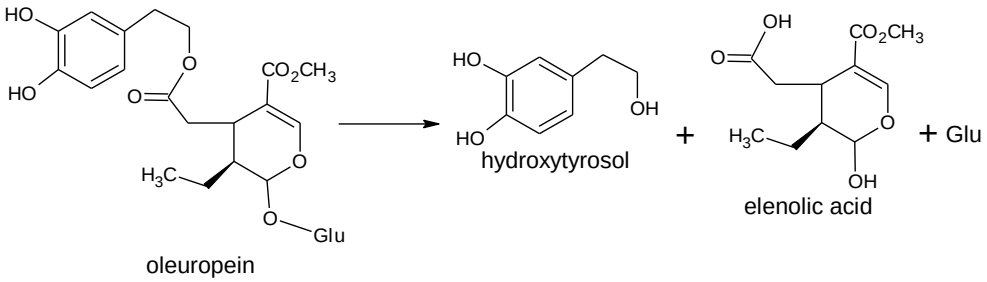


Figure 1.3: hydrolysis of the oleuropein

This class includes also tyrosol, caffeic acid and cumaric acid.
It's known that the highest antioxidant activity is possessed by the ortho-biphenols, such as hydroxytyrosol.[7]
The presence of the phenolic compounds determines a bitter and spicy flavour, but also a fruity taste; polyphenols are thermolabile and affected by the mechanic processing, so in a refined olive oil these compounds are absent.
Moreover a synergic action between the phenols and the α-tocoferol has been observed in the prevention of the degradation reactions of the olive oil.[8]

Tocopherols: they have an antioxidant activity. The most active is the α-tocopherol, also known as vitamin E.[9] They behave as primary antioxidant compounds, reacting directly with the radical, as shown in figure **1.4**

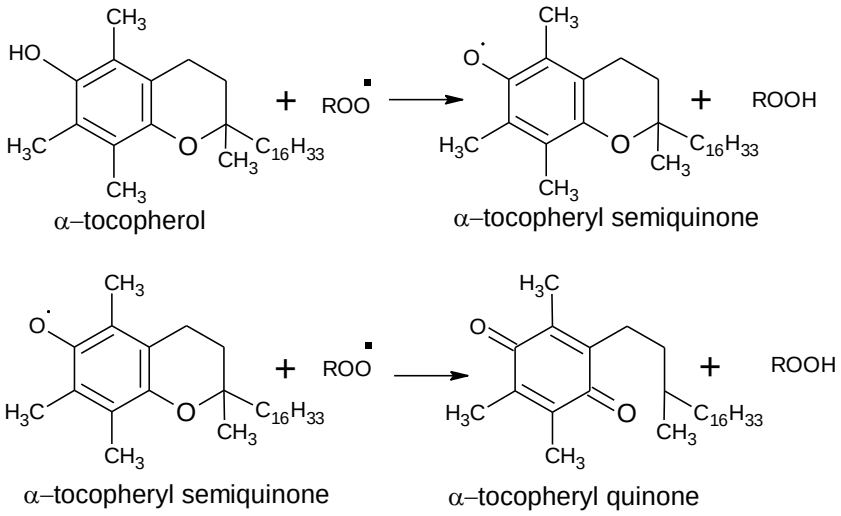


Figure 1.4: Radicalic reaction mechanism of the tocopherol.

Their activity is empowered by β-carotene, that acts as secondary antioxidant compound: it doesn't react directly with the radicalic species, but regenerates the primary antioxidant compounds that can newly prevent the oxidation.[10]
The technical processing of the oil (especially the refining) decreases inevitably their amount, with loss in the vegetable water.[11] It is worth of note that only the virgin olive oils contain remarkable concentrations of polyphenols and tocopherols; in this category of oils, the extravirgin olive oil has the highest concentration.

Sterols: these compounds present a common structure referable to the cyclopentan-perhydro-phenanthrene, with a double bond in the B-ring, a hydroxyl on the position 3 and an alkylic chain in the position 17 that differentiates the various sterols. (Figure **1.5**)
They have various roles in nature: they are precursors of cholesterol, steroid hormones, some vitamins (D group), and also fundamental constituents of the cell membranes and walls. In vegetals, the sterols composition is related to the botanic family of the species considered, so this fraction can be defined as a finger print of the olive oil: in the olive oil the β-sitosterol is predominant, in the sufflower and sunflower oils the Δ-7- stigmatenol is predominant.

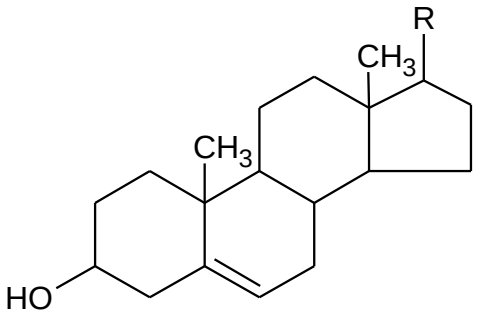


Figure 1.5: general structure of the backbone of the sterols.

For its particular composition a "nutraucetical" (nutritional and therapeutic) effect has been attributed to the olive oil,[12] because of its capacity to neutralize both the external free radicals, present also in the same oil, and also the endogenous ones, present at cellular level in the human tissues.

The protective properties of olive oils are related to:

- The phenolic fraction
- The presence of α-tocopherol, β-carotene and their synergic effect
- The prevalence of oleic acid, which requires less protection by the antioxidant micronutrients.

1.3 Classification and conventional analysis of the olive oil

The first olive oil definition dates back to the 1908 with the Italian law n 136. In 1925 the Royal Decree n 2033 was promulgated, that prohibits the marketing of mixtures of olive oil with other vegetable oils. Later, in 1936, another Royal Decree (XIV) defined six classes of olive oil: for the first time the free acidity is the main criterion for the class definition.
The definition "extra virgin olive oil" was born in 1960 with the so called "Solari" law, valid until the 1991, when the normative competence passed from each Country to the European Community.
In the same year the Reg. 2568/91 was promulgated, that modifies the previous Reg. 136 of the 1966, creates nine oil categories and regulates in a complete way the olive oils topic.

This regulations has been repeatedly modified and updated, while remaining reference legislation. In particular the Reg. 1513/2001 (since November 1, 2003) established 8 categories, removing the class of the "current olive oils".
The new classification is here reported:

VIRGIN OLIVE OILS: obtained from the olive fruit only by mechanic or physical processes, under conditions that don't cause alterations in the olive oil and that don't undergo any treatment other than washing, decantation, centrifugation or filtration, except the oils obtained by solvent or using adjuvants having a chemical or biochemical action, or re-esterification processes and any mixture with oils of other kinds.
These olive oils are classified as follows:

a) *Extra virgin olive oil*: virgin olive oil which has a free acidity, expressed as oleic acid, of no more than 0.8 g per 100 g of oil (0.8%)
b) *Virgin olive oil:* Virgin olive oil which has a free acidity, expressed as oleic acid, of not more than 2 g per 100 g of oil (2.0%)
c) Lampante Virgin Olive Oil: Virgin olive oil which has a free acidity, expressed as oleic acid, higher than 2 g per 100 g of oil (3.3%)

REFINED OLIVE OIL: the olive oil obtained from virgin olive oils by refining methods, with a free acidity, expressed as oleic acid, of no more than 0.3 g per 100 g of olive oil (0.3%).

OLIVE OIL- MIX OF REFINED OLIVE OILS WITH VIRGIN OLIVE OILS: Olive oil obtained by combination of refined olive oil with virgin olive oil, excluded the lampante olive oil, and having a free acidity, expressed as oleic acid, not higher than 1 g per 100 g of olive oil (1.0%)

CRUDE OLIVE- SANSA OIL: olive oil obtained from the olive sansa by treating with solvents or physical process, or classifiable as lampante olive oil except for certain specified characteristics and excluding the oil obtained by riesterification and mixture of oils of different nature.

REFINED OLIVE-SANSA OIL: oil obtained from crude olive-sansa oil by refining methods, with a free acidity, expressed as oleic acid, not higher than 0.3 g per 100 g of olive oil.

OLIVE-SANSA OIL: combination of refined sansa-olive oil and virgin olive oil, excluded the lampante olive oil, with a free acidity not higher than 1 g per 100 g of olive oil (1%)

Only four of the mentioned classes can be sold at the retail market, namely extra virgin olive oil, virgin olive oil, olive oil and olive-sansa oil.
It is worth of note that the refined olive oils (both from sansa and from olives) undergo a treatment that compromises the characteristic olive oil aroma. Moreover for both the olive oil and olive-sansa oil the regulation doesn't explain the minimal percentage of the added virgin oil: this means that by adding only 1% of virgin olive oil to a refined olive oil or to a refined olive-sansa oil, the resulting products will fit the commercial categories of olive oil and olive-sansa oil.
Although not strictly related to the product, we should mention those normative references that allow to improve and to control the industrial process, for example the EC directive 43/93, concerning the ISO standards and in particular the ISO 9000, or the norm that imposes the olive oil producers to fit to HACCP methods.
The present norm consists in 27 analytical chemical and sensorial parameters, that allow to classify the olive oil according the commercial classes.
The high number of parameters is due to the intense research and control activity developed in the recent years, mostly in Europe. The most clear definition of the olive oils is in the Reg. 2568/91, constantly updated by a committee of experts from various countries: in the Annex I it establishes the analytical limits for the parameters and the sampling criteria (Annex I *bis*); moreover a detailed set of decision-making schemes is given (Annex I *tris*) to help the choice of the analysis and the classification of the oil according the categories established in the same regulation. To the established analysis it is possible to associate a series of not regulated analysis.
The analytical parameters are divided in two groups: the ones concerning the genuineness and the ones concerning the quality. The parameters related to the quality give an indication about the quality of the olive oil. The free acidity values, the peroxide value, the UV spectrophotometric absorptions, the fatty acid composition, the amount of minor polar components, etc., all those parameters depend on some specific factors

such as cultivar, the extraction process, methods of storage, collection and transport of the olives and oils, the ripeness degree, etc.[13]
The genuineness parameters are physical-chemical parameters, generally used to verify the authenticity of the olive oils or the absence of different oils and substances foreign to the nature of the product.[13]
However we cannot separate the two groups, because some of the analysis give information about both the genuineness and the quality of the oil. Moreover new techniques are in development to discover the adulterations. But also the frauds become always more difficult to be detected, and some of the older analysis are no more useful in the fraud investigation, and it is necessary to replace them with new techniques, such as the gas-chromatographic analysis of the sterols.
The most frequent frauds concern the classification of the olive oil, that are sold as a product of a superior class. The virgin olive oils can also be mixed with other kind of seeds oils. Often some physical-chemical processes (refinement) are performed to hide these frauds, removing from the oil compounds such as free fatty acids, oxidation products, polycyclic aromatic hydrocarbons, residues of proteic substances, phosphatic fertilizers, mucilage, waxes, salts of alkali metals, alkaline earth metals, iron and copper, peroxides, excess of colouring substances (chlorophyll), other contaminating compounds.[14]
The most common refining processes are:

- degumming (removal of the lipid complex);
- de-acidification (removal of free acids);
- discoloration (removal of pigments);
- deodorization (removal of the glycerides with a high content of saturation degree).

Here we report some of the analytical parameters usually considered for the oil evaluation.

FREE ACIDITY DETERMINATION
It is the most ancient and diffused analysis of this product classification.
Even if it is not sufficient to define the quality of the olive oil, it gives a fast indication of its class.
This analysis indicates the amount of the fatty acid released by hydrolysis from the glycerol. The increase of its value happens in the first steps of the processing and is affected by the olive quality and their storage conditions.

Depending on the raw material, it cannot be easily modified by processing and refining the product without modifying also the diglycerides content.[15]
Early harvesting and short shelf life are fundamental conditions to obtain olive oils with low free acidity values. Also the temperature control during the processing and the separation of the oil from the vegetable water are extremely important.
The lipase can operate on monoglycerides, diglycerides and triglycerides,[16] preferentially at 35-40 °C and in presence of water.[17]
The measurement is performed (Annex II) by dissolving the olive oil in an alcohol-ether mixture, and then titrating with an ethanol solution of potassium hydroxide in presence of phenolphthalein. The free acidity is expressed as grams of oleic acid present in 100 g of oil.
The acidity can be easily corrected by the addiction of bases that can neutralize the free fatty acids and removing the products of this reaction by washing the olive oil. But this kind of fraud is easily revealed by the diglycerides measure: they cannot be washed away from the olive oil, revealing the deacidification process, and tend to isomerize, passing from 1,2 diglycerides, that decrease with time, to 1,3 diglycerides, that increase with time.[17]

DETERMINATION OF PEROXIDE:
This parameter evaluates the oxidation state of the olive oil. All the fatty substances are subjected to the oxidation that, if not controlled and limited, can modify profoundly the chemical structure of the triglycerides, with the formation of volatile compounds with an unpleasant flavour and smell. This phenomenon is called rancidity.
The olive oil oxidation depends on different factors that can act in two phases: during the cultivation, the harvest, the storage and processing of the olives or during the storage of the olive oil.[18]
The same process in the two cases have different causes. In a fresh olive oil the oxidation is due to the action of the lipoxidase,[19] that can bind an oxygen from the air to the unsaturated fatty acids of the triglycerides (enzymatic oxidation). This phenomenon is promoted by a bad health state of the olives or the fruit damages, that allow the contact between the oil and the enzyme, dissolved in the water phase, during the harvest, the storage and processing of the olive.
Also the temperature plays a key role in this process for the reaction rate. In fact, the enzyme maintains its activity also at low temperature (-40 °C), and

this explains the bad quality of the oils obtained from olives harvested late and damaged by frost.
However, the oxidation can happen also without the enzyme: the simple presence of the oxygen in the air during the olive oil storage can bring to the formation of hydroperoxides following a radicalic mechanism. Once the reaction starts, it goes on with the formation of new radicals, promoted by the light, the heat and the contact with some metals (iron, copper, nickel) that act like catalyst.
Hydroperoxides are very unstable compounds, rapidly decomposed in volatile compounds (aldehydes, ketones, alcohols), responsible for the rancidity.
An olive oil with a high concentration of hydroperoxides will go rancid faster than one with a lower content. From here the importance of the respect of the correct storage and production conditions is clear.
The olive oil has a good resistance against the oxidation because of its high content of natural antioxidant compounds, such as phenols. It has also a high percentage of monounsaturated fatty acids (almost totally oleic acid) and a low percentage of polyunsaturated fatty acids (linoleic and linolenic acid): the latter consumes the oxygen ten times more rapidly than the former and this is also the reason why the olive oils can be stored for a longer time than the seed oils, that have a higher concentration of polyunsaturated fatty acids.
The peroxides formation follows a bell-shaped trend in the time: after an initial increasing phase they reach the maximum, often over the law limit, and then decrease, converting themselves in aldehydes and ketons and alcohols. So a low content of peroxides can be found in a young, but also in a very old olive oil, with its characteristics totally compromised. A high level of peroxides in a young olive oil is certainly index of errors occurred during the productive phases.[20]
The utility of this analysis is restricted to the first months of the olive oil life, becoming less reliable with the time.[18]
The analytical protocol (Annex III) says to dissolve the olive oil in acetic acid and chloroform, then to treat it with a solution of potassium iodide. At the end the Iodine so produced is titrated with a standardized solution of sodium thiosulphate.
The virgin olive oil cannot contain more than 10 meq of O_2 per oil Kg. Refined olive oils must present a peroxides level lower than 5.

UV SPECTROPHOTOMETRIC ANALYSIS:

The natural polyunsaturated fatty acids present the two double bonds separated by a methylen group, namely two single bonds. The olive oil oxidation and the refining process lead to the so called double bonds conjugation, that's to say the shift of one position for one of the double bonds, so they are finally separated only for a single bond.

The phenomenon can be pointed out by measuring the UV absorption of the olive oil dissolved in a proper solvent and put in a UV spectrophotometer. (Annex IX). The absorption is expressed with some K coefficients: the higher the quantity of absorbed light, the higher the concentration of the conjugated double bonds.

A two double bond conjugation absorbs at 232 nm wavelength, a three double bond conjugation absorbs at 270 nm. But we must keep in mind that also the hydroperoxides absorb mainly at 232 nm and their increasing concentration will correspond to an increase of the K232 value. On the other hand aldehydes, ketones and the other products of the secondary oxidation absorb mainly at 270 nm, and their increase determines an increase of K270; moreover also these products can be decomposed after about 5-6 months, bringing a decrease of the absorption at 270 nm.[18]

So the UV spectrophotometry allows to identify the oxidation state of an olive oil and then to characterize it qualitatively. Moreover it allows to understand whether the olve oil underwent the refining process or not and to recognize a mixture of a virgin olive oil with any other kind of refined oil.

WAXES CONTENT DETERMINATION:

They are esters of fatty acids with high molecular weight alcohols, present in the external part of the fruit, with a protective function. During the processing of the olives they pass in solution: if their content overcomes the limit it means that the oil has been enriched with olive-sansa oil. In fact during the processing of the sansa oil the use of solvents bring in the solution also high quantity of waxes.[21]

Following the analytical procedure, an internal standard is added to the sample and then fractionated with chromatography on hydrated silicagel: the first, less polar fraction obtained, containing waxes, is then analyzed with gas chromatography, on a capillary column (Annex IV)

The wax limits are different in different classes of live oils: 250 mg/kg for virgin olive oils, 300 mg/kg for lampante olive oils, 350 mg/kg for the other oils.

DETERMINATION OF THE COMPOSITION AND THE CONTENT OF THE STEROLS

The progress of the analytical techniques allowed the study of the sterolic fraction and its variation according to the genetic origin, industrial processes, geographic factor, etc.

The sterol composition is the finger print of the vegetable oils, being strongly connected to the botanical family of the seed and the fruit from which the oil was extracted. This element is useful in the identification of the mixture of olive oil with seed oils rich of oleic acid (i.e. sunflower and sufflower), where the determination of the acidic composition is not effective for the target.[22,23]

Moreover these compounds, for the olive oil, don't undergo important variation in relation to the agronomic factors and the extraction procedure, while they can show significant variation after refining processes, so they are a very important index to distinguish virgin olive oils from refined oils.

The analytic method (Annex V) imposes to add to the oil the α-cholestanol as internal standard; then, after saponification with potassium hydroxide in ethanol, the unsaponifiable fraction is extracted with ethylic ether. From this fraction the sterolic component is separated by chromatography on basic silica gel plate; the sterols are then converted in trimethylsilyl ethers and then analyzed by gas chromatography on capillary column.

This method is sometime considered not so precise because unable to separate Δ5-avenasterol from the principal peak of the β-sitosterol. However the official method is sufficiently useful for the revelation of many adulterations.

DETERMINATION OF THE ERYTHRODIOL AND UVAOL

The erythrodiol and uvaol are triterpenic dialcohols located in the outer cuticle of the drupe that pass in small amounts in the olive oil obtained by mechanical extraction; these compounds are present in higher concentration in the olive-sansa oils because of the more efficient solubilization exerted by the extraction solvent. Their determination can help to point out the presence of the olive-sansa oil in the virgin olive oil.

The analytic procedure (Annex VI) provides the saponification of the fatty substance with potassium hydroxide in ethanol; then the unsaponifiable fraction is extracted with ethylic ether and purified with a column of alumina. The unsaponifiable fraction is than fractionated by thin layer

chromatography, isolating the sterolic fraction and the erythrodiol together. Then this fraction is converted in the corresponding trimethylsilyl ester; the mixture is so analyzed by gas chromatography. The result is expressed as percentage of the erythrodiol in respect of the erythrodiol and sterols together. The normes impose a limit to the content of erythrodiol and uvaol: if their content overcome 4.5% of the total amount of the sterols, the olive oil can be considered adulterated. However values included in the range 2.0-4.5% must be followed by further analysis, in particular concerning the wax content.

ANALYSIS OF THE METHYLIC ESTERS OF THE FATTY ACIDS

The olive oil has a percentage of fatty acids relatively defined, different from that of the other oleaginouses plants. The Regulation takes into consideration only some of the fatty acids, namely the less abundant, considered the most discriminant for the identification of possible illegal mixtures. The limits, equal for all the classes, are imposed in particular for the myristic acid (0.05%), linolenic acid (1.0%), arachidonic acid (0.06%), eicosenoic acid (0.4%), behenic acid (0.2%) and lignoceric acid (0.2%).

The method (Annex Xa) indicates the gas chromatography with capillary or packed column to determine the qualitative and quantitative composition of a mixture of methylic esters of fatty acids. The methylic esters (Annex Xb) can be prepared in two ways: transesterification, at room temperature, with methanolic solution of potassium hydroxide, or the methylation at high temperature with sodium methoxide in methanol, followed by esterification in acidic environment.

The gas chromatographic analysis of the fatty acids allows to identify the food frauds involving the seed oils by using also some characteristic markers of these oils. For example the mallow oil, typically introduced from India and very similar to the cotton oil, contains cyclopropenoic acids (non-linear acids). These markers are not present in the olive oil, hence a fraud with the addition of these oils is easily detected.

By the gas chromatographic analysis it is possible to identify some *trans* isomers, whose presence can give an important information about possible mixture with seed oils rich of oleic acid. The isomerization of the *cis* form to the *trans* form for the fatty acids can increase with the refining processes involving the heating of the oil (deodorization).[24] For the virgin olive oil, the *trans* isomers of the monounsaturated fatty acids (C 18:0) and the polyunsaturated fatty acids (C 18:2 + C 18:3) account for 0.05% of the total

acid content. The content of the *trans* isomers of the fatty acids with a number of carbon atoms between 10 and 24 can be determined by separation on a chromatographic columns with a particular polarity.[14]

DETERMINATION OF THE FATTY ACIDS IN POSITION 2

Analyzing the distribution of the fatty acids in the triglycerides of the olive oil we can see that the position 2 is preferentially occupied by unsaturated fatty acids, while the percentage of the saturated fatty acids in this position never overcomes 2%. In the chemically esterified oils, the saturated fatty acids distribution in position 2 is the same of the other positions.

The method can be applied to the oils that have a melting temperature lower than 45 °C because of the limit imposed by the use of the pancreatic lipase.

This method cannot be applied indifferently to the oils and to the fats containing a significant amount of fatty acids with 12 or less carbon atoms (coconut and palm seeds oil, butterfat) or unsaturated fatty acids (with more than four double bonds) containing 20 or more carbon atoms (fish oils and marine mammals oils), or fatty acids containing oxygenated groups different from the carboxylic group.

The method (Annex VII) suggests to neutralize the oil and to dissolve it in a solvent. The mixture is then purified by filtering through an alumina column. The eluted fraction is then submitted to the action of the pancreatic lipase for an established time and the formed monoglycerides are then separated by thin layer chromatography. At the end, the methyl esters are submitted to a gas-liquid chromatography analysis.

The natural olive oils can contain no more than 1.5% of palmitic acid in position 2 (percentage affected principally by the measure error), while the hydrogenated oils contain 10%.

DETERMINATION OF THE CONTENT OF TRILINOLEIN

In the triglycerides composition of the olive oils the trilinolein is present in small quantity. In the seed oils, considering the high value of the linoleic acid in the acidic composition, this triglyceride is more abundant. The method is applied to all the vegetable oils containing triglycerides of fatty acids with long chain. The method can identify the presence of small quantitative of semisiccative oils (rich of linoleic acid) in vegetable oils containing oleic acid as predominant unsaturated fatty acid, such as the olive oil.

The analytic procedure (Annex VIII) imposes the separation of the triglycerides, according to their equivalent carbon number, by using HPLC and the chromatograms interpretation.

DETERMINATION OF STIGMASTADIENES

The 3-5 stigmastadiene and 2-4 stigmastadiene are formed from the β-sitosterol when the oil undergoes discoloration with earth or in general when undergoes treatments involving heating,[25] so it is useful to understand whether the oil has been extracted at temperature higher than those recommended. The method is very useful to reveal the presence of refined vegetable oils in the virgin olive oil, since refined oils contain stigmastadienes, absent in the virgin olive oil.[13]

The analytic procedure (Annex XVII) is based on the isolation of the unsaponifiable fraction and the following separation of the steroidic component with hydrocarburic character with chromatography on silicagel column. The isolated fraction is then analyzed by gas chromatography on capillary column. The method can be applied to all the vegetable oils, but is reliable only if the content of these hydrocarbons ranges between 0.01 and 4.0 mg/Kg.

DETERMINATION OF THE TRIACYLGLYCEROLS WITH ECN 42

Using the fatty acid composition obtained by the gas chromatography the profile of triacylglycerols can be calculated theoretically, using the equivalent carbon number (ECN 42). This value can be compared with the experimental ECN 42 calculated on the basis of HPLC data.

If, for the virgin oils, the difference between the number of equivalent carbons experimentally measured and that theoretically calculated exceeds 0.2 we can assume that the sample is adulterated with seed oils (Annex XVIII).

We can distinguish three steps: the determination of the fatty acid composition by capillary gas chromatography, the calculation of the theoretical content of triacylglycerols with ECN 42 and the HPLC determination of triacylglycerols with ECN 42.

The method can be applied for the detection of small amounts of seed oils (rich of linoleic acid) in every class of olive oil.

DETERMINATION OF IODINE VALUE

The iodine value measures the unsaturation degree of the oil, and represents the amount (g) of iodine fixed in certain conditions by 100 g of dry and filtered oil. The analytical method (Annex XVI) provides that the substance is dissolved in the solvent and then the Wijs reagent is added. After a certain amount of time a potassium iodide water solution is added and the iodine produced is titrated with sodium thiosulfate.

The iodine number of the olive oils containing double bonds is not a quantitative evaluation of the total double bonds, but an empirical value, that is only an indicative value.

Following carefully the procedures we can obtain a good reproducibility.

A good olive oil must obtain a iodine value included in the range 79-88.

DETERMINATION OF HALOGENATED SOLVENT CONTENT OF the OLIVE OIL

The presence of halogenated solvents beyond the legal limits can be connected to the olive oil contamination during production, storage or transport.[13] The analytical determination (Annex XI) is done by gas chromatography using the technique of head space. In the olive oils the maximum limit allowed for halogenated solvents is 0.1 mg/kg for each solvent and 0.2 mg/kg as the total amount of all the halogenated solvents.

DETERMINATION OF THE CONTENT OF ALIPHATIC ALCOHOLS

In olive oils it is possible to determine the content of aliphatic alcohols one by one or totally. The analytical procedure (Annex XIX) is based on the addition to the olive oil of 1 - eicosanol as internal standard and the saponification with potassium hydroxide in ethanolic solution; the unsaponifiable fraction is then extracted with diethyl ether. From the unsaponifiable extract the fraction of the alcohol is separated by chromatography on basic silica gel; the alcohols recovered from the silica gel are transformed into trimethylsilyl ethers and then analyzed by capillary gas chromatography.

ORGANOLEPTIC EVALUATION OF VIRGIN OLIVE OIL

Keeping on mind that the olive oil quality is not only related to the chemical composition, but also to the organoleptic characteristics (colour, aroma, flavor), for the first time an organoleptic assay has been introduced in an European normative to evaluate qualities and defects of a food.

This test, even with its weakness due to the subjective component related to the assayer, represents the complement to the chemical analytical description of a virgin oil. In order to standardize as much as possible the method, the regulation (Annex XII) provides in detail all the procedures and the equipment required to carry out the taste test.
The regulation also provides a vocabulary of the most important sensations detectable in the virgin olive oil (bitter, sour, sweet, grass, fruity, almonds, artichokes, spicy, winey-vinegary, musty-humidity, heated, metal, sludge, rancid). The commissions established by a ministerial decree with the aim to test the organoleptic characteristics of the olive oils, give an assessment based on parameters set in the specification of the Denomination of Origin in question, in addition to those developed by the Community. The taste test is based on the evaluation of a group of experts consultants known as the "Panel", properly trained; they, coordinated by a master-Panel, verify the presence and the intensity of specific aromatic notes and any defects. In particular three tests are performed, related to the three senses cited before:

1. visive exam: to determine the clearness and the colour;
2. olfactory exam: the olive oil is smelt to perceive any odors associated with the quality of the oil
3. gustative exam: the olive oil is tasted to evaluate the intensity of the four basic tastes, i.e. sour, sweet, salty, acid and of course the combination of these tastes.

The evaluations are then converted in numeric values by a statistical system and the results are given as the median of the defects, considering the perceived negative attribute with the greatest intensity, and the median of the fruity.
In relation to the fruity concept we have to say that this flavour fades with the progressive maturation of the olive oil; for example, the bitter taste that causes the description as fruity is typical of the young oils obtained from olives harvested at the right moment, but this flavor fades with the time, becoming a fragrance more and more sweet and soft.
The method can be applied only to the virgin oils classification because it would be not useful to perform this test on olive oils qualitatively not adequate.

1.4 Olive oil in Italy

Italy is the second producer of olive oil in Europe and in the world, with a mean national production of 6 million quintals, two thirds of which are extra virgin olive oil. In Italy there are 37 Protected Designation of Origin (PDO) and a Protected Geographical Indication (PGI) recognized by the European Union.
In Italy the olive tree is present on almost 1 million of hectares as principal culture and on a slightly smaller area as secondary culture, in association with arable or other tree species (vines, citrus, almond, etc.).
Concerning the altimetry, the olive tree is diffused for the 11% on the mountains, for the 62% on the hills and for the 27% in plain.
For the characteristics of the plant, that requires a mild climate, the olive culture is widespread in the regions of the South–(79%) and Center Italy (19%) rather than in the Northern Italy (2%), where, however, it is growing, focusing particularly in some areas with more temperate microclimate, such as Liguria, or the hill areas around Garda Lake.

In Regulation (EC) 182 of March 6, 2009 amending the Regulation (EC) 1019/2002, the obligation to indicate on the label the origin of the olives used to produce extra virgin olive oil has been extended to all European Countries
This action is a guarantee for recognition of the origin of the olive oils produced in Italy and therefore for their quality, especially considering the large amount of this oil produced abroad and imported to Italy every year.
The label of each original Italian oil must report captions like “obtained from Italian olives”, “obtained from olives farmed in Italy”, “100% Italian olives”.

Table 1.2: Data from ISTAT, concerning the Italian olive oil production (the quantities are expressed in quintals) updated to January 2011

Regions	2006	2007	2008	2009	2010
Piedmont	67	104	107	139	133
Valle d'Aosta	-	-	-	-	-
Lombardy	7233	26418	7114	9732	9983
Liguria	33578	23799	31168	46430	36812
Trentino-Alto Adige	2256	2384	2436	2667	2250
Veneto	13456	13332	14158	13825	13975
Friuli-Venezia Giulia	805	224	285	428	342
Emilia-Romagna	8723	7403	11297	11668	10631
Tuscany	167427	137929	172659	191341	201360
Umbria	138747	108410	118207	90999	103400
Marche	39243	39356	47289	37655	44172
Latium	253180	226679	369739	271728	302553
Abruzzi	205930	158721	220302	185235	187480
Molise	66500	57200	57200	64225	59600
Campania	303726	367926	440964	425200	396058
Apulia	2154750	1906625	1903373	1523401	1648153
Basilicata	63843	63666	65.329	60.830	61.809
Calabria	2061436	2040522	2008255	1721692	1648310
Sicily	426940	476760	496698	454665	477342
Sardinia	84694	85154	101192	64791	63417
ITALY	6032534	5742612	6067772	5176651	5267780

2. Food analysis with NMR

2.1 Generalities

NMR spectroscopy is progressively expanding its application field in food chemistry, particularly in the study of the traceability aimed to analyze and guarantee the origin, the safety and the quality of the food products.
This is due to the ability of the high resolution NMR to resolve the spectra of complex mixtures and to quantify the components without preliminary separation and / or chemical derivatization. All these elements make the NMR techniques very sutable for the application in the metabolomic field, whose target is to identify different metabolites contained in the sample (cell or tissue). Different NMR approaches can be used to the foodstuffs analysis and their choice depends on the kind of study to perform.
If the primary target is the study of effect of a genetic modification, it will be possible to limit the analysis to the substrate or to the direct product, obtained because of this modification. This approach is called *target analysis* and is principally used for screening purposes.
Another analytic approach consists in limiting the identification and the quantification process to a restricted number of preselected metabolites in a biological sample. This process is called *metabolite profiling.* The choice of the metabolites can be due to criteria such as their membership to a class of compounds (carbohydrates, fats, amino acids) or to a particular metabolic pattern.
A third approach is represented by the complete analysis, the identification and the quantification of all the metabolites present in the considered system. This method reveals the metabolome of the analyzed biological system and for this reason is called *metabolomics.*
There is one additional approach, used when is not necessary to quantify the level of each metabolite, but it is sufficient to rapidly classify the samples according to their origin or biologic relation. It is called *metabolic fingerprint,* in reference to its ability to provide a metabolic signature of the sample without the need to identify and quantify the various components.[26]
Analytical techniques such as gas or liquid chromatography are very useful with the target analysis or the metabolite profiling when the components to

be determined are known and it's easy to identify and quantify them with the right standards. The NMR techniques are very suitable for the metabolomics and metabolic fingerprinting where the substances to be defined are not established a priori.
In this case the analysis is really easyer and faster than other techniques such as GC and HPLC.
The metabolomic analysis is very important in the comparative experiments, where the attention falls on the difference of the concentrations of some metabolites among the samples.
When the NMR spectra of a series of samples are acquired in the same conditions (temperature, sample concentration, pH, etc.) the NMR signal intensity is generally proportional to the concentration of the corresponding compound and can be used as it is. In this way, the extracted data are easy to be elaborated and undergo the multivariate statistical treatment, that allows an easy comparison. In this sense the NMR spectroscopy, combined with chemometric methods, has been applied successfully to the characterization of various vegetable products for the quality control, the authentication, determination of geographical origin and for the detection of alterations of the products.
The most used statistical analysis are the variance analysis (ANOVA), the principal component analysis (PCA) and the linear discriminant analysis (LDA).

2.2 Studies of characterization, classification and identification of adulterated vegetable products

The wine is the most known food studied by NMR. The analysis uses the isotopic method, namely the quantitative analysis of the isotopic distribution of ^{13}C, ^{15}N, ^{2}H, etc., on the functional groups. This method has been approved by the European Union as official analysis method.
NMR gave great contribution for the resolution of problems concerning the diagnostic and control also for the extra virgin olive oil
The official analytic methodologies for the oil analysis (Chemical analysis with extractions, UV spectroscopy, gas chromatography, enzymatic methods, panel test…) don't provide the necessary safeguards to face in a

reproducible and selective way topics such as the identification of the geographical origin and the adulterations.
The NMR spectroscopy allows to obtain a good chemical characterization of the olive oil, giving quantitative and qualitative information both on the principal components and the minor compounds, with the same, small experimental error, performing a single experiment that represents the fingerprint of the oil.
Furthermore the NMR experiment is also quick to be performed and requires no extraction or derivatization.
Its combination with the multivariate statistical analysis gives information about the effect of different factors on the composition of the olive oil.
In the recent years the potentialities of the NMR technique, concerning the identification of adulterations and the geographical origin of the olive oil, have been shown.[27,28,29]
The NMR technique has also been validated in Europe as an analytical method to highlight fraudulent additions of hazelnut oil to the olive oil.[30,31,32] The presence of the hazelnut can be determined on the basis of its low content of linolenic and palmitic acid and the squalene. This method allows to reach a detection limit of the hazelnut oil addition of 12%.
A lot of studies concerning the geographic origin of the olive oil have been carried out. In a study of some olive oils from four different Italian zones (Lucca, Arezzo, Lazio and Garda), the combination of the ^{1}H-NMR analysis and the multivariate statistical analysis gave a differentiation of the samples, regardless of the cultivar of the oils, also considering different seasons of production.[29]
A contemporary study showed the possibility of the NMR technique to differentiate olive oils coming from three different areas of the same region (Tuscany), suggesting that is possible to use this technique for the certification of the protected designation of origin (PDO) of the extra virgin olive oils.[33]
In a study of the olive oils from Latium[34] the information obtained from the ^{1}H-NMR and ^{13}C-NMR spectra, subjected to different statistical methods, allowed the differentiation of the oils from the three geographical areas of the region, mostly on the basis of the different concentration of the aldehydes, terpenes and squalene. In the same study some differences in the level of the sitosterol allowed the characterization of the oils from a pedoclimatic point of view and the content of the oleic acid and saturated fatty acids seemed to be strongly affected by the irrigation practice.

In a recent study some oils from Liguria have been chosen as a case study in a large population of olive oil samples collected over three consecutive years from the Mediterranean area. By the application of the statistical methods of classification and discrimination to the data from the ^{1}H-NMR experiments it has been possible to build some models able to discriminate the olive oils from Liguria from the others, defining the former group as characterized by specific markers that confirm the provenance. These results prove the contribute of the NMR methodology to the DOP characterization of the olive oils.[35]

Using the ^{13}C-NMR and gas chromatography together with the multivariate statistical analysis on olive oil samples collected in Sicily and containing three different cultivars it has been possible to demonstrate the strong relation existing between the fatty acid composition and the cultivar of the oils.[36]

Also the beer has been submitted to a metabolic profiling study: 50 beers differing for their type and label underwent ^{1}H-NMR experiments and the data were submitted to the statistical analysis. The PCA showed the separation of four groups differing for the prevalence of different carbohydrates, such as destrines, maltose and glucose. Moreover it has been observed that the groups were distinguishable also for the different composition of the aromatic component, index of different fermentation processes in the production of different types of beer.[37,38]

The high resolution NMR techniques have been used for the analysis of extracts and materials with vegetable origin, such as truffles[39] and strawberries.[40] The identification of a large group of compounds belonging to the hydrosoluble fraction, such as carbohydrates, amino acids, organic acids, phenols, but also the organic fraction, such as phospholipids, fatty acids and sterols brought the characterization of their chemical composition.

Fruits and fruit juices have so far been extensively studied in terms of metabolomics both for the simplicity in sample preparation and for the great importance they play in human and animal feeding.

Metabolic profiling studies with NMR techniques have been carried out on the apple, mango, orange and tomato juices.

These studies have shown that combining ^{1}H-NMR with multivariate statistical analysis we can distinguish different apple juices[41] and recognize mixed orange juices from pure ones.[42] The detailed analysis of the high resolution ^{1}H-NMR spectrum and bidimensional NMR spectra allowed the

assignment of more than 90 signals of nineteen metabolites and the quantitative and qualitative analysis of two different tomato cultivars (Red Setter e Ciliegino).[43]

Using the NMR techniques, different metabolic profiling studies demonstrated the possibility to distinguish different species and strains of plants but also different tissues within the same plant.[44,45]

With these analysis it has been possible to study the presence of modifications in the metabolic profile of genetically modified (GM) organisms. For example wild type tomatoes and transgenic cultivars containing regulatory genes from maize can be distinguished on the basis of substances involved in specific metabolic pathways.[46]

The introduction in *Nicotiana Tabacum* of genes from *E. Coli* and *Pseudomonas fluorenscens* provided the capability to synthesize the salicylic acid, associated with resistance against the tobacco mosaic virus. Using the NMR-PCA analysis Verpoorte and coworkers showed a clear metabolism differentiation in the leaves (and stems).[47]

Another metabolic fingerprint study has been carried out on the GM maize, allowing the differentiation of three modified lines from the control one based on the identification of the discriminant metabolites.[48]

Good separation between transgenic and "wild type" wheat can be obtained through statistical analysis of metabolic fingerprinting obtained by 1H-NMR techniques.[49]

The same has been done with GM peas, with the aim to control the undesired effects seen in the plants grown in greenhouses: the analysis showed metabolic differences overcoming the natural interindividual variability.[50] A metabolic profiling study has been carried out also on the GM potatoes.[51]

All these examples demonstrate how the NMR techniques, combined with the multivariate statistical analysis, can be useful and reliable in the control of quality, origin and safety of food.

3. Statistical analysis of the data

3.1 Introduction

The basis of any statistical analysis is the "good quality of the data". We define data of good quality those related to the problem we want to analyse, with a low noise and containing the highest possible amount of information. Moreover in a statistical analysis we have to consider all the variables that simultaneously govern the observed system, taking full advantage from the information contained in the data to analyze.
A correct statistical analysis rarely can be univariate, taking in consideration just one variable. The chemiometry is the discipline that allows a multivariate approach, analyzing how the variables co-vary.

3.2 Organization of the data

The data are organized in "raw" matrices where the rows correspond to the samples (cases) and the columns are the variables; so each samples is characterized by *c* variables (table **3.1**)

Table 3.1: raw matrix of data

	Variables				
Samples	X_{11}	X_{12}	X_{13}	…	X_{1c}
	X_{21}	X_{22}	…	…	X_{2c}
	X_{31}	…	…	…	…
	…	…	…	…	…
	X_{r1}	X_{r2}	…	…	X_{rc}

The values of the *c* variables dentify the position of each sample in a space of *c* dimensions, so we can imagine the sample as a point in the hyperspace of the variables (figure **3.1**). Even if it not so frequent, we can represent in the same way the variables in the space with *r* dimensions of the samples. In the case of a matrix composed by a single sample or a single variable its name is vector.
The points corresponding to different samples will be more or less scattered in the multidimensional space. This variability represents the chemical information we have to understand. A variable that remains constant for all the samples is absolutely irrelevant for the chemical interpretation and cannot be used to characterize and distinguish the objects, or to foresee their properties.

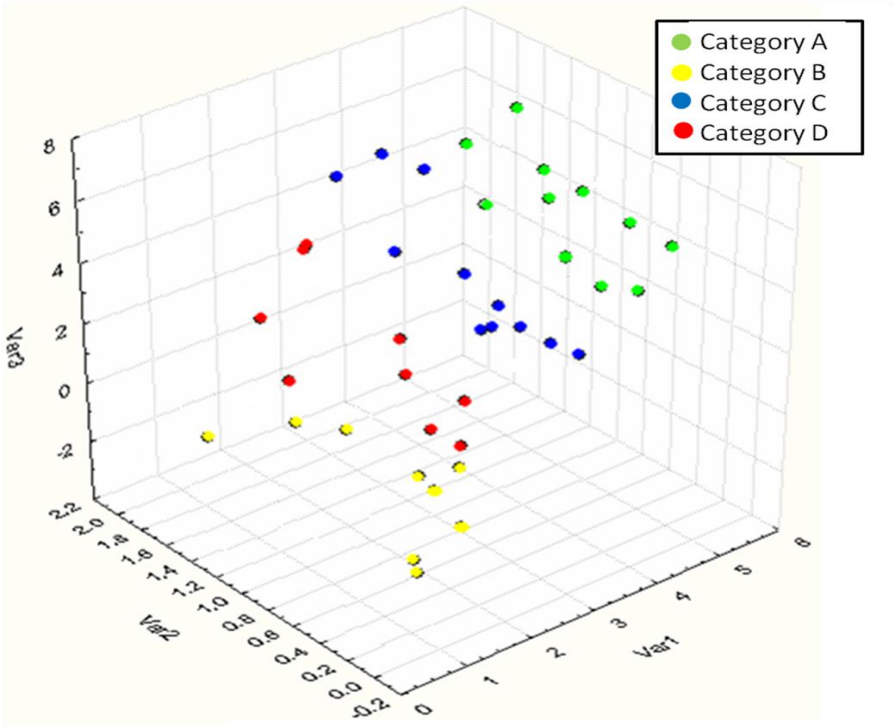


Figure 3.1: Distribution of the data in a three-dimensional space

Often the samples or the variables are associated to additive information with a descriptive character that can be useful for the subsequent analysis.

For the samples, these information is called label or index of category (i.e. cultivar, geographic origin, harvesting year, etc.), while for the variables there can be some variables "labels" that gather the same variables in groups (i.e. signals from carbonyl groups, signals from methyl groups). The presence of additional information is important because it increases the quantity of information that can be extracted from this data. For example, the variables of figure **3.1** allow the location in the space of the samples of four categories, but without the additional information they wouldn't have discriminant capability.
In this phase of the treatment of the data it can be useful to assign to the different samples a codified nomenclature that, at the same time, is concise and contains the most information possible about the various labels so that when the samples will be displayed in a multidimensional space it will be easy to identify any grouping.

3.3 Explorative analysis of the data

The exploratory analysis (or pretreatment) is of fundamental importance because it allows to have a preliminary idea about the trends and structure of the data, the correlation between the variables; moreover, it allows to identify and eliminate any possible anomalies that may affect all the subsequent analysis.
By this analysis we can:

- Remove undesired source of variability
- Remove (or integrate) of the missing data
- Partially remove the noise of the data
- Gain information on samples with a high dispersion of the mean value (outlier).

The elimination of source of undesired variability consists in the scaling of the data. This step is necessary when the variables have different units of measurement, and usually it involves the columns. Often, in fact, for each sample variables of different nature (hence, with different units) are reported , that give a significant contribution to the total variance.
It is helpful to remove the weight that is artificially imposed, since many analytical instruments recognize greater influence in variables with the

wider dispersion of data. The scaling data is performed using the range scaling, so the new value of the data will be

$$x'_{ij} = \frac{x_{ij} - x_{\min,j}}{x_{\max,j} - x_{\min,j}}$$

The range scaling is very sensitive to the outliers; the most used transformation is the autoscaling, and the variable will assume the following value:

$$x'_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}$$

The scaling of the data brings all the variables to have the same weight. This pretreatment must not be performed when the difference among the measured value is important for the definition of the problem we are considering, so the choice whether to perform or not the scaling must always be guided by the chemical sense.
More rarely the scaling is performed on the rows content. The most common transformation consists in constraining the sum of the terms of each line or their squares to assume a constant value

$$\sum_j x'_{ij} = k$$

Then we must establish what to do if some data are missing. For the ^{1}H-NMR data it happens often that the water signal, due to all the exchangeable protons, moves along the spectrum hiding some important signals and preventing their measurement. Since many multivariate statistical methods cannot work in presence of missing data, it will be necessary to provide their substitution. If they are replaced with zero the variance will be overestimated; if they are replaced with the mean value the variance will be underestimated. Usually methods that reconstruct the

missing data using the *Nipals* algorithm or regression techniques (*Linear Trend at Point*) are preferred.
If one or more compounds in a sample are absent (not because of specific interference or complications in the measurement, but because the compound is not actually present), we must consider that the information should not be discarded or undervalued.
With the oils, for example, the absence of the volatile compounds such as aldehydes or terpenes can be index of refining (deodorization). In this situation we must use the zero value even if, in a more rigorous way, we must use a value that includes the noise always associated to the analysis. In the NMR analysis we can calculate the signal-to-noise ratio and in this way recover the noise value (NOISE).
With regard to the anomalous data, often identified as *outliers,* they can have different origins: an intrinsic anomaly of the sample; an error in the sample preparation; an error in reading the variables; transcription error, miscalculation. Sometimes, after the identification of the outliers, it is possible to identify and correct any reading, typing or computation errors, and in some cases, where possible, it's even desirable to repeat the analysis/measurement.
In the worst case, when the anomaly is inherent to the sample, we can decide to reject the sample, to avoid to affect the outcome of the entire analysis.
The methods to identify the outliers are different and can be divided with a first approximation in multivariate and monovariate methods. It is also to be said that the use of one of the techniques doesn't excludes the others.

Descriptive statistics (monovariate method): the central trend and variability index easily and quickly give a useful information about the presence of any anomaly. Among the central trend index we find the mode (the most populated category), the median (the central value of the distribution), the mean (M) ("barycenter" of the distribution). This parameter, being very easy to manipulate and very sensitive to the external data, is extremely suitable for the identification of any outlier:

$$M = \frac{\sum x_i}{N}$$

Where N is the number of the examined cases.

Among the variability indices that measure the dispersion grade, namely the tendency of the cases to detach from the central value or to lay around the mean value we must remember: the *variance* (σ^2) (the mean of the squares of the differences between the observed values and the mean value)

$$\sigma^2 = \frac{\sum (x_i - M)^2}{N - 1}$$

and the standard deviation (σ) (square root of the variance) that is often used to evaluate the repeatability and reproducibly of the analysis. Differing from the variance, the measurement unit is the same of the original data.

Box-Plot (univariate method): graphic representations that describe in a compact and immediate way the distribution of data. There are two types of Box-Plot: those using the mean and standard deviation, and those using the median and quartiles; the latter (figure **3.2**) give better results because the median is less affected by outliers and extreme values.
In the box plot the data having values included between the first and the third quartile (Q_1 and Q_3) are placed in a rectangle. Then the difference Q_3-Q_1, named H; the limits are placed at Q_3 + 1.5H and Q_1-1.5H: the data external to these limits are outliers; the data external to the range Q_3 + 3H and Q_1 - 3H are defined as extreme data.

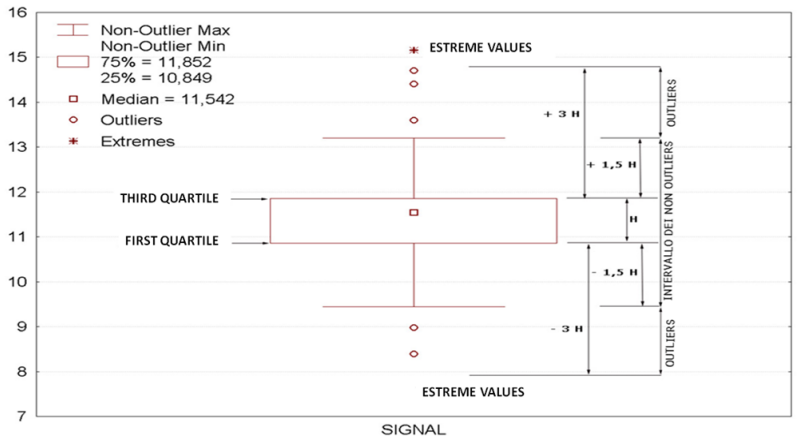


Figure 3.2: graphic representation of the box-plot

Principal component analysis (PCA) (multivariate method): it can be used also for the identification of the outliers. It has an advantage: it's fast and easy, but, differently from the other methods, it takes in consideration all the variables. However we must pay attention to the structure of the data, in order to avoid coinfusion between outliers and data placed at the limits of a trend (figure **3.3**).

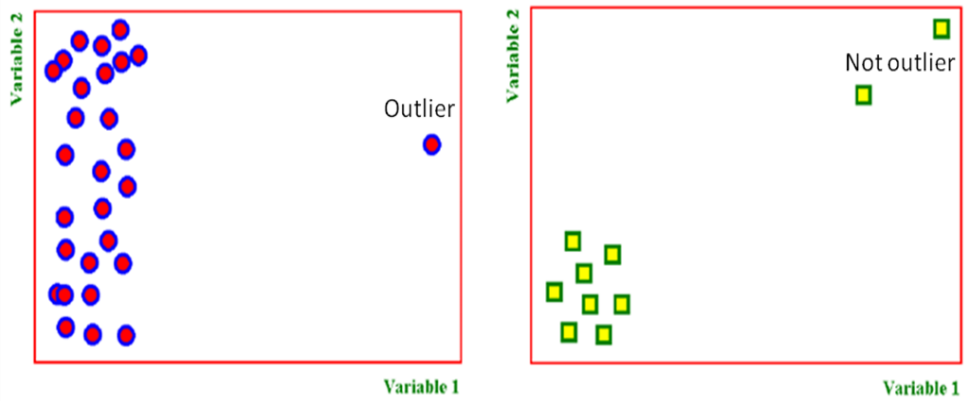


Figure 3.3: an example of different structures that a serie can assume in a bidimensional space.

With the outlier analysis, the first part of the explorative analysis of the *raw matrix*, that works on the cases, is concluded, giving the *correct matrix.* Then the correct *matrix* will undergo the next analysis (ANOVA, PCA, PDA, etc.), so this is the moment to test the normality and the correlation of the variables.

Some analyses, such as analysis of correlation coefficient, analysis of variance and linear discriminant analysis, require the assumption that data distribution is normal (Gaussian), having the classic bell-shaped pattern defined by the following function

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-M}{\sigma}\right)^2}$$

In this case it is always better to verify these assumptions, even though it can be hardly determined whether a given distribution is normal or not, especially when there are few samples available. It is useful, anyway, to

have an idea of which variables have an ideal condition and which are in a forced condition, that must be considered during the data interpretation. The normal distribution is characterized by the symmetry of the Gaussian curve, verified by the *Curtosi's index* (*Kurtosis)*

$$Ku = \frac{N(N+1)\sum(x_i - M)^4 - 3\sum(x_i - M)^2 \sum(x_i - M)^2 (N-1)}{(N-1)(N-2)(N-3)\sigma^4}$$

and by the *Skewness index,* whose values, in a normal distribution, are close to zero.

$$Sk = \frac{N\sum(x_i - M)^3}{\left[(N-1)(N-2)\sigma^3\right]}$$

In this phase it is useful also to control the correlation matrix, to identify any correlation among the variables (high correlation coefficients).

This information can be used in different ways:

- *making predictions*: if a variable is related to another, we can predict with a certain margin of error one of the variables using the other one;
- *reduce the uncertainty of prediction*: the correlation between the variables indicates that if some information is repeated, this can be used to decrease the uncertainty of estimates on the value of a third variable;
- *reduce the number of variables*: the correlation between the variables indicates that one of the two variables gives the same information provided by the other. In this case we talk about redundancy, and in some cases it is convenient to eliminate one of two experimental variables to reduce costs;
- *verify the prerequisites for the subsequent multivariate analysis*: for many multivariate techniques, such as linear discriminant analysis (LDA), a prerequisite for their application is that the variables are independent from each other, i.e. the variables need not to be related. A measure of the degree of linear association between the variables is the *tolerance,* for which often the statistical software sets a threshold; the smaller the tolerance the higher the dependency. The variables with low values of tolerance should be excluded from the linear discriminant analysis, since they do not give additional information and increase the noise.

3.4 Analysis of the Variance

The analysis of Variance, also known as ANOVA (Analysis Of Variance) allows to establish whether a certain variable can discriminate between two or more factors. For example, for the olive oils, they can come from two or more areas, or be two or more cultivars. This statistical technique compares groups of data evaluating the internal variability of these groups and the variability among the groups. To evaluate the variability we can use the variance, as an index of data dispersion compared to their average value. High variance means that the data of a group have values very different, low variance means that the values of the group are very similar.
This important analysis provides a useful information that may allow an increase of rate, effectiveness and efficiency of all the subsequent multivariate analysis (TCA, PCA and LDA in particular). Without the information provided by the ANOVA we can perform hundreds of tests, without knowing which variables could be used and also risking to include sources of noise arising from unnecessary variables that have no discriminating power.
We have different types of ANOVA.[52] Rigorously the simplest ANOVA analyzes one variable and one factor (and is called univariate unifactorial ANOVA). ANOVA is also very flexible and can be extended to multivariate and multifactorial situation. In the multivariate ANOVA (also known as MANOVA, Multivariate Analysis Of Variance), we can study the differences among different populations when, for each element, more than one variable is measured, and the variables are considered contemporarily (combined effects).
In practice, the ANOVA is an extension of the t-test, with the advantage that it reduces of the error associated to the repetition of the t-test.
In ANOVA, the average are evaluated considering whether the variability between groups is greater than the variability within groups, the Fisher test or F ratio, expressed as the ratio of variance between groups and intra-group variance. We can also use the p-level, defined as the measure of the credibility of a result, i.e. the probability of error that we can commit, accepting as true the observed results.

The ANOVA analysis is based on the condition of normal distribution of the values of the variables in each population and of equality of variances within the different populations. However, this statistical analysis is quite robust and can be applied even if the data are not normal and the variances are equal, with the condition that the different samples have the same size dimensions of the different samples are equal.
In performing the test, Student-Newman-Keuls (SNK or procedure of Newman and Keuls) is often used, by which you can assess the real discriminating power in case of a multiple comparison, that is when the label contains more than two options (for example if we have more than two cultivars).
It may happen that, using the analysis of variance in the most classical way, the resulting p-level sujests that the considered variable is highly discriminant, but the more accurate SNK test indicates that the situation is different: for instance, when comparing 3 cultivar, SNK tests may indicate that such variable is able to discriminate the cultivar 1 from 2 and 1 from 3 cultivars, but not cultivar 2 from 3. Generally this test is conducted using as a threshold value the p-level of 0.05.

3.5 Principal component analysis[53]

The sample can be considered as a point with *c* dimensions in the variables hyperspace. A complete representation of the problem would require the visualization in a space having a number of dimensions corresponding to the number of the variables. If we have more than three dimensions we are not able to make this visualization.
Moreover an univariate graphical analysis is unable, in most cases, to provide detailed information about the relationships among the samples and about the most informative variables; in fact, the relevant information is associated to the changes of a variable in respect to the others (namely the way the variables co-vary)
The principal component analysis (PCA) allows to identify, starting from the data, some privileged "directions" along which maximal information is focused. In this way we obtain some new abstracted variables, named principal components (PC). The PC number to be used to describe the data

will be much lower than the starting determinations, reducing the number of the space dimensions. But how these new variables can be obtained? Each principal component is a linear combination of the experimental variables expressed by the following relations:

$$PC1 = a_{11}x_1 + a_{12}x_2 + ...+ a_{1c}x_c$$
$$PC2 = a_{21}x_1 + a_{22}x_2 + ...+ a_{2c}x_c$$
$$PCn = a_{n1}x_1 + a_{n2}x_2 + ...+ a_{nc}x_c$$

Where *a* is a coefficient called *loading,* that represents the contribute of each variable to the description of the considered principal component and x is the value of the measured variable. The coefficients assigned to each variable in relation to each PC are calculated in such a way to maximize the variance along that direction. Precisely, we calculate the direction along which we have the maximum of the variance (PC1). At this point the following component is calculated in order to explain the maximum of the residual "variance". The second principal component must be orthogonal to the first one. The process goes on until all the PC are calculate.
This process is comparable to a rotation of the axes in the dimensional space keeping still the points-samples in order to maximize the variance.
The variance along each PC is indicated with λ_i and is named *eigenvalue* of the principal component. So we can quantify the percentage of the total variance (%EV) explained by each PC.

$$\%EV(i) = \frac{\lambda_i}{\sum_{k=1}^{N} \lambda_K}$$

and the overall percentage of variance explained by the first j principal components.

$$\%CV(j) = \sum_{i=1}^{j} \%EV(i)$$

Now we have a number of PC equal to the number of starting variables and we must select which PC use. Many criteria are based on the eigenvalues: in

the table **3.2** the eigenvalue of the first component is almost 7, so this variable contains the information corresponding to 7 original variables. Moreover it explains 40.6% of the total variance. The last PC are uninformative and contain mostly the "noise".

Table 3.2: in the columns the number of the PC and the corresponding λ, % of the variance explained by the single PC and the overall percentage of variance explained by the used principal components.

PC	λ	%EV	%CV
1	6.9010	40.59	40.59
2	2.2290	13.11	53.71
3	2.0180	11.87	65.58
4	1.3320	7.84	73.41
5	0.9168	5.39	78.80
.....	....	...	...
14	0.0603	0.35	99.95
15	0.0071	0.04	99.99
16	0.0016	0.01	100
17	0.0001	0.00	100

Some criteria used to establish the number of PC to use are:

- *the mean eigenvalue criterion*: based on the consideration that all the PC characterized by an eigenvalue higher than the mean eigenvalue can be significant; often it is reduced to the choice of all the PC having an eigenvaue higher than 1.
- *the percentage of variance explained:* the operator choose to use the number of PC sufficient to explain the desired percentage of total variance.

In the space designed by the PC the samples have coordinates called *scores* while the variables' coordinates are called *loadings*.

The analysis of the loadings allows the interpretation of the analytical data and, in particular, of the relationships among the variables.
The plot of loadings allows the evaluation of the influence of each variables on the PC and the correlation among the variables.
Considering the graph of figure **3.4**, some variables, such as terpenes, give a positive contribution to the principal component 1, others, such as the unsaturated fatty acids, give a negative contribution and some others, such as the oleic acid (var 5) don't give any contribution.
The second principal component, on the other hand, explains the separation among the samples with a high content of oleic acid (var5) and those with a high content of palmitic and stearic acid (var7).

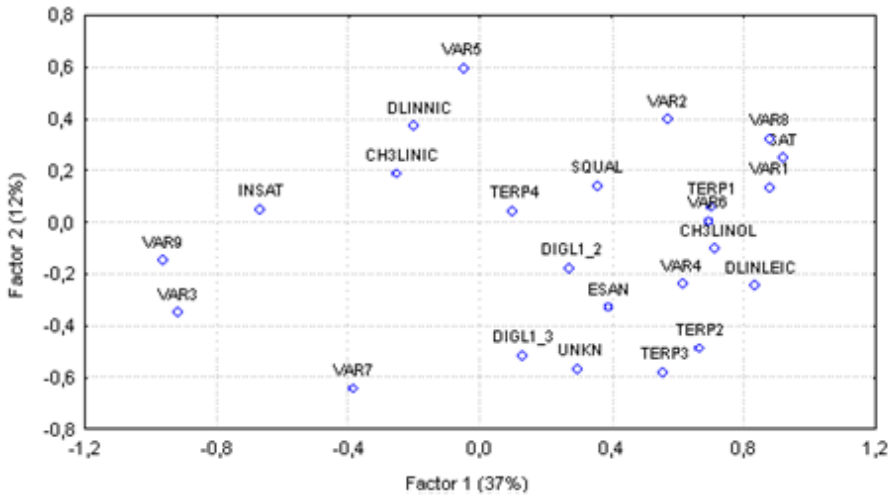


Figure 3.4: plot of loading

The relationships among the variables are purely phenomenological and don't mean any causality: we cannot say that a decrease of the oleic acid implies an increase of the palmitic acid, but we can say that along the PC2 the increase of the level of the palmitic acid goes together with the decrease of the oleic acid.
The PCA is used also for the research of the outliers or the treatment of the missing data in the matrix. Notably the PCA aims to maximize the separation between the objects, that is between the samples and not between groups, moreover there is no a priori assumption on the existence of groups.

3.6 Clustering analysis

The multidimensional data can present some grouping in the space that are not so easy to see. The identification of these groups is made possible by the *clusters.* This method is not to be confused with the classification methods, that works hypothesizing a priori the existence of the groups. On the contrary, in the cluster analysis the groups are not known a priori (actually we cannot even say that these groups really exist); the cluster analysis wants to identify the presence of groupings using the concept of *similarity.*[54]
It is necessary to associate a numerical size to the qualitative concept of similarity (or dissimilarity) among the objects. From an operative point of view, it is necessary to represent the dissimilarity between two samples using a distance measurement. In particular, since we can represent the samples as points in the space of the variables, it is also true that objects close to each others will have similar properties (small differences in the variables values).
There are different type of distances:

- the Euclidean distance

$$d_{st} = \sqrt{\sum_j \left(x_{sj} - x_{tj}\right)^2}$$

- the Manhattan distance

$$d_{st} = \sum_j \left| x_{sj} - x_{tj} \right|$$

- the Mahalanobis distance

$$d_{st} = \left(\mathbf{x}_\mathbf{s} - \mathbf{x}_\mathbf{t}\right)^{\mathrm{T}} \mathbf{S}^{-1} \left(\mathbf{x}_\mathbf{s} - \mathbf{x}_\mathbf{t}\right)$$

where S^{-1} is the covariance matrix

The first one can give some wrong results if the variables are correlated. This problem can be resolved by a distance measure that takes in consideration the correlation (Mahalanobis).
In addition to the chosen distance measure, the methods differ for the criteria of definition of the groups. In general there are two classes of methods:

- Hierarchical (figure **3.5**)
  - *agglomeration methods*: initially each observation (sample) is considered as a separate cluster, the grouping of the samples proceeds by combining the pairs of clusters increasingly similar, until all objects are combined into a single group;
  - *divisive methods:* the approach is the reverse of the previous one. In fact, we start by considering all objects as belonging to a single cluster, which is gradually divided into smaller groups until each object represents a single cluster.

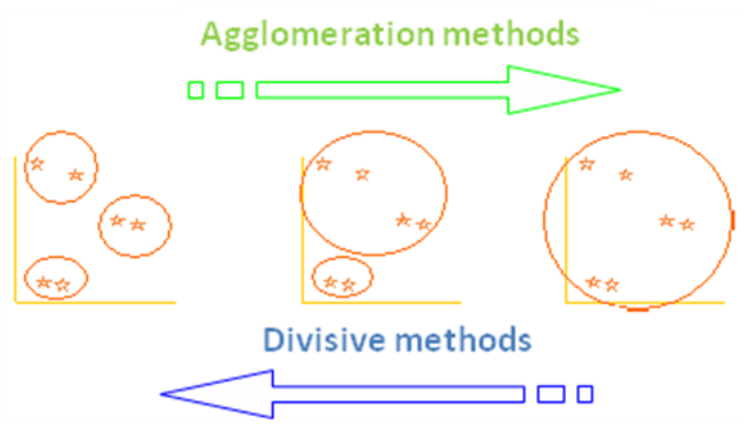


Figure 3.5: representation of the hierarchical clustering methods

- Non-hierarchical methods: these methods try to group the objects in a defined number of clusters, moving them from a cluster to another until the best grouping, according to the fixed number of clusters, is reached. (figure **3.6**)

Figure 3.6: representation of the non hierarchical clustering methods

Once established the method to use, there are still some criteria to define: how the individual samples or clusters can be further grouped and how the similarity

(or dissimilarity) between pairs of clusters can be calculated. Again there are different approaches:

- *Single bond method:* the distance among the clusters corresponds to the smaller distance between couples of samples of different clusters.
- *Complete bond method:* the distance between two clusters is defined as the largest distance between couples of different clusters.
- *Centroid method:* the distance between two clusters is defined as the distance between the centroids of two clusters;
- *Average bond method:* the distance between two clusters is defined as the average of the distances between all the possible pairs of samples in different clusters.

Once chosen the aggregation method for the clusters, we can represent the analysis by building a graph of the progressive union of the clusters, obtaining the plot of figure **3.7**, named dendrogram.

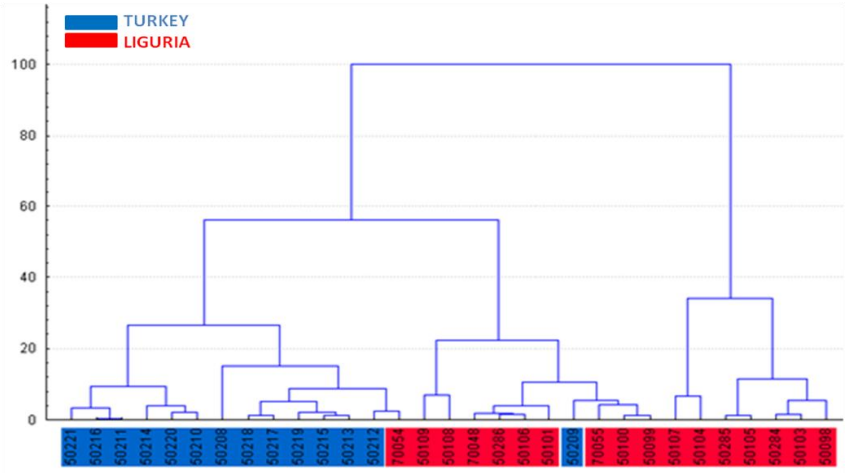


Figure 3.7: example of dendrogram obtained performing a TCA

The choice of the grouping criteria is fundamental, because the use of different criteria brings clusters with different number and composition.

3.7 Linear discriminant analysis

The classification methods have the target of building a model that can identify the belonging class of each sample on the basis of a certain number of independent variables.
Differing from the cluster analysis and the principal component analysis, the classification methods require that the belonging classes are established a priori and for each series of samples the corresponding class is known so it will be possible to build the model. The LDA is one of the simplest method to use, using the linear combination of the starting variables.
The first result of a classification analysis is the so called "confusion matrix". The cases are assigned with a certain probability to one of the classes, according with the Bayes rule (to obtain the best classification results, each sample has to be assigned to the class with the highest belonging probability.)
The samples with known assignment (training set) are used to calculate the coefficients of each variable to build the *classification function*

$$f_i(\mathbf{x}) = a_{i0} + a_{i1}x_1 + a_{i2}x_2 + \ldots + a_{ic}x_c$$

In table **3.3** a confusion matrix is reported: the rows contain the true classes, the columns contain the predicted classes. The example reports the olive oils from different geographic areas, submitted to LDA: 38 samples from the Latina province are correctly included in this class, while 2 remaining oils are more similar to those from Viterbo, thus they are included with a higher probability in a wrong class.

Table 3.3: confusion matrix of the samples from different geographic areas.

		Predicted classification				
		Latina	Frosinone	Rome	Viterbo	Rieti
Observed classification	Latina	38	0	0	2	0
	Frosinone	0	20	0	0	0
	Rome	0	0	19	1	0
	Viterbo	0	0	0	20	0
	Rieti	0	1	0	0	17

The classification matrix of the cases (table **3.4**) presents in the second column the true belonging class of each samples, in the third one the class with the highest probability of attribution. In the remaining columns the attribution with decreasing probability is shown. The samples 22 an 80 are classified with higher probability to the group of Viterbo rather than to the correct one.

Table 3.4: classification matrix of the cases; from the third column the collocation of the samples to the corresponding group has a decreasing probability.

Sample code	Observed class	LT	FR	RM	VT	RI
		1	2	3	4	5
1	LT	LT	VT	RM	FR	RI
2	LT	LT	VT	FR	RI	RM
3	FR	FR	RI	RM	VT	LT
4	LT	LT	VT	FR	RI	RM
5	LT	LT	VT	FR	RM	RI
6	RI	RI	FR	RM	LT	VT
…	…	..	…	…	…	…
*22	LT	VT	LT	RM	FR	RI
…	…	…	…	…	…	…
*80	LT	VT	LT	RM	FR	RI

*= samples that are attributed to classes different from those observed.

The attribution of a sample to a group rather than to another can be explained by looking at the matrix of the Mahalanobis distance of each sample from the centroid (vector containing the mean values of all the variables for each group), reported in table **3.5**. The samples 22 and 80 are closer to groups different from those observed, and so placed in classes different from the expected ones.

Table 3.5: Matrix of the Mahalanobis distance of the samples from the centroid of the group.

Sample code	Observed class	LT	FR	RM	VT	RI
		1	2	3	4	5
1	LT	53.25	144.56	1212.70	97.07	149.41
2	LT	23.35	82.12	103.08	56.50	101.91
3	FR	80.42	17.59	59.02	59.95	31.52
4	LT	10.42	44.03	60.02	29.63	54.68
5	LT	33.00	93.44	99.14	76.27	99.45
6	RI	85.46	28.98	62.37	84.23	22.37
…	…	…	…	…	…	…
*22	LT	34.34	56.69	47.16	26.63	71.90
…	…	…	…	…	…	…
*80	LT	25.90	60.94	59.83	23.45	69.11

*= samples placed in classes different from the expected.

The next step is to convert the numeric models in graphic models that can help the interpretation of the examined data set. To do this, starting from the classification functions we obtain the discriminating functions:

$$d_1 = c_{11}x_1 + c_{12}x_2 + \ldots + c_{1c}x_c$$

The coefficient of the linear combinations are determined similarily than the principal components. As discrimination criterion the Fisher F-ratio is used and we have to search the single direction with the highest value of this coefficient. This direction is known as first discriminant function. The next discriminant function is given by the direction perpendicular to the first and so on.

To establish which functions are to be used, we can calculate the variance for each function or use the *Wilks' Lambda* (the higher the value, the higher the discriminant power).

It is necessary to evaluate how the method can behave with samples not used during the building phase. This phase is named validation of the method. During this phase, a fraction of the samples is not used for the calculation of the model and is kept apart as an *evaluation set* or a *test set*.

The different validation techniques differ primarily in the manner in which objects are divided between training and evaluation set. Some examples are:

- *leave-one-out*: N-1 samples used to build the model, 1 sample left out of the model to test it;
- *Leave*-more-out: as the leave-one-out but leaving out a higher number of samples;
- *bootstrap*: N objects are extracted from the data set with the possibility to consider a sample more than once.

4. Materials and methods

4.1 Sample preservation

Olive oils have to be shelved in conditions that prevent the alteration of the product, protected from light, oxygen and heat in containers of the proper material.[55,18,56]
Olive oil samples are kept in a room safe from the light, minimizing the headspace (the volume full of air between the oil surface and the bottle cap), or, when possible, substituting the air with nitrogen, to reduce the oxygen content. In this way, we can prevent the autooxidation reaction, that is triggered principally by both these factors.
The temperature is included in the range 10-15 °C, to prevent the olive oil solidification and the consequent decrease of the phenols concentration at low temperature, and the triggering of the oxidative reactions at higher temperature.

4.2 Instrument and solvents

NMR 5 mm tubes Kontes (disposable type);
Gilson pipette, type Microman M25, 25 µL (range 3-25 µL);
Gilson pipette type Pipetman P200, 200 µL(range 30-200 µL);
Gilson pipette type Pipetman P1000, 1 mL (range 0.2-1 mL);
Deuterated Dimethylsulfoxide (DMSO-d6), 99,9% purity, from ARMAR (Döttingen, Switzerland), stored in 1 g vials at room temperature, safe from light and moisture;
Deuterated Chloroform (CDCl3), 99,8% purity, from Euriso Top (France), stabilized with silver.

4.3 Protonic NMR spectrum (^{1}H-NMR)

Samples preparation

Olive oil (20 µL), DMSO-d_6 (20 µL) and (700 µL) $CDCl_3$ are put in a 5 mm NMR tube. Tubes are then sealed with blowtorch. If only chloroform is used, the sample is not sufficiently stable, so the spectrum must be acquired immediately; if the DMSO is added the sample is more stable. However, the spectrum must be acquired within an hour, considering that the added solvents may cause degradative phenomena.
If the $CDCl_3$ solvent has a high acidity, changes of the chemical shift in the protonic spectrum are observed, preventing the right lecture of the resonances (figure **4.1**). In this case, a deacidification procedure must be performed.

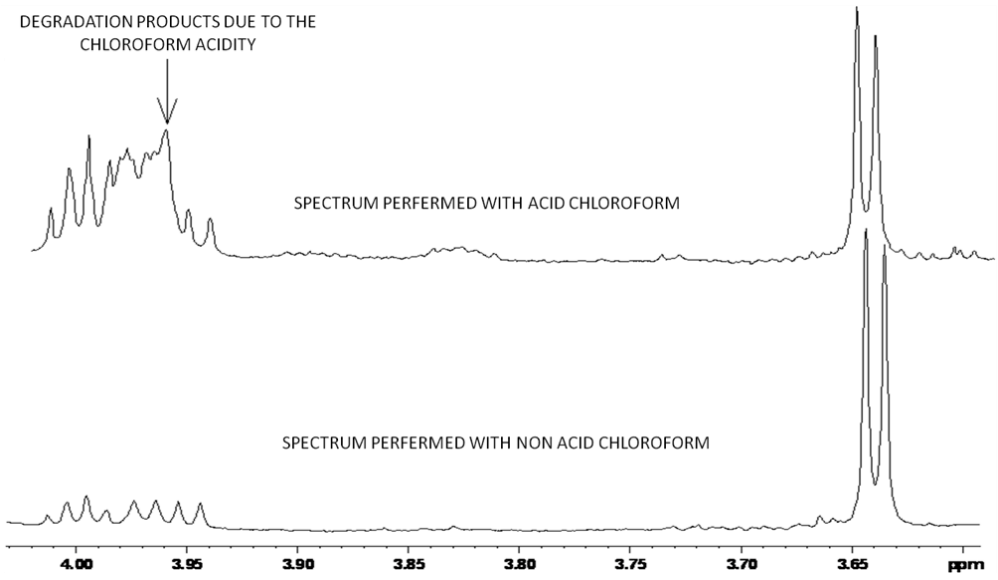


Figure 4.1: Comparison between spectra performed with acid chloroform and not acid chloroform

Even the position of the water signal can help to understand whether the chloroform quality is good or not: in the case of chloroform of good quality, the water signal must fall between 2.8 ppm and 3.6. If the chloroform is acid the water signal is down fielded.

Spectra acquisition

The ^{1}H-NMR spectra are performed on a Bruker AVANCE AQS600 spectrometer operating at the protonic frequency of 600.13 MHz (14.3 T). The instrument is equipped with a 5 mm probe for the protons (^{1}H), properly tuned on the characteristic protonic frequency of 600.13 MHz. Before the acquisition of the spectrum of a sample, it is important to control the homogeneity of the magnetic field. A solution with the 1% of acetone in $CDCl_3$ can be used as standard, looking at the $CHCl_3$ impurity always present in the $CDCl_3$ solvent. In an homogeneus magnetic field, the peak widths are equal to or less than the following values:

- half height width (Δν ½): 0.5 Hz;
- 5.5‰ height width: 8 Hz;
- 1.1‰ height width: 15 Hz

The $CHCl_3$ and $(CHD_2)_2SO$ present in the deutered solvents ($CDCl_3$ and DMSO-d_6) used for the sample preparation, gives in the ^{1}H spectrum, a signal at 7.260 ppm and 2.526 ppm, respectively. These signals can be used to verify the spectrum resolution: the $CHCl_3$ signal must be symmetric (figure **4.2**) while the signal of DMSO must be a quintet shaped (figure **4.3**).

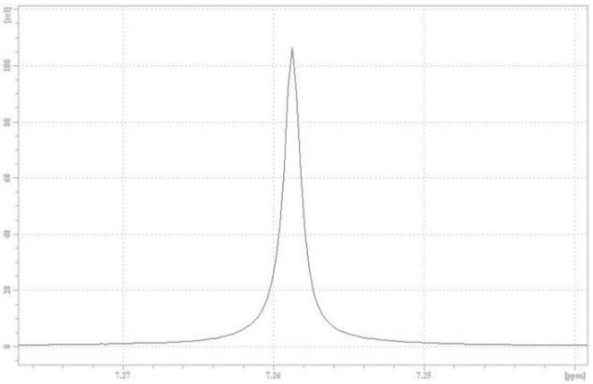


Figure 4.2: the chloroform signal ($CHCl_3$)

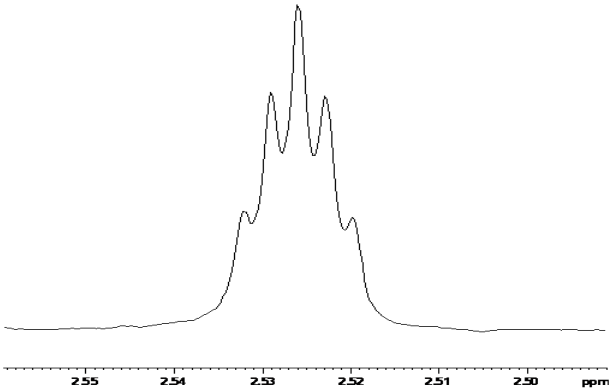


Figure 4.3: the dimethylsulfoxid signal

The ^{1}H spectra have been acquired using the following parameters:
Time Domain (TD): 64 K;
spectrum width (SW): 18.5 ppm;
O1: 4580 Hz;
Receiver Gain: 16;
90 degrees pulse;
Delay (D1): 0.5 sec;
Acquisition Time (AQ): 2.96 sec;
Number of Scans (NS): 1 K;
Dummy Scans (DS): 4;
Temperature: 300 K.

Spectra processing

The ^{1}H spectra are processed using the program TOPSPIN (vers. 1.3). using the following parameters:

- LB = 0.3 Hz (exponential function);
- SI = 64 K (zero filling);
- manual correction of the phase;
- spectrum calibration on the signal due to the methylenic protons in position 2 respect to the carbonyl (2.251 ppm);
- semi-manual correction of the base line (Cubic Spline Baseline Correction);
- Semi-manual reading of the intensity using a dedicated software (Sobolev-D'Imperio). In order to use this software is necessary to set the value (in ppm) for each peak, the range (in ppm) within which the peak is present, the maximum value and the minimum value of the signal intensity, a value of reading sensibility.

Signals identification

The ^{1}H spectrum of an olive oil sample shows not only a set of resonances due to the most concentrated components, but also signals due to the minor components. Here, we report the signals selected in this work to perform multivariate statistical analysis.

Aldehydes

In figure **4.4** is shown the aldehydes spectral region:

- hexanal 9.704 ppm;
- *trans* 2-hexenal 9.454 ppm

The aldehydes are generated by the lipoxigenases. The action of these enzymes can give to the olive oil pleasant characteristics, since they are responsible of the formation of the aldehydes in the olive fruit.[57,58] On the other hand, acting on the olive oil they bring the formation of the compounds responsible of the rancidity.[59]

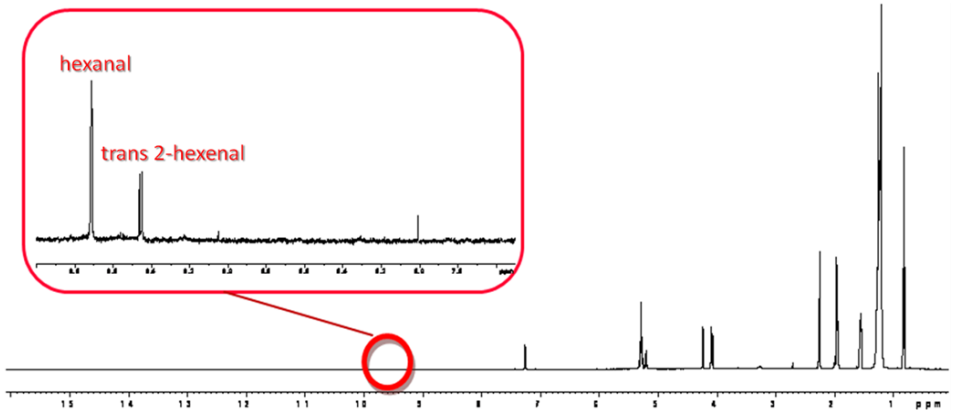


Figure 4.4: the aldheydic region

Terpens

In figure **4.5** is reported the terpens area:

- terpene 4 (4,887 ppm);
- terpene 3 (4,655 ppm);
- terpene 2 (4,602 ppm);
- terpene 1 (4,531 ppm).

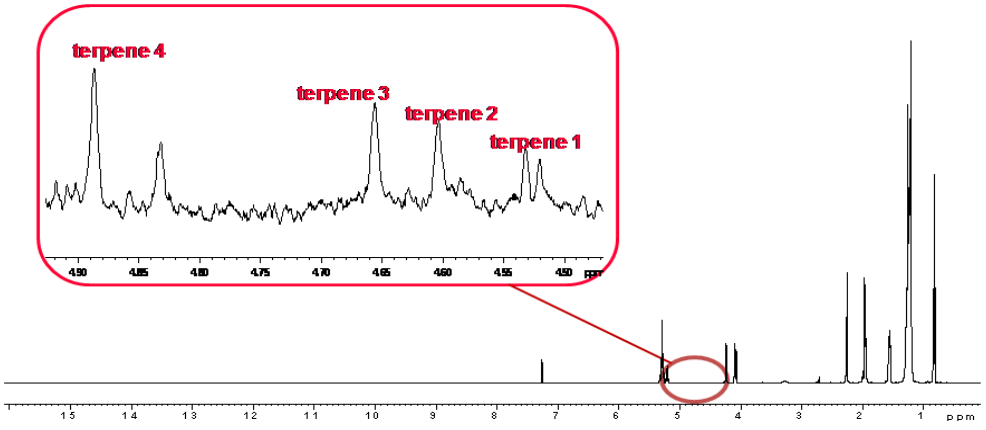


Figure 4.5: the terpenes area

Diglycerides

In the olive oil the triglycerides are degradated by the lipase to 1,2 and 1,3 diglycerides and their concentration depends on the ripeness degree of the olives and on the storage condition of the olive oil.

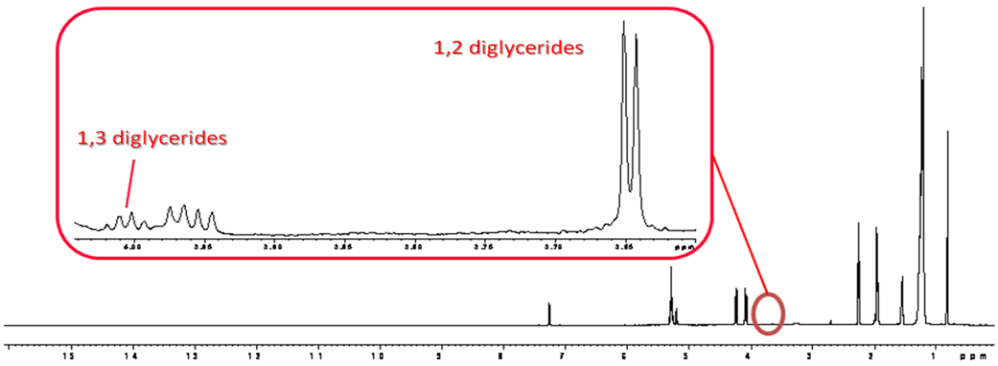


Figure 4.6: the diglycerides area

In the ^{1}H NMR spectrum, two signals due to these compounds are well visible:

- methylenic proton in β-position of the chain linked to the glycerol (*sn* 1,3 diglycerides) (3,997 ppm) (quintet);
- methylenic proton in position 3 of the glicerol (*sn* 1,2 diglycerides) (3,645 ppm) (doublet).

Diallylic protons of polyunsaturated fatty acids

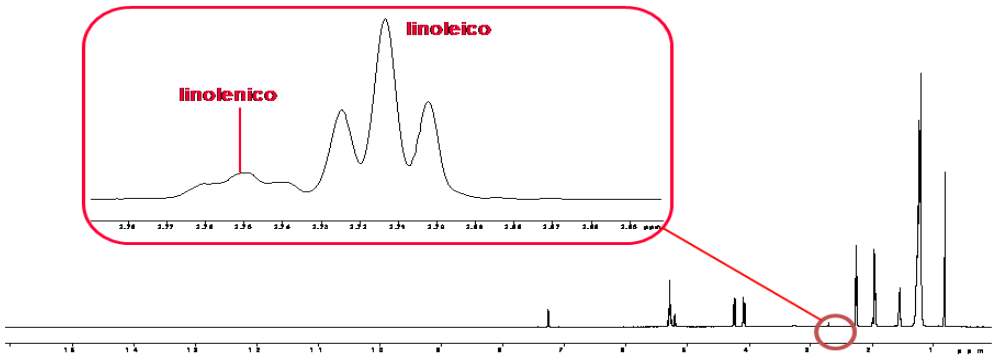


Figure 4.7: Diallylic protons of ^{1}H spectral region of an olive oil

- diallylic protons of the linolenic acid fatty chains (2.746 ppm);
- diallylic protons of the linoleic acid fatty chains (2.710 ppm).

In particular the determination of the abundance of linolenic and linoleic acids in olive oils is very important because of their role in the identification of the adulteration of the olive oil by the addition of hazelnut oil.[30,60]

Protons of the methylenic group in α to the carboxylic group of the fatty acids

Signal at 2.251 ppm (figure **4.8**) used during the processing for the horizontal calibration of the spectrum and for the normalization of the read signals.

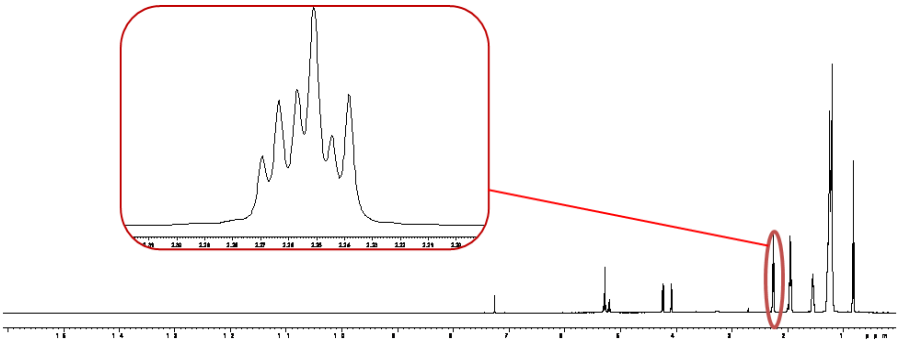


Figure 4.8: reference peak (^{1}H NMR spectra) due to methylenic protons bound to C2 normalized to 1000

Squalene

The signal at 1. 621 ppm reported in figure **4.9** is due to the methyls 17 and 18 of the squalene. It's an important signal because its absence is index of refining or adulteration of olive oils.

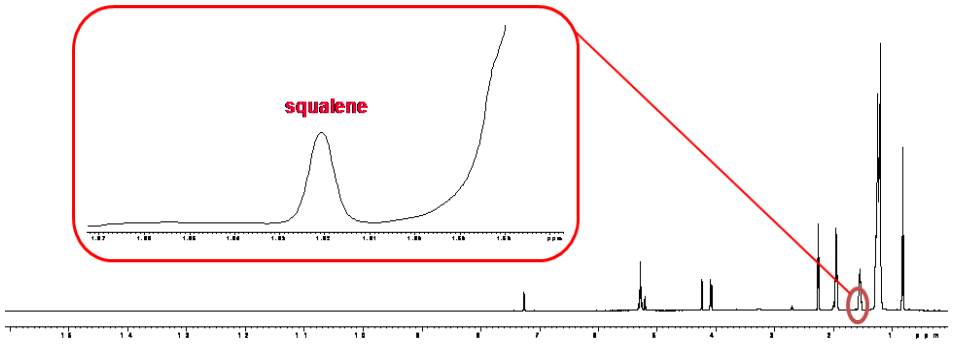


Figure 4.9: signal of the methyls 17 and 18 of the squalene

Methylenic protons beyond the position 2 of the β position of the fatty acids chains

As shown in figure **4.10**, it is possible to distinguish the resonance of the protons of:

- unsaturated fatty acids (1.245 ppm);
- saturated fatty acids (1.198 ppm).

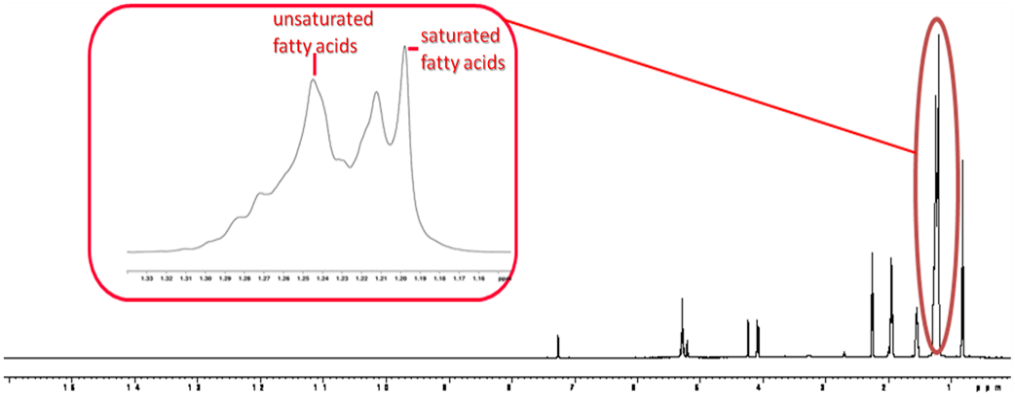


Figure 4.10: methylenes beyond the β position of the saturated and unsaturated fatty acids chains

Wax

Waxes (esters of long chain fatty acids with high molecular weight alcohols) are substances present in the exterior part of the fruit, with a protective function. When the olives are processed, a fraction of these waxes goes into

solution. If the wax content exceeds a certain value, it means that the oil has been adulterated with the olive-sansa oil.
In figure **4.11** the signal assigned to these compounds (0.978 ppm).

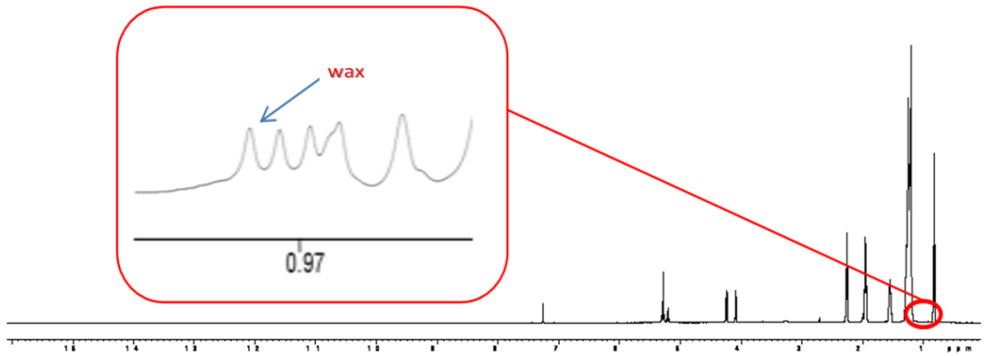


Figure 4.11: signal due to the wax

Protons of terminal methyls of the fatty chains

In a NMR spectrum the protons of the terminal CH_3 of the fatty acids have different resonance frequencies:

- protons of terminal CH_3 of the Δ-3 polyunsaturated chains, such as the linolenic acid (0.918 ppm);
- protons of terminal CH_3 of the mono-unsaturated and saturated fatty acids chains, such as the stearic, oleic and linoleic acid (0.844 ppm);

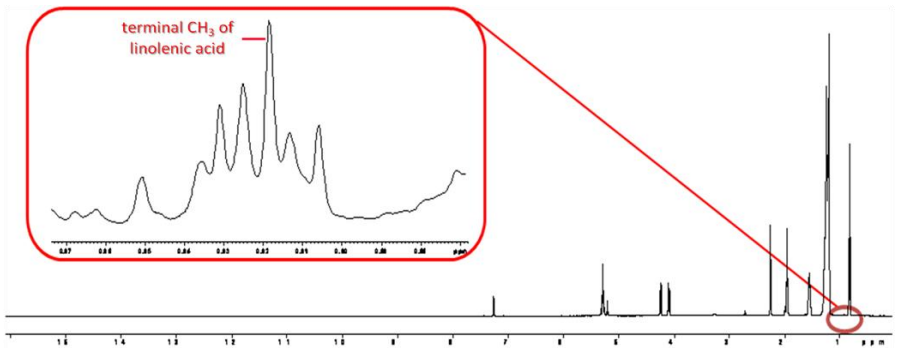


Figure 4.12: signal due to the terminal CH_3 of the linolenic fatty acid

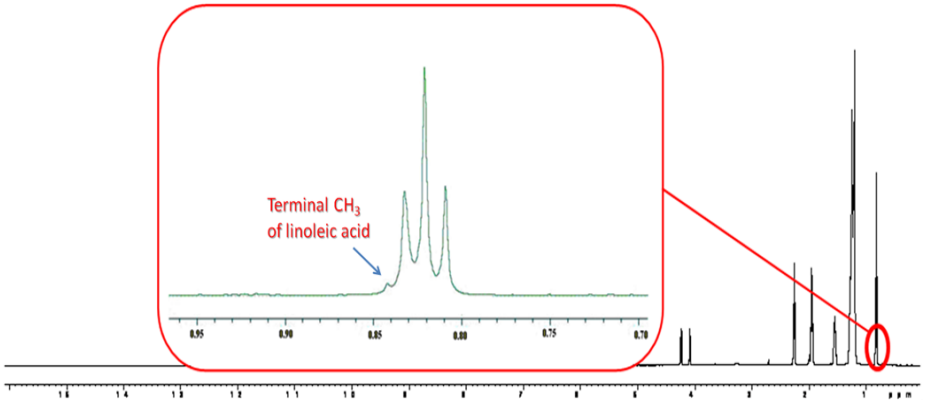


Figure 4.13: signal due to the terminal CH_3 of the linoleic fatty acid

β-Sitosterol

In the vegetable species, the sterols have a fundamental importance, since their composition is related to the botanic family of the plant. Their presence is shown by the presence signals of the methyls 18 in the region of the spectrum between 0.57 e 0.70 ppm. (figure **4.14**). In the olive oil the β-sitosterol is the principal steroid, identified by the signal at 0.624 ppm.

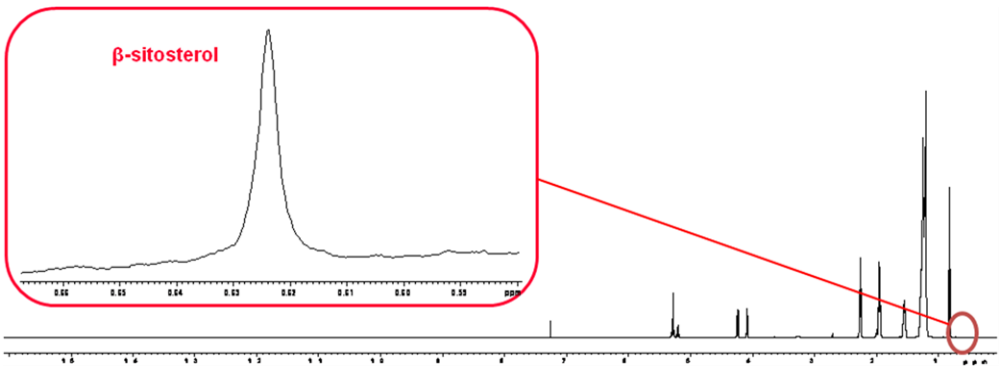
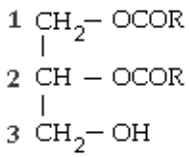


Figure 4.14: signal of the methyl of β-sitosterol 18.

Assignation of the mono- and di-glycerides in the ^{1}H NMR spectrum of the olive oil

In the olive oil there are triglycerides that are degraded by enzymatic hydrolysis to diglycerides; these are then hydrolyzed to mono-glycerides.
In an extravirgin olive oil there are no monoglycerides. Old oils or lampante oils present a concentration of diglycerides lower than the extravirgin oils and a appreciable concentration of monoglycerides
From here the interest in the determination of the ^{1}H NMR signals due to mono and diglycerides, using mono and bidimensional spectra of standard compounds such as:
1,3 dilinolenine mixed with 1,2 dilinolenine;
1-monoleina mixed with 2-monoleine.

<u>1,2 diglycerides signals</u>



The protons of the glycerol chain of the 1,2 diglycerides give 4 different signals (figure **4.15**):

- signal at 5,070 ppm of the proton in 2 of the glycerol;
- signals at 4,290 ppm and 4,170 ppm of each proton in 1 of the glycerol;
- signal at 3,645 ppm of the proton in 3 of the glycerol.

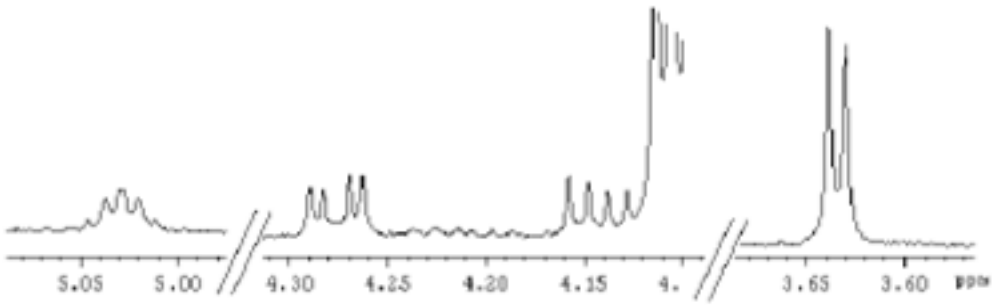
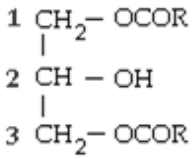


Figure 4.15: 1,2 diglycerides signals

Signals of the 1,3 diglycerides



The protons of the glycerol chain of the 1,3-diglycerides give 3 different signals (figure **4.16**):

- signals at 4,100 ppm and 4,060 ppm of one of the two protons in position 1 and 3 of the glycerol; notably the CH_2 protons in position 1 are not chemically equivalent because of the different neighbourhood and so they give two different signals. The same is for the CH_2 protons in position 3;
- signal at 3,999 ppm of the proton in position 2 of the glycerol. This proton couples with the protons in 1 and 2 of the glycerol with coupling constants almost equal, and its signal is a quintuplet.

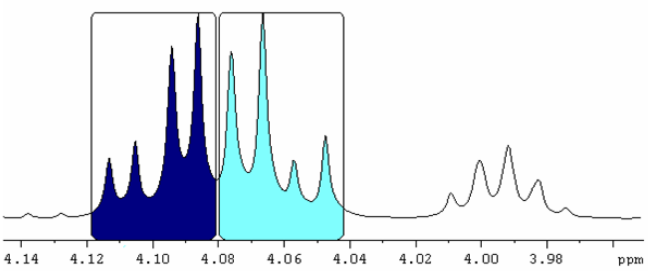


Figure 4.16: signals of the two protons in position 1 and 3 of the glycerol and the proton in position 2 of the glycerol

Notably in the olive oil ^{1}H NMR spectrum the signals at 4.100 ppm and at 4.060 ppm are covered by the doublet at 4.05 ppm due to the protons in 1 and 3 of the triglycerides (figure **4.17**). Moreover the signal at 3.999 ppm is not always a quintuplet, because partially covered by the ^{13}C satellite of the triglycerides at 4.05 ppm (figure **4.18**)

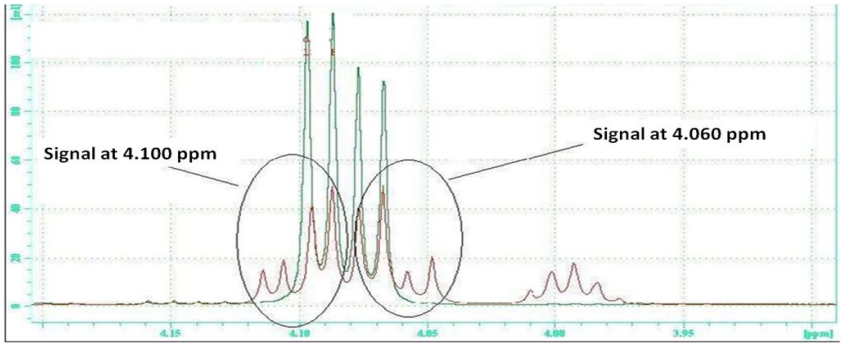


Figure 4.17: signals due to the protons in 1 and 3 of the triglycerides

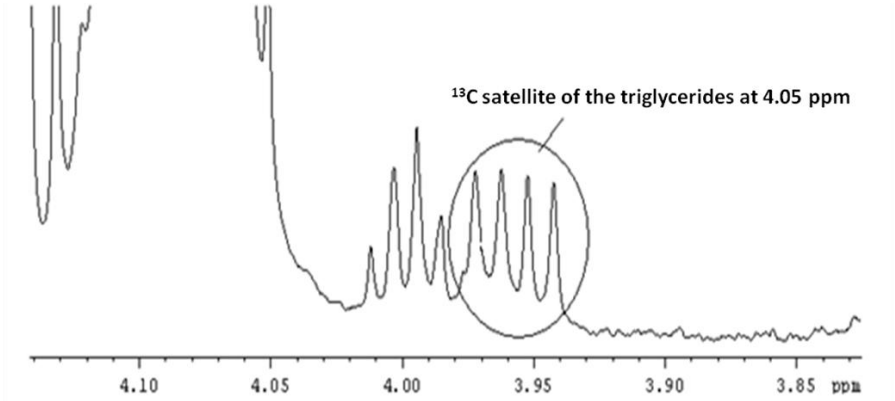
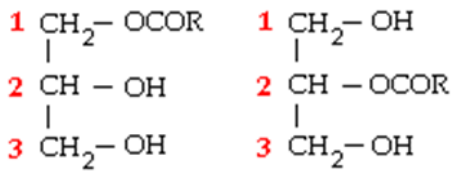


Figure 4.18: signal due to the proton in position 2 of the glycerol partially covered by the ^{13}C satellite of the triglycerides

Signals of the monoglycerides

Figure **4.19** shows in yellow the signals of the 1-monoglycerides and in green the signals of the 2-monoglycerides.



1-mono and 2-monoglycerides

The first give the following signals:

- signal at 4.070 ppm of the protons in position 1 of the glycerol. In the ^{1}H-NMR spectrum of a lampante oil this signal is covered by the signal of the triglycerides;
- signal at 3.873 ppm of the proton in 2 of the glycerol;
- signal at 3.604 ppm of one of the two protons in position 3 of the glycerol;
- signal at 3.509 ppm of one of the two protons in position 3 of the glycerol;

The 2-monoglycerides are characterized by the following signals:

- signal at 4.853 ppm of the proton in 2 of the glycerol. it is a signal difficult to be seen also in the ^{1}H-NMR spectrum of a lampante oil.
- signal at 3.740 ppm of the proton in position 1 and 3 of the glycerol.

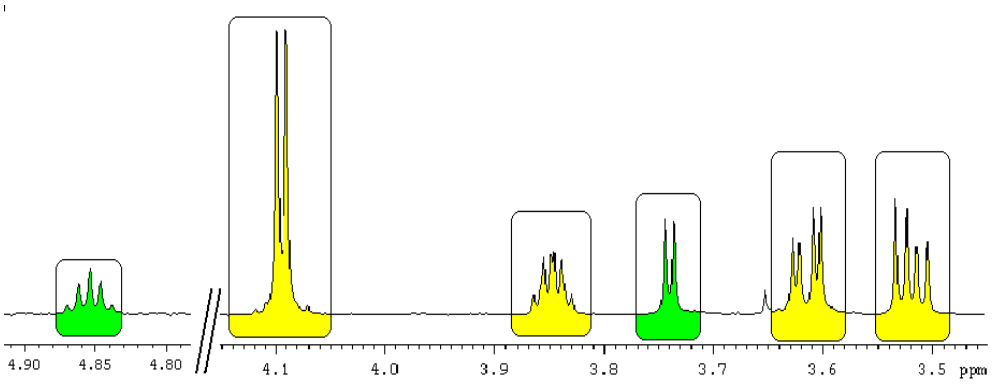


Figure 4.19: the signals of the 1-monoglycerides (in yellow) and the signals of the 2-monoglycerides (in green).

It must be remembered that the presence of these signals is principally due to the degradation of the triglycerides performer by the lipases, so this portion of the spectrum can explain whether the sample is an extravirgin oil or not. (Figure **4.20**)

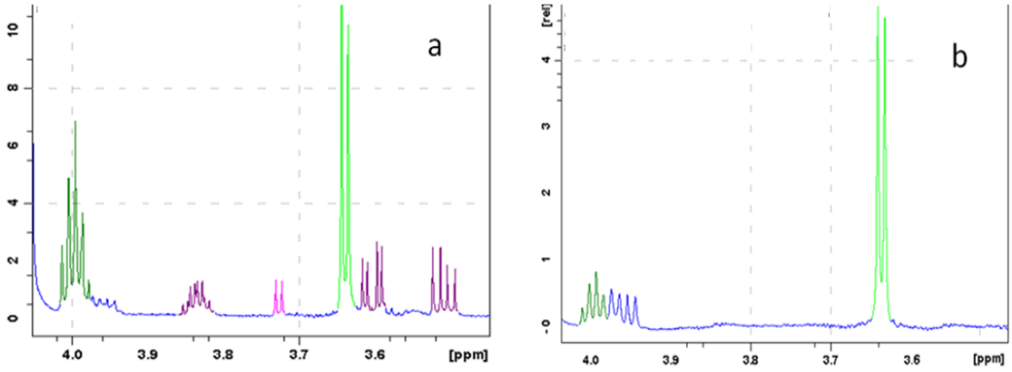


Figure 4.20: dark green: 1,3 diglycerides, light green: 1,2 diglycerides, purple: 1 monoglycerides, pink: 2 monoglycerides.

4.4 Statistical analysis applied to the signals of the ^{1}H-NMR spectra

The intensity of the signal to which we are interested (table **4.1**) is reported in an excel table and then normalized with respect to the reference signal, the peak of the methylene protons in β to the carbonyl (figure….); this signal, since independent from the nature of the fatty acids and from their position on the glycerol, represents the total amount of the fatty substance of the examined sample.
The intensity of each peak is normalized with respect to this peak, whose intensity is set equal to 1000. The normalization is a fundamental step for each statistical analysis because it allows to make significant comparisons among the samples of the data set. This procedure applied to the NMR spectrum allows to eliminate the variation associated to the total quantity of the sample analyzed and also to have values proportional to the molar ratio among the compound of the sample.

Table 4.1: signals of the ^{1}H-NMR spectra of the olive oil

N	Variable	NMR resonances (ppm)
1	Hexanal	9.704
2	T-2-hexenal	9.454
3	Terpene 4	4.885
4	Terpene 3	4.661
5	Terpene 2	4.609
6	Terpene 1	4.541
7	*sn1,3*-Diglycerides	3.988
8	*sn1,2*-Diglycerides	3.636
9	Linolenic ac. (diall.)	2.746
10	Linoleic ac. (diall.)	2.710
	Reference peak (α carboxyl protons of acyl chains)	**2.251**
11	Squalene	2.251
12	Unsaturated. F.A.	1.620
13	Saturated. F.A.	1.244
14	Wax	1.197
15	Linolenic ac. (CH3)	0.978
16	Linoleic ac. (CH3)	0.910
17	Sitosterole	0.843

The values of the normalized intensities of the signals are the variables considered in the statistical analysis. The obtained matrix, called raw matrix, undergoes the statistical explorative analysis and is necessary to obtain the histograms for each variable (signal of the spectrum) corresponding to the components of the oil. (table **4.1**)
The analysis of the histograms allows a direct comparison of the content of each compound in the considered samples and the visualization of particular trends for the variables. Already at this stage we can identify variables more or less important for the possible differentiation or classification according to the characteristics of the samples (i.e. the classification by cultivar,

geographic origin, altitude, etc.). The information obtained at this stage can be the starting point for the statistical multivariate analysis or can be used to find a confirmation of its results.
The matrices are then imported to the programs used for the specific statistical multivariate analysis (STATISTICA data analysis software system, Version 6.0, StatSoft, Inc. 2001).
With the Principal Component Analysis (**PCA**) we find the principal properties of the variables, and the data are simplified and reported in a bi- or tri-dimensional space. In this way we can visualize the samples distribution relative to two or three principal components that resume and include the contribute of the *n* starting variables.
The samples are distributed in the space following the directions that maximize the variance and, since no a priori hypothesis is made, we can visualize any natural grouping or separations of the samples that can be pointed out with a proper labelling of the samples.
The maximization of the separation among the samples the PCA can be performed also using only variables concerning compounds whose histograms pointed out a characteristic variance and that are probably more discriminant for a particular differentiation. For this kind of information we can also use the Analysis Of the Variance (**ANOVA**) that evaluates the discriminant power of the variables. The lower the *p. level* of each variable, the higher its discriminant power according to the desired discrimination. The highest reliability is considered as corresponding to a *p. level* minor than 0.05.
The Linear Discriminant Analysis (**LDA**) maximize the separation among the groups hypothesizing their existence a priori. The samples are included in a group with a certain limit of confidence (95%, visualized in the graph with the ellipses of confidence). The LDA can be performed also using the results of the ANOVA. In general, a preliminary investigation is required to discard any variables whose values are mutually dependent, since they could increase the noise, without giving any additional information. The correlations between the variables are evaluated by the correlation matrix using a parameter called tolerance, whose limit value is set as 0.05: if the value of the correlation index for two variables is higher than 0.95 they are correlated and one of them is going to be discarded.
The Clusters Analysis (TCA) is performed to obtain a classification of the samples in groups according to their similarity with no a priori hypothesis. Due to the strong influence of the numerous methods for calculating

distances and agglomeration of objects, generally this analysis is used as confirmation of the results of the PCA and LDA. The distances are calculated using the Euclidean method, while the clusters are grouped following the "*unweighted pair group*" method of linkage.

5. Results and discussion

This work has been carried out within a partnership between the Institute of Chemical methodologies of the CNR of Rome and UNAPROL (Italian Olive Consortium). The aim of this three years lasting collaboration is the "Realization of control systems to guarantee the respect of the standards of authenticity, quality and marketing of the Italian extra virgin olive oil". This program is part of a series of activities provided by the traceability program of UNAPROL (reg.ce.867/08), involving more than 470 chains and 6000 farms on the whole Italian territory.

The aim of this project is to create an analytical control system that can support the chain traceability systems, certified according the UNI EN ISO 22005/08 norm, allowing the certification and the valorisation of high quality Italian extra virgin olive oils.

UNAPROL has performed sampling during two harvesting years, 2009/10 and 2010/11, involving respectively 8 Italian regions (Lombardy, Tuscany, Sardinia, Apulia, Calabria, Sicily, Latium and Molise) with 153 samples for the first year, and 9 Italian regions (Lombardy, Tuscany, Sardinia, Puglia, Calabria, Sicily, Latium, Molise and Liguria) with 122 samples for the second year.

The aim of this part of my PhD thesis, carried out in the NMR laboratory "Annalaura Segre" (Institute of Chemical Methodologies – Area della Ricerca di Roma), was to perform ^{1}H NMR analysis of the UNAPROL olive oil samples and, by means of statistical elaboration of the data, to characterize these olive oils according to their geographical origin, genotype and pedoclimatic factors and also to collect some nutritional information.

The analysis protocol (see Material and Methods), developed in the CNR laboratory of Rome during the past fifteen years, consists in:

- ^{1}H-NMR analysis
- processing of the spectrum
- statistical analysis of the data

Once performed the ^{1}H-NMR experiments, the spectra were processed and, for each one, the intensity of 17 selected signals, corresponding to some characteristics components always present in the olive oil, was recorded.

The statistical treatment of the data includes a preliminary explorative analysis to identify anomalous data and similarities or differences in the

variables trend by histograms graphs, followed by a multivariate statistical analysis.
Following this protocol, it was possible to the study the influence of some parameters such as geographical, genetic, pedoclimatic, ecologic and agronomic factors on the olive oil chemical composition of the olive oils.
Here we report the results obtained studying the olive oils region by region and year by year, and also comparing the two harvesting years productions.
Because of the low number of samples collected in Tuscany (5 olive oils) and Lombardy (4 olive oils) during the 2009/10 harvesting year, the olive oils from these regions have not been taken in consideration in the statistical elaboration.
A statistical analysis has been carried out also for all the collected samples, to point out differences and similarities among the regions following different classification criteria.

5.1 Characterization of Italian olive oils

Italian Region	Number of Samples	
	2009/10	2010/11
Lazio	33	10
Puglia	32	31
Calabria	28	25
Sicilia	24	8
Molise	18	10
Sardegna	9	13
Toscana	5	9
Lombardia	4	6
Liguria	-	10

A first principal component analysis (PCA) was performed on the data set of both the harvesting years (Figure **5.1**), including samples from 8 Italian regions for 2009/10 and 9 region for 2010/11 (samples from Liguria were available only for the second year),and doesn't show a clear seasonal effect.

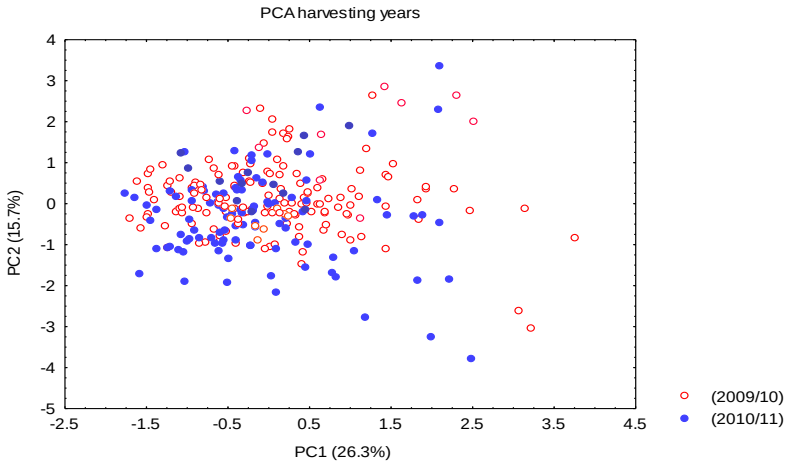


Figure 5.1: PCA of the entire database for both the harvesting years (2009/2010)

5.1.1 Peninsula vs islands

We carried out the same analysis labelling the samples according to the three geographic macroareas of Italy, i.e. mainland Italy, Sicily and Sardinia, (Figure **5.2**). In this way it is possible to recognize a clear seasonal effect for the islands and, though less evident, also for the samples from the mainland Italy.

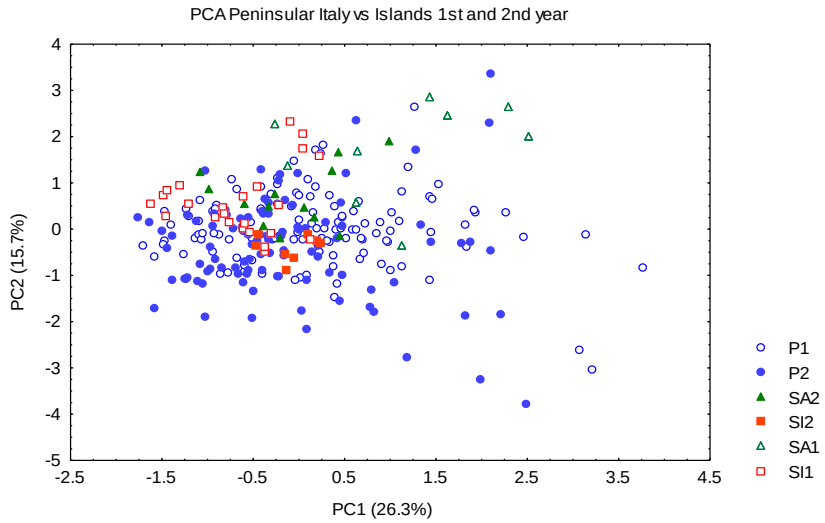


Figure 5.2: PCA of all olive oils from both the harvesting year (1,2) distinguished by geographic macroareas. Peninsular Italy: P1, P2; Sicily: SI1 and SI2; Sardinia: SA1, SA2.

Figure **5.3** shows the linear discriminant analysis (LDA) performed on the same data set, with a clear seasonal effect also for the mainland Italy samples.

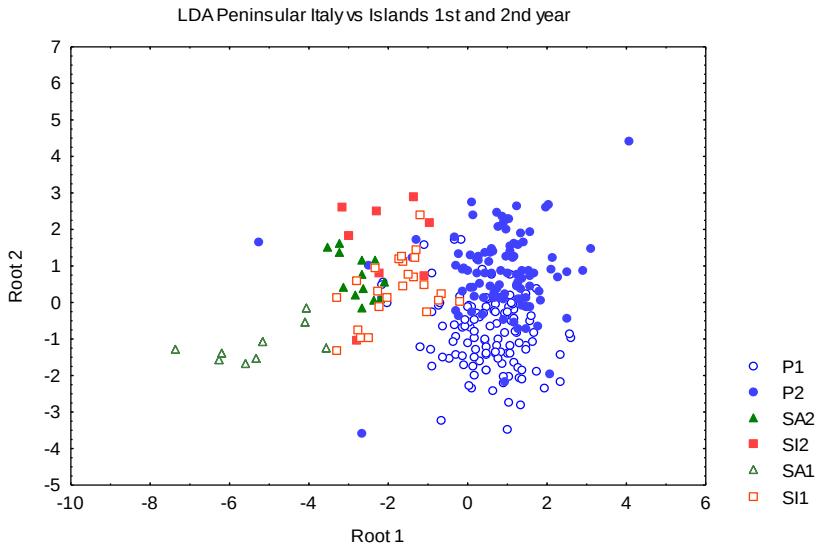


Figure 5.3: LDA of the olive oils from both the harvesting year distinguished by geographic macroareas: Peninsular Italy: P1, P2; Sicily: SI1 and SI2; Sardinia: SA1, SA2.

Repeating the LDA on the data sets of each year apart we can easily see how these three groups, Peninsula, Sicily and Sardinia, are clearly separated in both the harvesting seasons. (Figure **5.4** and **5.5**)

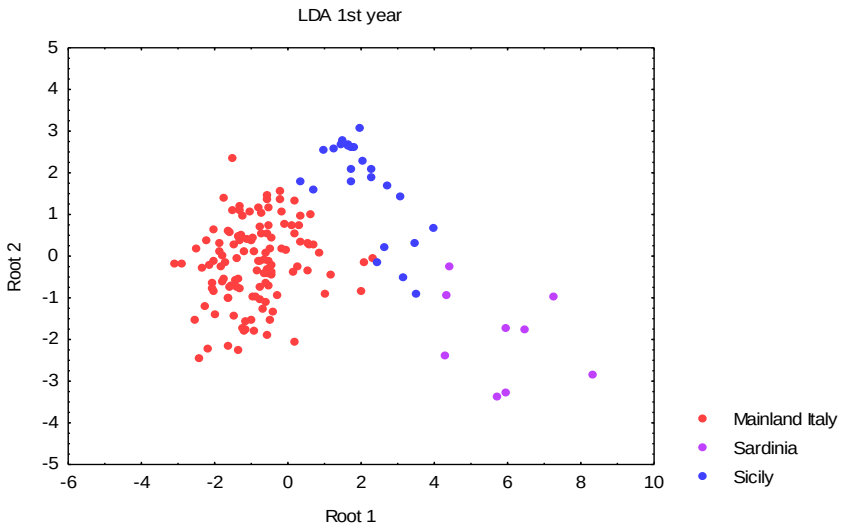


Figure 5.4: LDA of the olive oils from the first harvesting year distinguished by geographic macroareas (mainland Italy, Sicily and Sardinia).

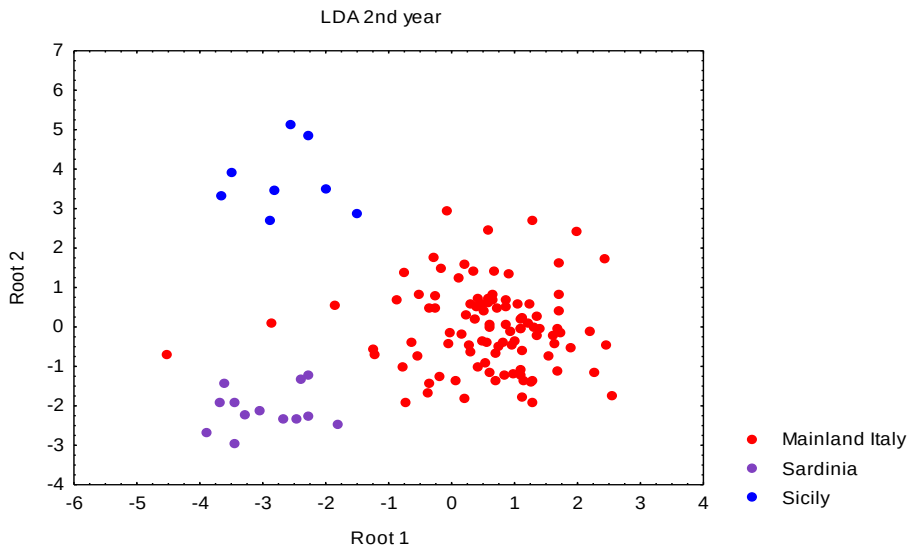
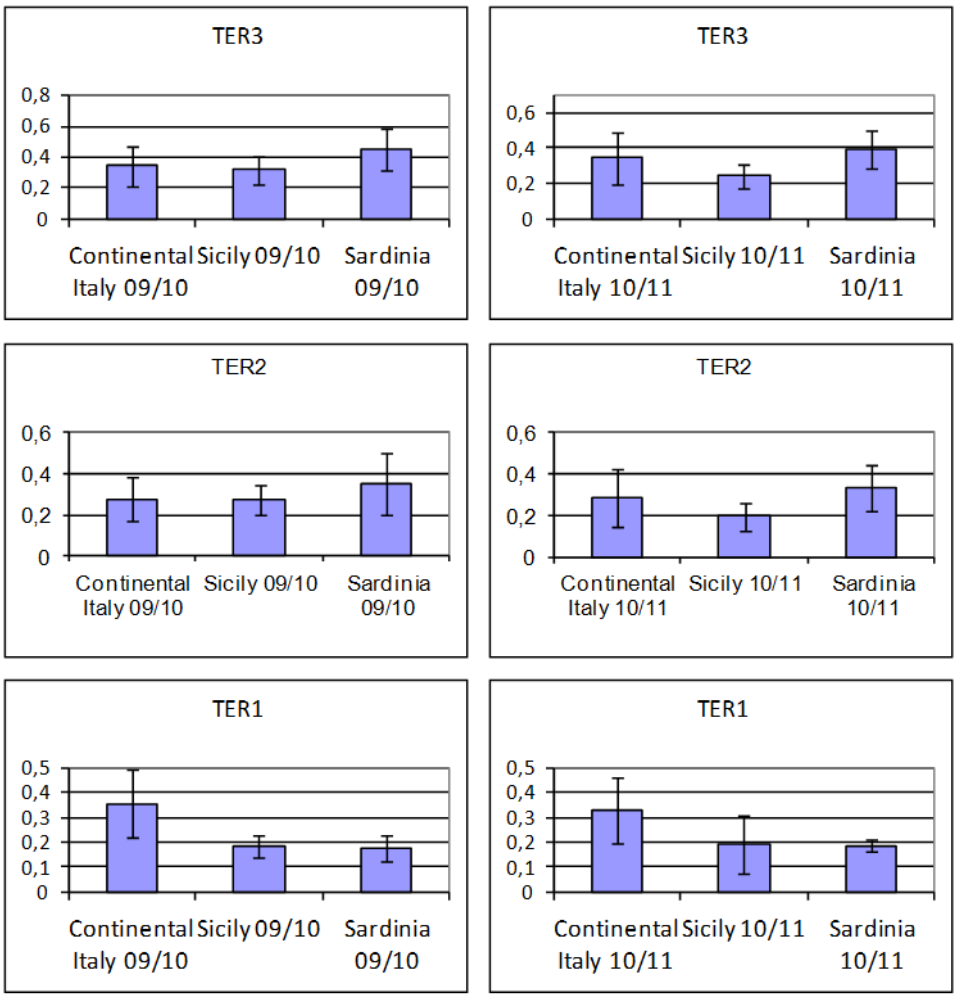


Figure 5.5: LDA of the olive oils from the second harvesting year distinguished by geographic macroareas (mainland Italy, Sicily and Sardinia).

This discrimination depends on the level of terpenes 3 and 2, highest in Sardinia and minimum in Sicily, terpene 1, that is maximum in the mainland Italy and very low in both the islands, and linoleic acid and squalene, minimum in the mainland Italy and maximum in Sardinia, as shown in the graphs of figure **5.6**



continues...

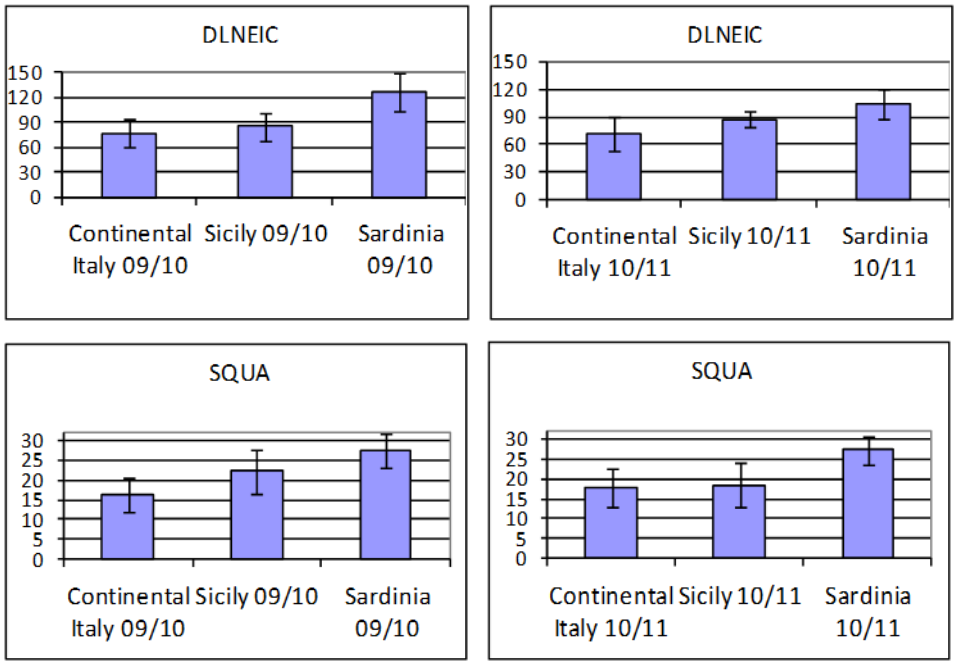


Figure 5.6: mean values of the signal of terpene 3, 2 and 1, diallylic protons of linoleic acid and squalene in the olive oils of the two harvesting years, distinguished by geographic macroareas (mainland Italy, Sicily and Sardinia).

The variables affected by the harvesting year are: the aldehydic compounds and terpene 4, low in the samples from Sicily and high in those from Sardinia for the first year, contrary to what appends in the second year, and linolenic acid, whose concentration in the peninsular samples increases in the second harvesting year. (Figure **5.7**)

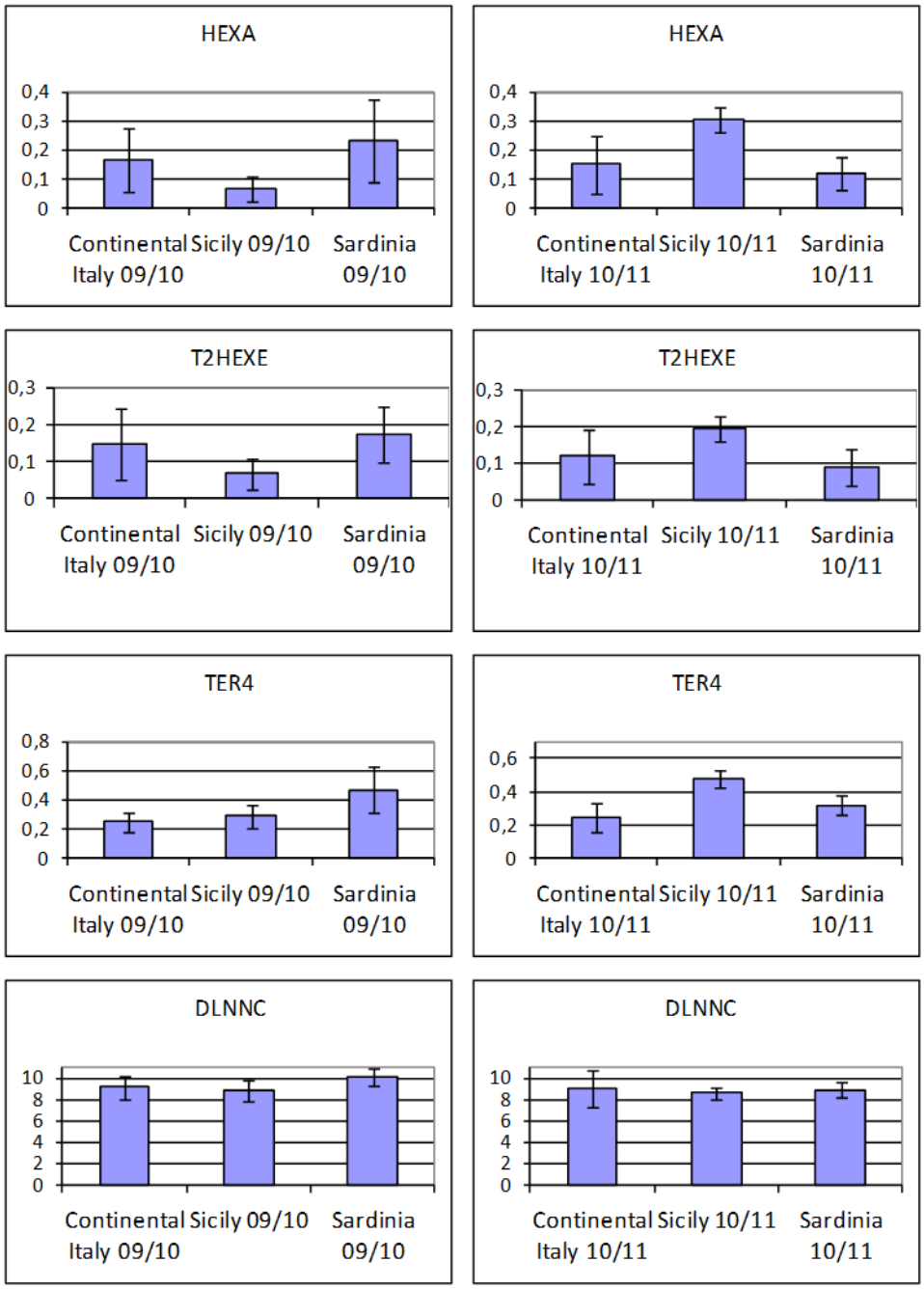


Figure 5.7: mean values of the signal of hexanal, T-2-hexenal, terpene 4 and diallylic protons of linolenic acid in the oils of the two harvesting years, distinguished by geographic macroareas (mainland Italy, Sicily and Sardinia).

5.1.2 North, Centre, South, Islands

We also tried to analyse the database according to four macroareas:

- North (Lombardy and Liguria, the latter only for the second year)
- Centre (Tuscany, Latium and Molise)
- South (Apulia and Calabria)
- Islands (Sicily and Sardinia)

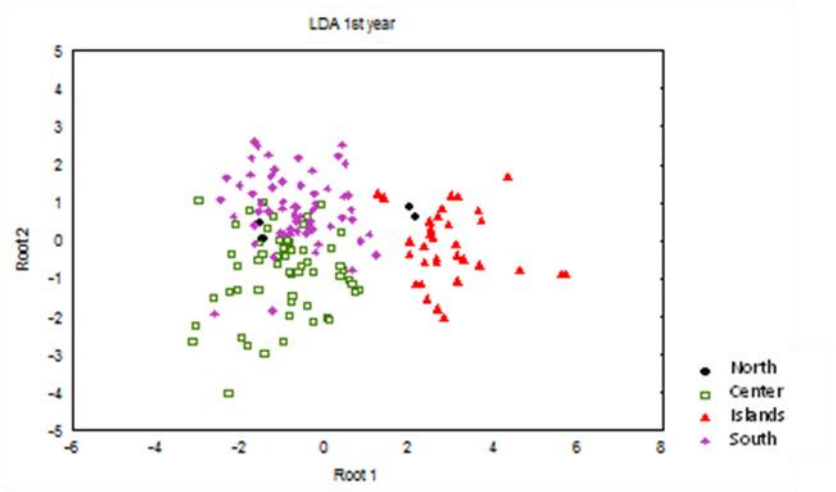


Figure 5.8: LDA of the olive oils from the first harvesting year distinguished by geographic macroareas (North, Centre, South, Islands).

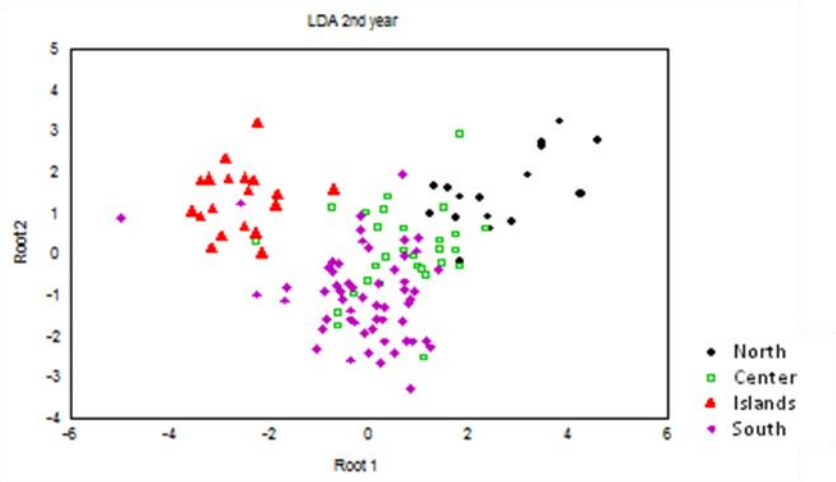


Figure 5.9: LDA of the olive oils from the second harvesting year distinguished by geographic macroareas (North, Centre, South, Islands).

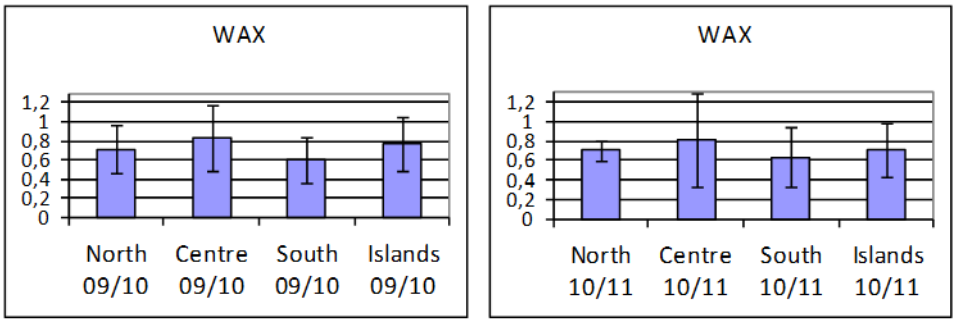


Figure 5.10: mean values of the signal of wax in the oils of the two harvesting years, distinguished by geographic macroareas (North, Centre, South, Islands).

The two LDA show a good separation among three groups. In both the years the olive oils from the Southern Italy present the lowest concentration of wax (figure **5.10**), while the olive oils from the islands show the highest contents of linoleic acid and squalene and the lowest level of terpene 1 (figure **5.11**). Notably the level of the linoleic acid shows an interesting trend: it has the lowest value in the Northern samples, increases in samples from the centre of Italy and in olive oil sample from Southern Italy and reaches the maximum in olive oils from the islands. This trend was observed for both the harvesting years.

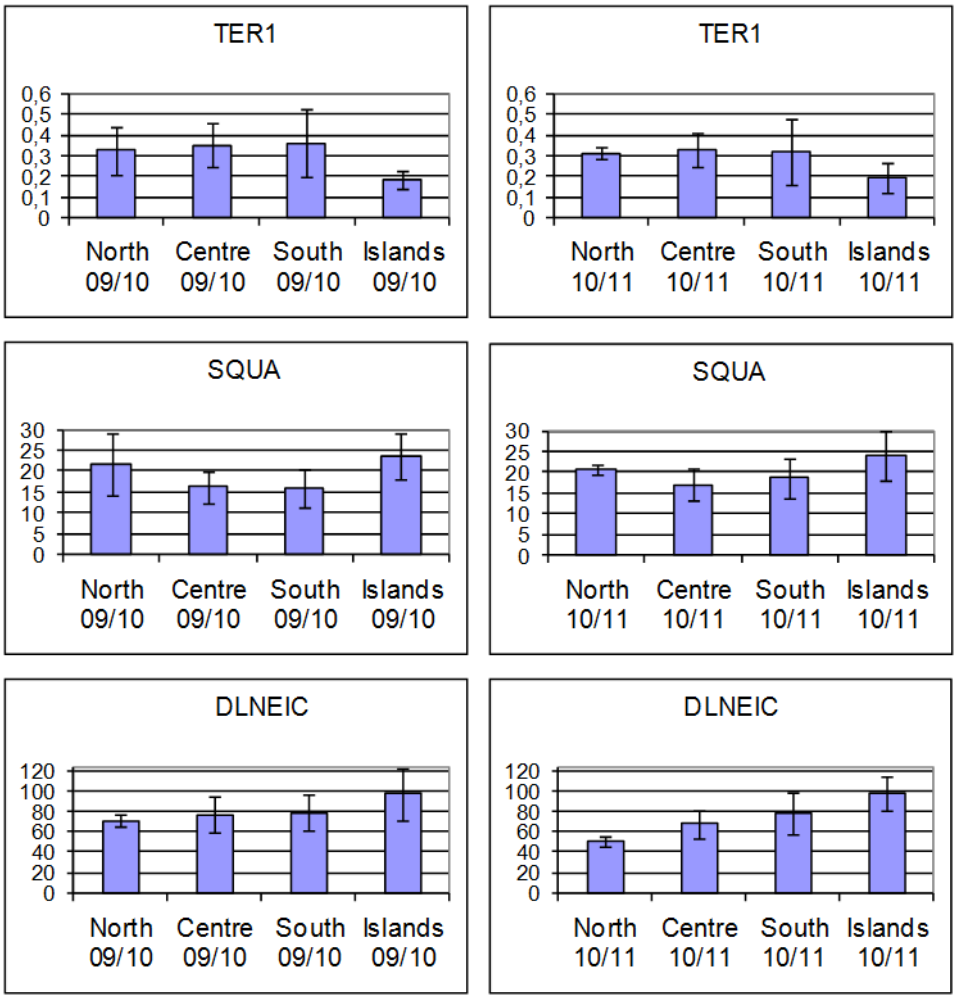


Figure 5.11: mean values of the signal of terpene 1, squalene and diallylic protons of linoleic acid in the oils of the two harvesting years, distinguished by geographic macroareas (North, Center, South, Islands).

5.1.3 Single regions and single harvesting years

Finally the LDA has been carried out on the olive oils from each harvesting years, and considering each region separately. The results are shown in figure **5.12** (first year) and **5.13** (second year).

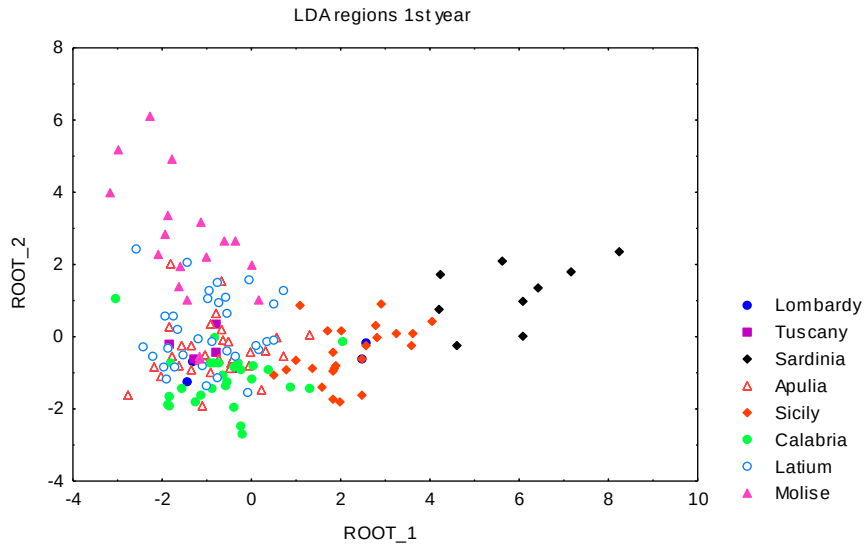


Figure 5.12: LDA of the olive oils of the first harvesting year, labelled by region.

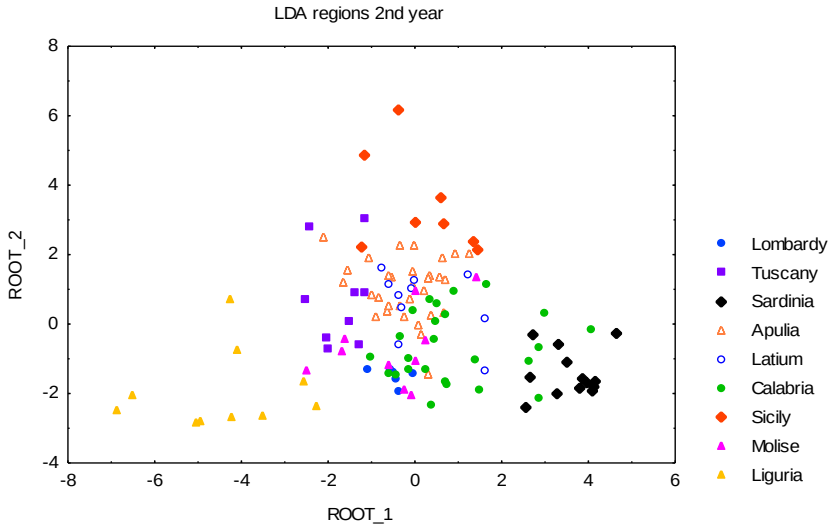
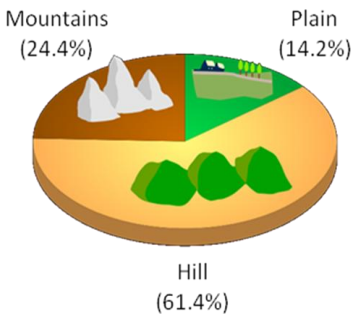


Figure 5.13: LDA of the olive oils of the second harvesting year, labelled by region.

In both the harvesting years the samples coming from Sicily and Sardinia are well separated from those from the peninsular regions. Also the olive oils from Molise and Calabria are quite separated from the others, however this effect is less evident in the second year.
In the second year samples from Liguria are well separated and totally distinguishable from the others.

5.2 Characterization of olive oils from Sicily



Olive trees reached this Island with the Phoenician between the IV and the VIII cent. b.C., and later with the ancient Greeks, after a century. The olive trees cultured reached the greatest diffusion, in Sicily as in the rest of the Mediterranean basin, during the Roman Empire. During the decay of the Roman Empire and the Turkish domination the olive culture was substituted by the culture of citrus fruits.
After the Norman domain, and during the Middle Ages, the olive tree culture regained its place, becoming a fundamental part of the Sicilian economy.[61]

Three POD are officially recognized for the Sicilian olive oils: Valli Trapanesi (Colour: green with golden highlights; aroma: clear aroma of olive, with some grass trace; flavour: fruity, lightly spicy and bitter), Val di Mazara (Colour: golden, with intense green highlights; aroma: fruity, sometimes with almonds notes; flavour: fruity, smooth, with a sweet aftertaste; area: Palermo and Agrigento), Monti Iblei (colour: green; aroma: intensely fruity aroma of freshly picked olive with a hint of tomato leaves and grass; flavour: medium fruity with persistent notes of green tomatoes and other vegetables, medium spicy).[62]

Table 5.1: oils from Sicily with the respective information

HARVESTING YEAR	CULTIVATION SITE	ALTITUDE (m above sea level)	CULTIVAR
09/10	Realmonte (AG)	200	Biancolilla
09/10	Realmonte (AG)	200	Biancolilla
09/10	Agrigento (AG)	180	Nocellara del Belice
09/10	Agrigento (AG)	180	Nocellara del Belice
09/10	Sciacca (AG)	80	Cerasuola
09/10	Sciacca (AG)	80	Cerasuola
09/10	Sciacca (AG)	190	Nocellara del Belice
09/10	Sciacca (AG)	190	Nocellara del Belice
09/10	Sciacca (AG)	140	Cerasuola
09/10	Sciacca (AG)	140	Cerasuola
09/10	Lucca Sicula (AG)	280	Biancolilla
09/10	Lucca Sicula (AG)	280	Biancolilla
09/10	Palazzo Adriano (PA)	670	Biancolilla
09/10	Palazzo Adriano (PA)	670	Biancolilla
09/10	Calamonaci (AG)	180	Biancolilla
09/10	Calamonaci (AG)	180	Biancolilla
09/10	Sciacca (AG)	250	Cerasuola
09/10	Sciacca (AG)	250	Cerasuola
09/10	Sambuca di Sicilia (AG)	200	Nocellara del Belice
09/10	Sambuca di Sicilia (AG)	200	Nocellara del Belice
09/10	Licata (AG)	100	Biancolilla, Nocellara del Belice
09/10	Licata (AG)	100	Biancolilla
09/10	Sciacca (AG)	100	Cerasuola
09/10	Sciacca (AG)	100	Cerasuola
10/11	Sciacca (AG)	120	Cerasuola
10/11	C.da Batia Sciacca (AG)	180	Cerasuola
10/11	Villafranca Sicula	130	Biancolilla
10/11	Comitini (AG)	270	Nocellara del Belice
10/11	Agrigento	300	Nocellara del Belice, Biancolilla
10/11	Sciacca (AG)	120	Cerasuola

10/11		270	Carolea
10/11	Aragona (AG)	400	Nocellara del Belice

5.2.1 Statistical analysis of the olive oils of the first year

The samples of the first year are 24, and only 2 of them come from Palermo. The others are from Agrigento.
They were sampled in pairs from the same locality and farm, differing by the date of harvest or milling. In the PCA the olive oils of each couple are close to each other, confirming the similarity of their composition. (figure **5.14**)

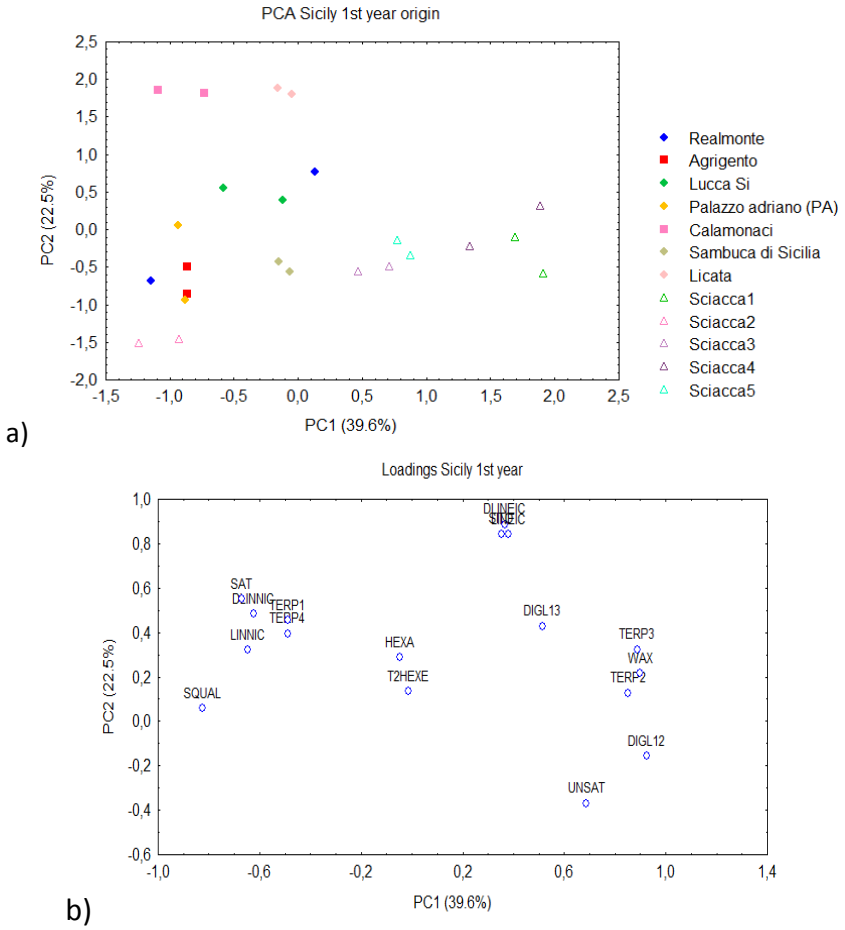


Figure 5.14: a) PCA of the olive oils from Sicily of the first harvesting year labelled according to their geographic origin. B) Plot of loadings of the PCA.

In a previous paper the usefulness of the NMR analysis for the distinction of different cultivars in the Sicilian olive oils has been reported, in particular referring to the different composition of fatty acids, detectable by ^{13}C-NMR experiments and gas chromatography.[36] In our case, interesting results using the ^{1}H-NMR analysis were obtained. In fact a good classification according to the cultivar criterion is obtained using PCA, TCA and LDA.

The PCA of figure **5.15** shows the differentiation of the samples labelled by the cultivar.

Olive oils of Cerasuola cultivar, clearly differentiated on the right side of the graph along the PC1 (39.6% of the total variance), are characterized by a very low content of squalene, saturated fatty acids and terpenes 1 and 4 and high concentration of unsaturated fatty acids, wax, sitosterol and terpenes 2 and 3.

The remaining two groups, whose olive oils are of Biancolilla and Nocellara cultivars, are placed in the left side of the graph, separated from each other along the PC2 (22.5% of the total variance). Both of them are rich of squalene and saturated fatty acids. Biancolilla olive oils present also a higher concentration of terpene 1 and terpene 4 and a lower content of unsaturated fatty acids, while Nocellara olive oils show lower concentration of sitosterol, wax, terpenes 2 and 3 and linoleic acid.

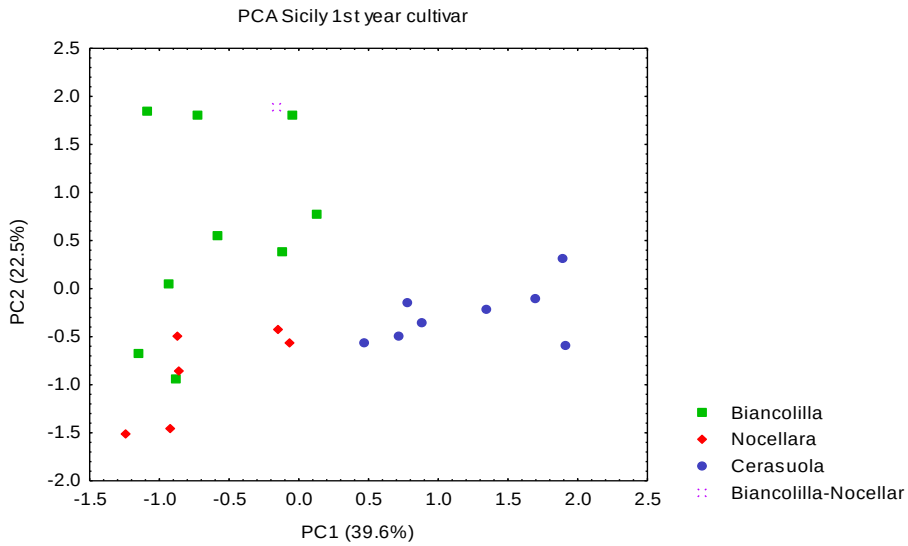


Figure 5.15: PCA of olive oils from Sicily of the first harvesting year, labelled according to the cultivar.

In the TCA we obtain a dendrogram where the samples are grouped by branches placed at different levels according to the similarity degree.
The TCA of the Sicilian olive oils (figure **5.16**) confirms the strong differentiation among the cultivars Cerasuola and the other two, Biancolilla and Nocellara. From the first branching they already set a totally independent group. The second branching separates a second group containing only samples of Biancolilla cultivar and the only bivarietal oil of Biancolilla/Nocellara cultivars, and a third group with the remaining samples of Biancolilla cultivar and all the olive oils of Nocellara Cultivar.

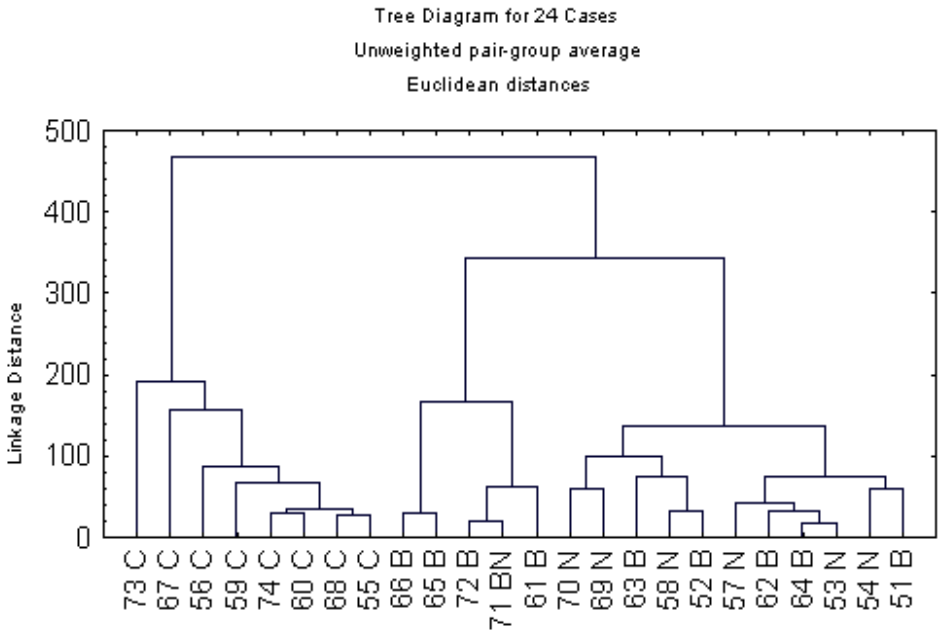


Figure 5.16: TCA of the olive oils from Sicily of the first harvesting year, labelled by cultivar (C=Cerasuola, N=Nocellara, B=Biancolilla).

The Linear Discriminant Analysis of the monovarietal samples provides a perfect separation of the three supposed groups (figure **5.17**). The group of the olive oils of Cerasuola cultivar is the most spaced from the others, but we can generally see an excellent sensitivity and specificity of the analysis, without superimposition of the confidence ellipses, nor external or shared samples.

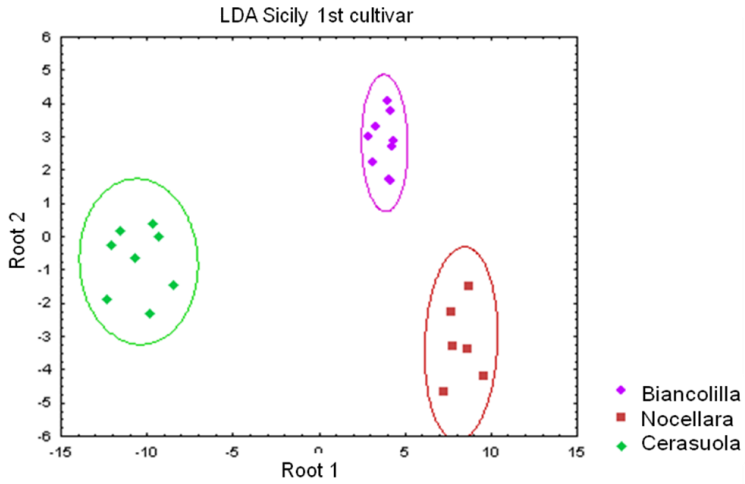


Figure 5.17: LDA of Sicilian olive oils of the first harvesting year, grouped by cultivar: Cerasuola, Biancolilla and Nocellara.

We also evaluated the effect of the ripening process on the olive oil composition, using the ripeness degree data available for all the Sicilian oils of this harvesting year. (table **5.2**)

Table 5.2: Ripeness degree of the Sicilian olive oils of the first harvesting year

Harvesting year	Ripeness Index	R.I.	Harvesting year	Ripeness Index	R.I.
09/10	2,10	2-3	09/10	2,70	2-3
09/10	2,10	2-3	09/10	2,95	2-3
09/10	1,40	1-2	09/10	1,85	1-2
09/10	1,40	1-2	09/10	2,70	2-3
09/10	3,34	3-4	09/10	2,11	2-3
09/10	3,36	3-4	09/10	2,29	2-3
09/10	1,71	1-2	09/10	2,63	2-3
09/10	1,83	1-2	09/10	2,76	2-3
09/10	2,18	2-3	09/10	2,35	2-3
09/10	2,39	2-3	09/10	2,56	2-3
09/10	2,55	2-3	09/10	3,20	3-4
09/10	2,31	2-3	09/10	3,10	3-4

As we can see in the PCA of figure **5.18**, the increase of the ripeness degree is associated to an increase of the terpene 2 and 3 and wax and to a decrease of linolenic acid and terpene 1. Moreover a minor ripeness degree corresponds to a lower concentration of sitosterol and linoleic acid, while a higher ripeness degree is associated to a higher level of *sn* 1,2 diglycerides and a low concentration of squalene and saturated fatty acids.

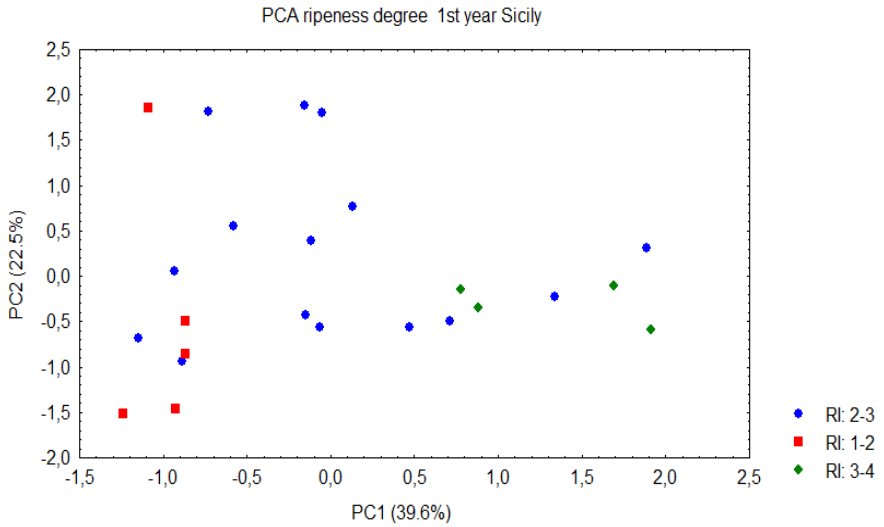


Figure 5.18: PCA of the oils from Sicily of the first harvesting year, labelled by the ripeness index.

5.2.2 Statistical analysis of the olive oils of the second year

All the 8 samples of the second harvesting year come from the Agrigento province, so it was possible to perform a PCA concerning the only genetic factor (figure **5.19**).

Only one of the 8 samples is a mix of two cultivars, all the others are monovarietal. The three samples from Sciacca are monocultivar Cerasuola, and they present a higher value of terpene 3 and lower level of squalene and saturated fatty acids, comparing to the other Sicilian samples. This is visible also in the olive oils of Cerasuola cultivar of the first harvesting year.

Notably the only bivarietal sample, with Nocellara and Biancolilla cultivars, is placed exactly between the monovarietal olive oils of these cultivars.

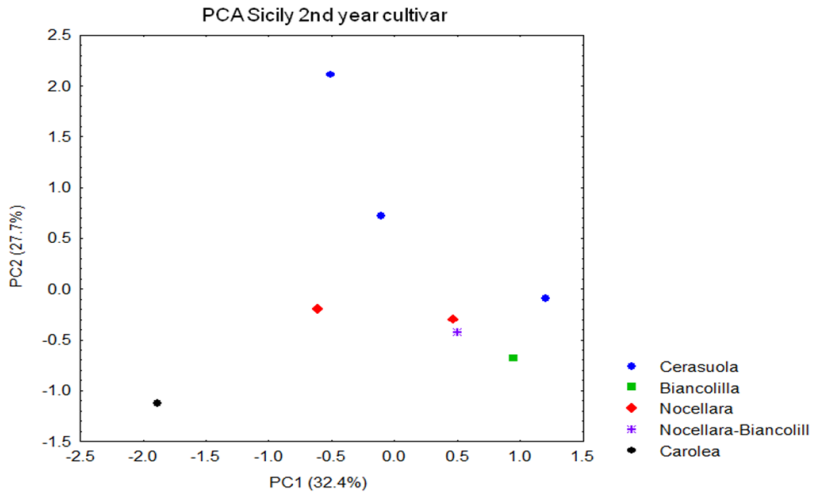


Figure 5.19: PCA of the Sicilian olive oils the second harvesting year, labelled according to the cultivars.

5.2.3 Comparison between olive oils of the two harvesting years

Comparing the two harvesting years we noticed that Sicilian olive oils present low concentration of terpene 1 and 3 and high level of linoleic acid (figure **5.20**), respect to the national mean values. The same results about terpene 3 and linoleic acid have been observed in the Sicilian olive oils of the three harvesting years monitored by the TRACE project, 2004/05, 2005/06 e 2006/07.[35] Moreover, still referring to the three season of the TRACE project and the two of the UNAPROL project, we noticed that, excepted the first UNAPROL season, 2009/10, when the terpene 2 concentration is similar to the national mean value, in all the other years its content is lower in the samples from this region than in the others.

Similarly the squalene in Sicily presents a higher value than in other regions during all the considered harvesting years excepted the last one, 2010/11, when its concentration is comparable to the national mean value.

The aldehydic compounds, instead, can't be considered as characterising variables for the olive oils from this region, being strongly affected by the seasonal effect: during the two UNAPROL seasons they increase their level, passing from the first to the second harvesting year, contrarily to what happens in the rest of the Italy.

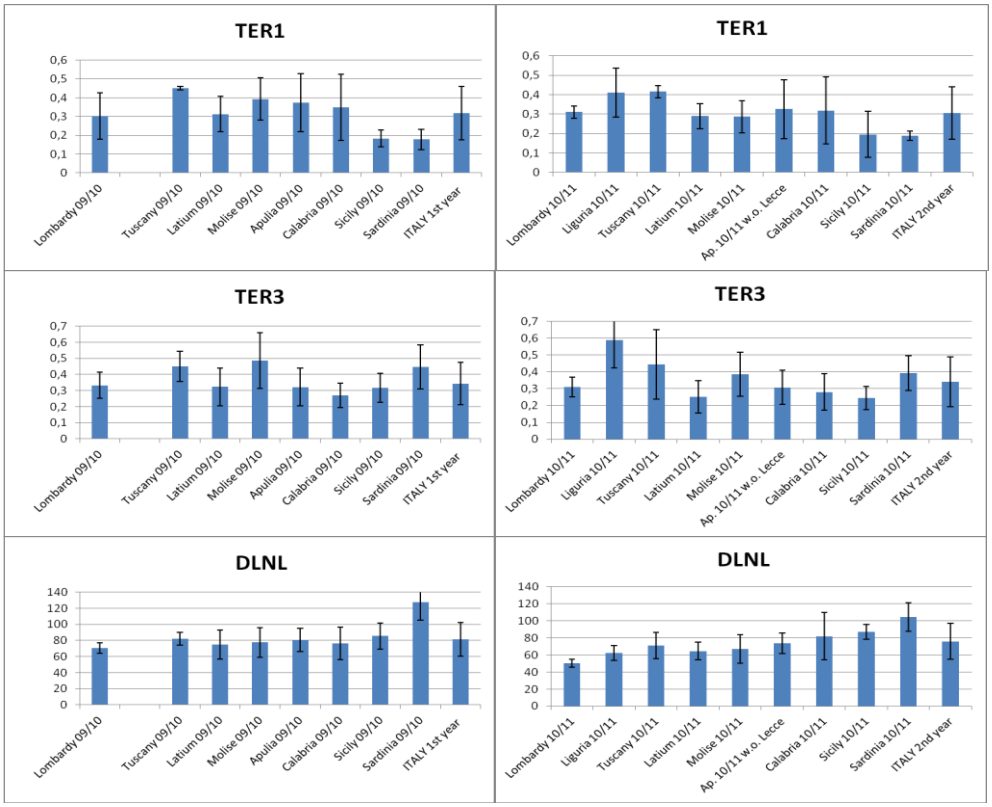


Figure 5.20: mean values of the signal of terpene 1, 3 and of diallylic protons of the linoleic acid in the olive oils from the different regions of Italy of both the harvesting years.

Performing a PCA on the olive oils of both the harvesting years we can clearly see a strong separation of the two groups, with no superimposition. (figure **5.21**)

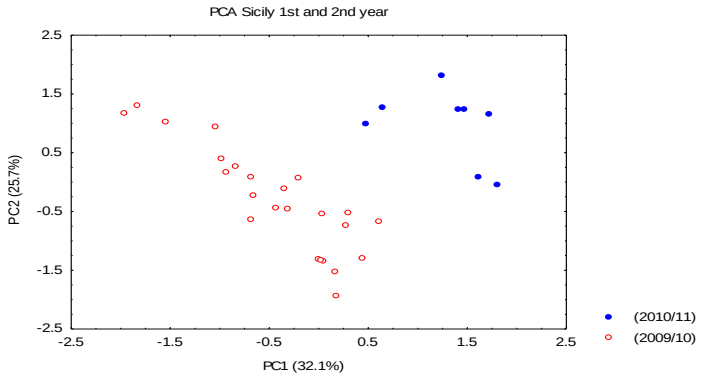
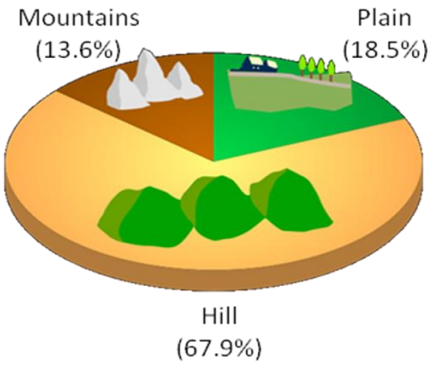


Figure 5.21: PCA of the olive oils from Sicily of both the harvesting years.

5.3 Characterization of olive oils from Sardinia



Olive trees are one of the most spread cultivation in Sardinia, involving more than 90% of the territory. Spain played a very important role since it has introduced olive tree cultivars such as Palma, Majorca and Sivigliana, and it contributed to the regulation of the culture.
The most famous native cultivars are: Tonda and Bianca di Cagliari, Pizz'e carroga, Bosana, Terza, Ceresia, Olianedda, Semidana and Corsicana.
The province of Sassari is the most productive, especially near Sorso, Sennori and Alghero. The typical cultivar in the Northern of the island is Bosana, also known as Tonda Sassarese. The olive oils from this cultivar have an intense aroma, more or less gentle; sometimes the flavour is sweeter, with notes of legumes, sometimes is more herbaceous.
In the Southern island, the Cagliari province, near Dolianova the cultivars Tonda di Cagliari and Pizz'e Carroga give extremely precious olive oils, with a good fluidity and a bold, but very fine aroma.
The POD "Sardegna" is reserved to the virgin olive oil produced in the selected areas of the region and it contains for the 80% the Bosana cultivar, Tonda di Cagliari, Nera (Tonda) di Villacidro, Semidana and their synonyms
The remaining 20% contains the other cultivars of the region, that must not modify the final characteristics of the product (colour: from green to yellow, varying with time; aroma: fruity; flavour: fruity with bitter and spicy notes).[63]

Table 5.3: oils from Sardinia with the respective information

HARVESTING YEAR	CULTIVATION SITE	ALTITUDE (m above sea level)	CULTIVAR
09/10	Sassari (SS)	350	Bosana
09/10	Sassari (SS)	500	Bosana
09/10	Alghero (SS)	0	Bosana
09/10	Sorso (SS)	100	Bosana
09/10	Sennori (SS)	350	Semidana 70%, Bosana 30%
09/10	Dolianova (CA)	100	Oliva da olio
09/10	Dolianova (CA)	100	Oliva da olio
09/10	Dolianova (CA)	100	Tonda di Cagliari
09/10	Dolianova (CA)	100	Tonda di Cagliari
10/11	Bancali (SS)	225	Bosana
10/11	Il Bosco, Sorso Sassari (SS)	136	Bosana
10/11	Sennori (SS)	277	Bosana 80%, Semidana 20%
10/11	Cliorras, Usini (SS)	200	Bosana
10/11	Fenujedda ittiri (SS)	400	Bosana
10/11	Alghero Alghero(SS)	0	Bosana
10/11	Bainza Pais Uri (SS)	300	Bosana
10/11	Sorso (SS)	136	Bosana
10/11	S. Maria di Lucardu (SS)	250	Semidana, Tonda di Cagliari, Bosana
10/11	Ittiri (SS)	400	Bosana
10/11	Loc. Mugori Dolianova (CA)	50	Tonda di Cagliari
10/11	Loc sv Pranu Dolianova (CA)	50	Tonda di Cagliari
10/11	Agrogergei Gergei (CA)	150	Maior, Tonda di Cagliari, Pizz'e Carroga, Nera di Gonnos

5.3.1 Statistical analysis of the olive oils of the first year

The 9 olive oils from Sardinia come from the Cagliari province (4 samples) and from the Sassari province (5 samples): 2 samples from Cagliari are olive oils from Tonda di Cagliari cultivar and 2 are generically identified as obtained from Oliva da Olio; 4 samples from Sassari come from Bosana cultivar whereas the remaining one is a bivarietal olive oil of Bosana-Semidana cultivar.

The differences between the two groups of olive oils are due to the geographical origin and to the genetic factor.

By labelling the samples by cultivar we obtained the PCA shown in figure **5.22**. The graph points out the clear separation of the samples along the PC1 (46.6% of the total variance): in the right side of the plot we find the samples from the Cagliari province, with a high concentration of squalene, linoleic acid and sitosterol, saturated fatty acids and terpenes 2 and 3. The samples from the Sassari province, with Bosana cultivar, are placed in the left side of the graph, clearly apart from the olive oils from Cagliari.

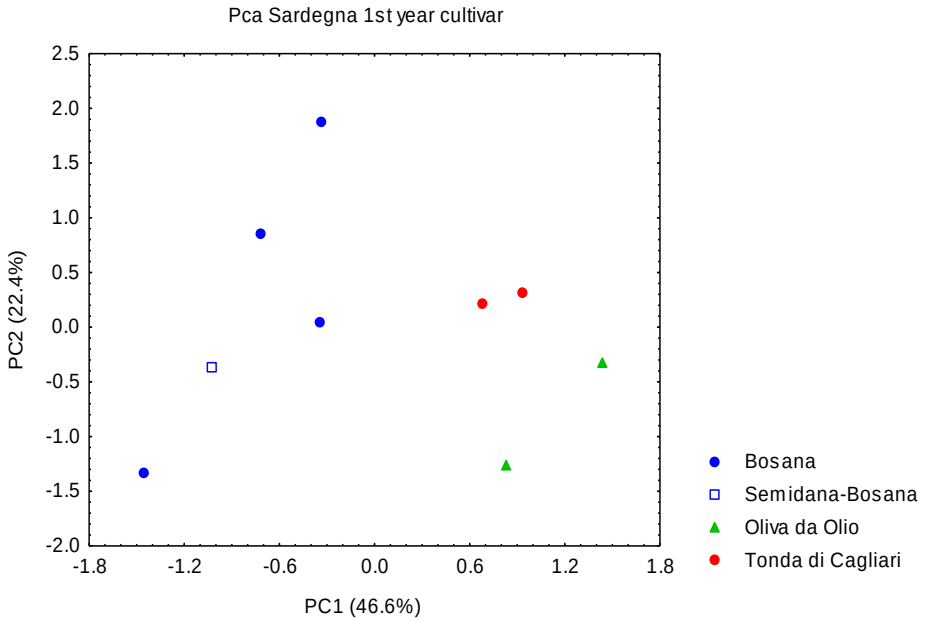


Figure 5.22: PCA of the oils from Sardinia of the first harvesting year, labelled by cultivar.

5.3.2 Statistical analysis of the olive oils of second year

Samples of the second harvesting year are from the Sassari province (10 samples) and from the Cagliari province (3 samples).
The PCA shown in figure **5.23** points out a clear separation between the two groups along PC1 (31.9% of the total variance)

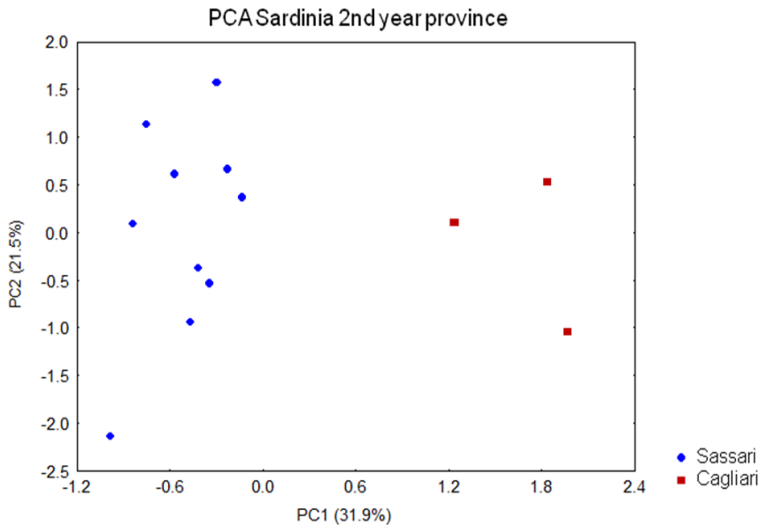


Figure 5.23: PCA of the oils from Sardinia of the second harvesting year, labelled by geographic origin.

As in the previous harvesting year the cultivar composition follows the geographic origin: the 3 samples from Cagliari contain the Tonda di Cagliari cultivar, 2 are monovarietal olive oils whereas one sample is a mixture of this cultivar with other cultivars. Only one olive oil from the Sassari province has a 30% of Tonda di Cagliari cultivar, the others contain the Bosana cultivar, totally absent in the olive oils of the first group.
Relabelling the samples by the cultivar, we notice in the same PCA (figure **5.24**) how the monovarietal olive oils of Bosana cultivar are completely apart from the monovarietal oils of Tonda di Cagliari cultivar.

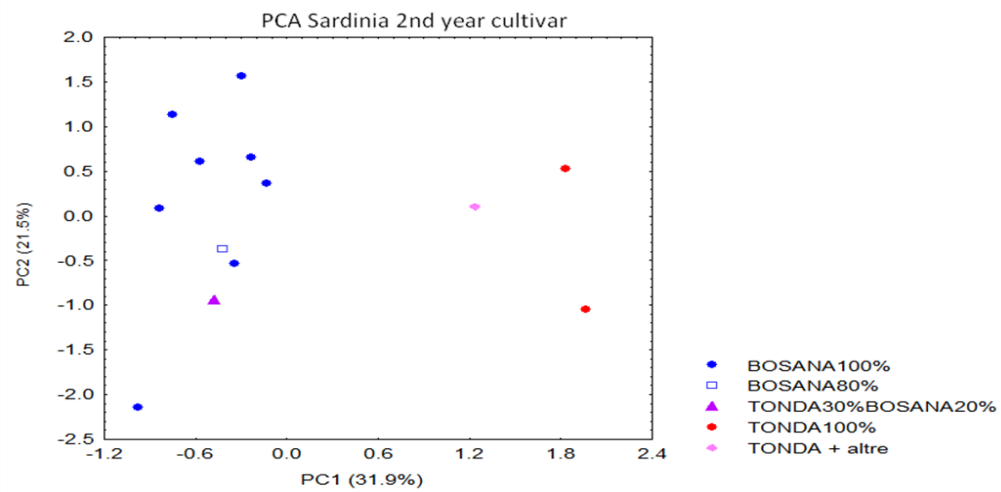


Figure 5.24: PCA of the oils from Sardinia of the second harvesting year, labelled by cultivar.

The olive oils coming from the Cagliari province containing the Tonda di Cagliari cultivar present a higher concentration of squalene, sitosterol, saturated fatty acids and linoleic acid, according to what we found in the first harvesting year. On the other hand, the olive oils from the Sassari province, containing mostly the Bosana cultivar, are characterized by a slightly higher content of terpene 4.

5.3.3 Comparison between olive oils of the two harvesting years

The olive oils from Sardinia show a very high value of linoleic acid, more than those from Sicily. As in Sicily, the concentration of aldehydic compounds is extremely variable, but contrary to that region, in Sardinia their concentration decrease from the first to the second harvesting year.
For this reason the aldehydic compounds cannot be considered as characterizing variables for the oils from Sardinia. On the contrary the characterizing variables are squalene, terpenes 2, 3 and 4, having a very high concentration in these olive oils, while terpene 1 present a very low concentration.
In the PCA shown in figure **5.25** we can clearly see the seasonal effect on the olive oils from Sardinia, considering each province individually: in fact overall the olive oils of the second year are shifted to lower values of the PC1 (42.2% of the total variance), still remaining apart the groups with different geographic origin. The most affected variables are the aldehydic

compounds, the linoleic acid and the wax, whose concentration decreases in the second year. Also the terpenic compounds are affected by the seasonal effect: in the oils from the Cagliari province we have a strong decrease of the terpenes 3 and 4 passing from the first to the second year, while in the oils from the Sassari province only the terpene 4 decreases in the second year.

Notably, samples of the second harvesting year are grouped by the geographic origin better than those of the first year, more dispersed in the graph.

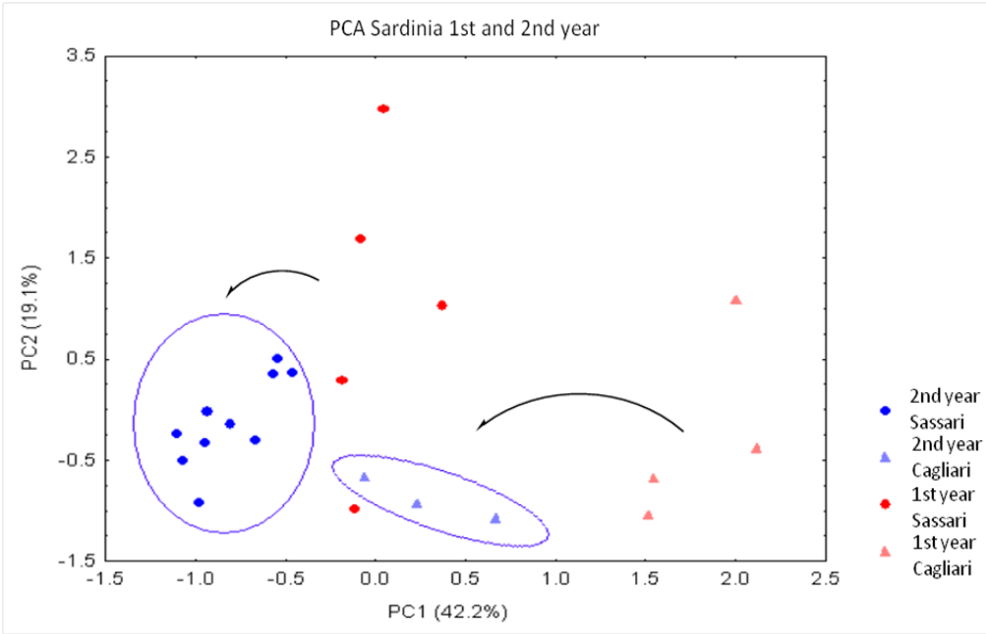
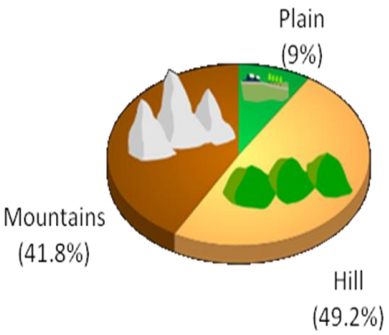


Figure 5.25: PCA of the olive oils from Sardinia of the two harvesting years. The geographic origin of the samples is also indicated.

5.4 Characterization of olive oils from Calabria



In this region the olive tree is present since the 8th-7th cent. b.C., when the ancient Greeks brought this resource in this region, but it's only with the ancient Romans, their technical innovations and the improvement of the procedures that this culture spread.
The most suitable areas for olive trees culture are those near Cosenza, Lamezia Terme and Reggio Calabria (Piana di Gioia Tauro and Locride).
The most common native cultivars near Cosenza are Carolea, Coratina, Tondina and Ottobratica. In the Piana di Gioia Tauro we found mostly the Ottobratica, Sinopolese, Roggianella and Leccino cultivars. In Locride area cultivars like Geracese (or Grossa di Gerace) and Nocellara are present.
Three PODs are present in Calabria region: Bruzio, Lametia, Alto Crotonese.
The first one is also associated to the following geographic references: "Fascia Prepollinica" (Tondina, min. 50%, Carolea, max. 30%, Grossa di Cassano, max. 20%; colour: green with yellow highlights; aroma: medium fruity; flavour: fruity), "Valle di Crati" (Carolea, min. 50%, Tondina, max. 30%, Rossanese or Dolce di Rossano, max. 20%; colour: from green to yellow; aroma: medium fruity; flavour: fruity), "Colline Joniche Presilane" (Rossanese or Dolce di Rossano, min. 70%; colour: gold with green highlights; aroma: lightly fruity; flavour: fruity with taste of almond), "Sibaritide" (Grossa di Cassano, min. 70%, Tondina,max. 30%; colour: yellow

with green highlights; aroma: lightly fruity; flavour: lightly fruity a little bit bitter).
The Lametia POD olive oil is obtained from the Carolea cultivar (90% min.) and its colour goes from green to straw-yellow, its aroma is fruity and it has a gently fruity flavour.[62]

Table 5.4: oils from Calabria with the respective information

HARVESTING YEAR	CULTIVATION SITE	ALTITUDE (m above sea level)	CULTIVAR
09/10	Rossano (CS)	250	Rossanese
09/10	Rossano (CS)	350	Carolea (Nicastrese)
09/10	Rossano (CS)	200	Rossanese
09/10	Acri (CS)	650	Carolea (Nicastrese)
09/10	Bisinano (CS)	450	Carolea (Nicastrese)
09/10	Bisinano (CS)	500	Frantoio
09/10	Acri (CS)	550	Leccino
09/10	Roccaforte del Greco (RC)	800-900	Sinopolese , Carolea
09/10	Rossano (CS)	60	Coratina
09/10	Rossano (CS)	10	Coratina
09/10	Corigliano C. (CS)	50	Carolea
09/10	Rossano (CS)	40	Coratina
09/10	Cerchiara di Calabria (CS)	90	Tondina
09/10	Cerchiara di Calabria (CS)	100	Grossa di Cassano
09/10	Cerchiara di Calabria (CS)	120	Grossa di Cassano
09/10	Roggiano Gravina (CS)	250	Roggianella
09/10	Delianuova (RC)	450	Sinopolese
09/10	Delianuova (RC)	450	Sinopolese
09/10	Caulonia (RC)	500	Geracese
09/10	Stignano (RC)	20	Carolea
09/10	Cittanova (RC)	300	Ottobratica
09/10	Cittanova (RC)	300	Carolea
09/10	Cittanova (RC)	400	Ottobratica
09/10	Oppido Mamertina (RC)	200	Ottobratica, Sinopolese

09/10	Taurianova (RC)	220	Roggianella
09/10	Caulonia (RC)	100	Geracese
09/10	Placanica (RC)	100	Nocellara Messinese
09/10	Corigliano C. (CS)	60	Rossanese
10/11	Cerchiara di Calabria (CS)	150	Tondina, Cassinese
10/11	Cerchiara di Calabria (CS)	150	Cassinese
10/11	Cerchiara di Calabria (CS)	90	Tondina
10/11	Santa Sofia d'Epiro (CS)	500	Carolea
10/11	Rossano (CS)	30-60	D.di Rossano, Nocell. del Bel, Frant.
10/11	Rossano (CS)	70	Dolce di Rossano
10/11	Rossano (CS)	70	Dolce di Rossano
10/11	Rossano (CS)	80	Dolce di Rossano
10/11	Acri (CS)	650	Carolea
10/11	Bisignano (CS)	450	Carolea
10/11	Bisignano (CS)	500	Carolea
10/11	Acri (CS)	550	Leccino
10/11	Rossano (CS)	250	Rossanese
10/11	Rossano (CS)	200	Rossanese
10/11	Rossano (CS)	350	Carolea
10/11	Rossano (CS)	100	Rossanese
10/11	Rossano (CS)	100	Rossanese
10/11	Delianuova (RC)	450	Sinopolese
10/11	Cittanova (RC)	300	Ottobratica
10/11	Cittanova (RC)	300	Carolea
10/11	Oppido Mamertina (RC)	200	Ottobratica, Sinopolese
10/11	Caulonia (RC)	100	Geracese
10/11	Stignano (RC)	20	Carolea
10/11	Caulonia (RC)	500	Geracese
10/11	Roccaforte del Greco (RC)	800-900	Sinopolese, Carolea

5.4.1 Statistical analysis of the olive oils of the first year

During the first harvesting year 28 olive oils have been collected: 12 from the province of Reggio Calabria and 16 from that of Cosenza.

The PCA of figure **5.26** shows that olive oils are not grouped according to their geographical origin. Anyway, samples from Reggio Calabria province have a high level of aldehydic compounds, saturated fatty acids and sitosterol, while those from Cosenza have a high concentration of unsaturated fatty acids.
The two olive oils from Reggio Calabria separated from the other elements of the group come from Caulonia, and present the highest level of linoleic acid for this region.

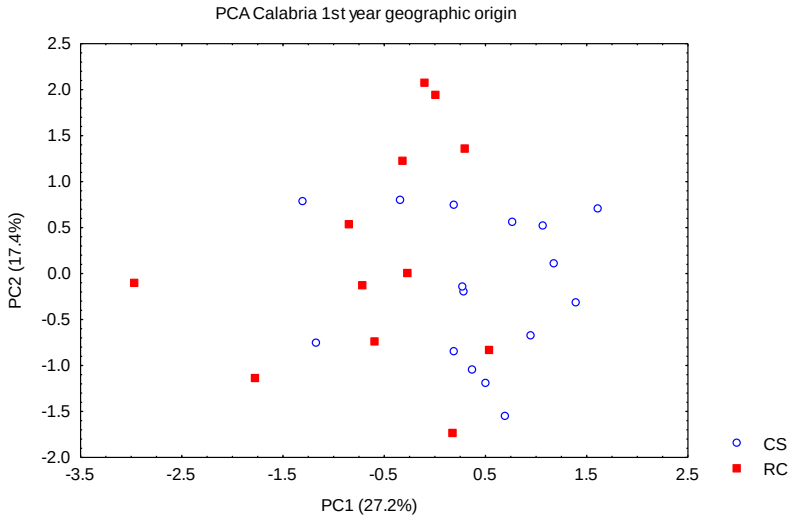


Figure 5.26: PCA of the oils from Calabria of the first harvesting year labelled by geographic origin.

The PCA reported in figure **5.27** shows the effect of the genetic factor. The olive oils are all mono or bivarietal.
Two samples from Geracese cultivar, apart from the others on the left side of the graph, are characterized by a higher content of terpene 1 and linoleic acid and low concentration of wax.
We can also distinguish a group of four Carolea oils at low values of the PC2 (17.4% of the total variance), with a very high content of squalene and terpene 4, that is lower in the two remaining samples, placed in the superior region of the graph. Notably, the Carolea samples come from both the provinces, Cosenza and Reggio Calabria.
The olive oil with mixed cultivars Sinopolese and Carolea is near the 4 olive oils from Carolea mentioned above: all samples present high value of terpene 4. The two monovarietal samples Sinopolese, very close to each

other, show a high concentration of aldehydic compounds and low content of squalene. The two olive oils of Ottobratica cultivar, are placed at the centre of the graph at the same value of the PC1, both having a high concentration of sitosterol and low concentration of terpene 4 and squalene, and they are separated only along the PC2 because of the different content of hexanal, higher in one of them, and linoleic acid, higher in the other one.
The three olive oils of Coratina, placed in the right bottom of the PCA graph, are characterized by low levels of hexanal and sitosterol; the latter is instead more abundant in the two olive oils with Grossa di Cassano cultivar, at the centre of the graph.

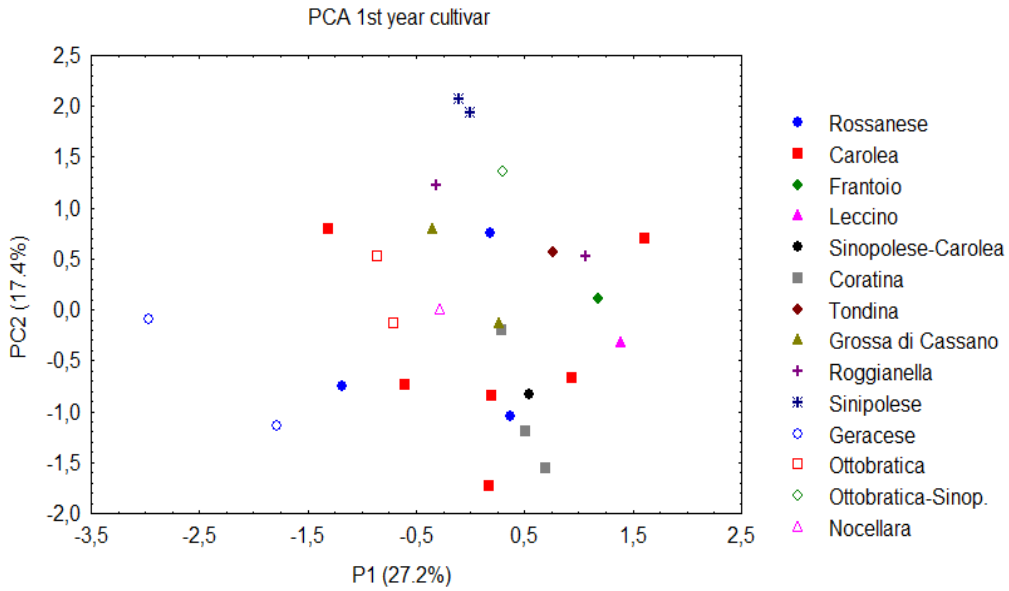


Figure 5.27: PCA of the oils from Calabria of the first harvesting year, labelled by the cultivar.

In the PCA plot of figure **5.28** we can see that the altitude is not a discriminant parameter in this case, not allowing the separation of the two groups of olive oils, even if the samples from hill altitude present a higher content of hexanal and of terpene 2 and lower concentration of unsaturated fatty acids and linolenic acid.

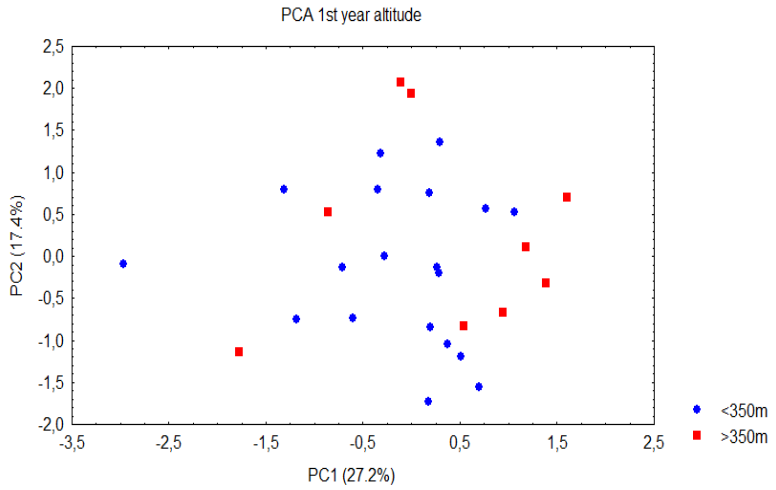


Figure 5.28: PCA of the olive oils from Calabria of the first harvesting year labelled by the altitude of production.

5.4.2 Statistical analysis of the olive oils of the second year

Twenty-five samples have been collected during the second harvesting year: 17 from the Reggio Calabria province and 8 from the Cosenza province.

The PCA of figure **5.29** shows the effect of the geographic origin: only three samples from the Reggio Calabria province are not overlapped with those from the other province: they are the two sample from Caulonia and the one coming from Stignano. These olive oils present a high content of saturated fatty acids, terpene 4, squalene and the maximum content for this region of linoleic acid. The olive oils from Cittanova, Delianova, Oppido Mamertino and Roccaforte fall among the samples from the Cosenza province.

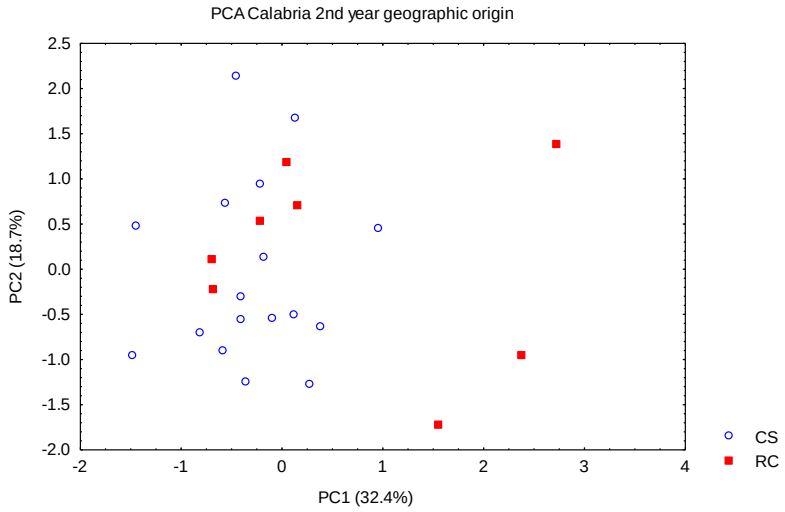


Figure 5.29: PCA of the oils from Calabria of the second harvesting year labelled by geographic origin.

The effect of the genetic factor is shown in figure **5.30**, in a PCA including only monovarietal olive oils: 7 samples of Rossanese cultivar, 7 samples Carolea, 2 samples of Geracese and only one olive oil of Tondina, Cassinese, Leccino, Sinopolese and Ottobratica cultivar.
Notably the two samples of Geracese cultivar are clearly separated from all the others (as seen in the first harvesting year) and so the Carolea olive oils, all but one, and those of Rossanese cultivar. The only sample of Carolea cultivar apart from its group is that coming from Stignano.
Olive oils of Carolea cultivar are characterized by a higher concentration of terpene 3 and wax and a lower content of terpene 1, linolenic and linoleic acid and sitosterol. The sample from Stignano, apart from the rest of this group, shows not only a high concentration of terpene 3 and wax, low levels of terpene 1 and linolenic acid, as the other Carolea samples, but also a high level of sitosterol, linoleic acid, squalene and hexanal. The two olive oils of Geracese cultivar have a high value of trans 2-hexenal, terpenes 1 and 4, saturated fatty acids, squalene and mostly of linoleic acid. On the other hand the Rossanese cultivar gives olive oils poor of terpenes 2 and 3 and wax and rich of hexanal.

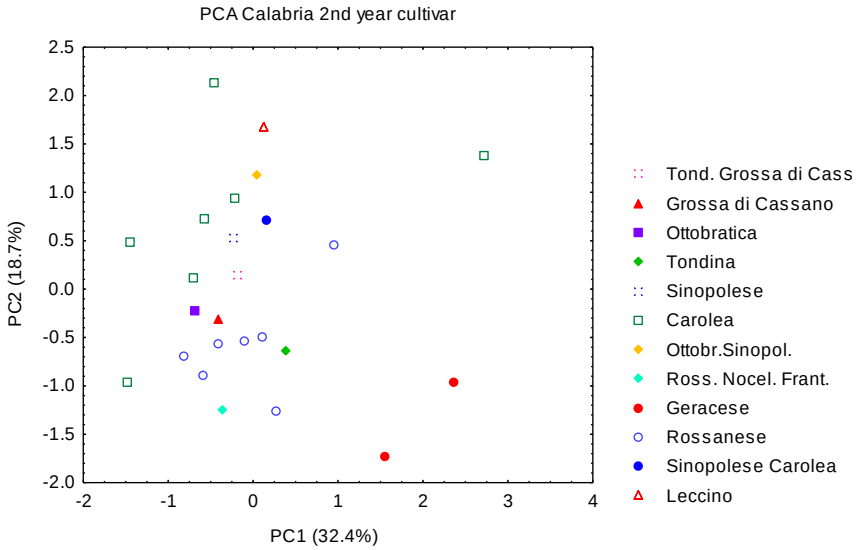


Figure 5.30: PCA of the olive oils from Calabria of the second harvesting year, labelled by the cultivar.

The altitude effect, shown in figure **5.31**, evidences how the samples produced in hill farms are different from those produced in plain, with the exception of only two samples: one is the only olive oil of Rossanese cultivar coming from hill altitude, the other one is one of two samples coming from Caulonia. Generally for this region olive oils from hill farms are richer of wax and terpene 3, and olive oils from plain farms contain more linoleic and linolenic acid which is high also in the olive oils of the first year.

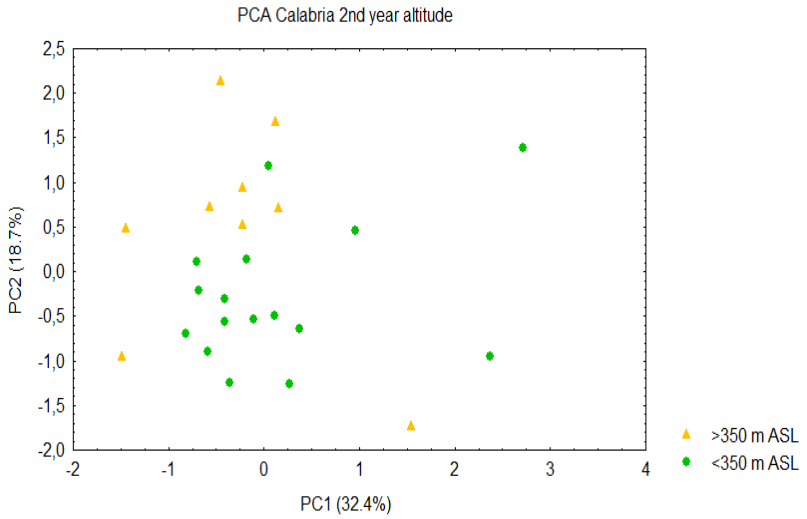


Figure 5.31: PCA of the oils from Calabria of the second harvesting year labelled by altitude.

5.4.3 Comparison between olive oils of the two harvesting years

As shown in the PCA of figure **5.32**, no seasonal effect is visible for the olive oils of this region: none of the variables is subjected to an appreciable variation. This is the only Italian region where the seasonal effect is not relevant, even considering each province or area or any other factor.

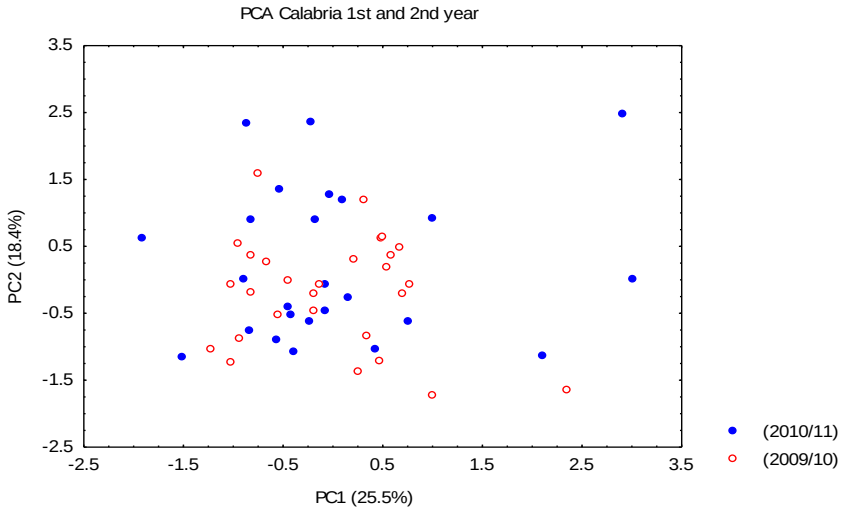


Figure 5.32: PCA of olive oils from Calabria of both the harvesting years.

5.5 Characterization of olive oils from Apulia

Apulia is one of the most productive regions of Italy and presents 4 PODs:

- Collina di Brindisi: the Northern part of the province of Brindisi produces an excellent olive oil, born from a thousand-years-old experience (the cultivation of the olive tree was introduced here by the ancient Greeks and Romans);
- Dauno: the Foggia province was called Daunia by the ancient Romans (from the name of the Illyrian population here settled), the first ones to import the olive cultivation in this land;
- Terra di Bari: the olive tree cultivation dates back to the Neolithic (5000 b.C.) and it is the most typical tree in this area. In the Middle Age this oil was very esteemed by the merchant of Venice, that exported this precious product in the whole Europe and it still represents a historical, cultural and economical heritage of this region.
- Terra d'Otranto: this was the medieval name of Salento, and the oil produced in this area has a very ancient tradition: 8000 years ago the first habitants of this land already cultivated the olive tree, followed by Massapi, Phoenicians, Greeks and Romans. The area involved in this cultivation includes the whole province of Lecce, part of that of Taranto and Brindisi.[62]

Table 5.5: oils from Apulia with the respective information

HARVESTING YEAR	CULTIVATION SITE	ALTITUDE (m above sea level)	CULTIVAR
09/10	Vico del Gargano (FG)		
09/10	Carpino (FG)		
09/10	Carpino (FG)		
09/10	Poggio Imperiale (FG)	70	Peranzana, Leccino, Frantoio, S. Caterina
09/10	Apricena (FG)	70	Peranzana, Leccino, S. Caterina
09/10	Apricena (FG)	75	Peranzana, Frantoio, S. Caterina
09/10	Apricena (FG)	75	Peranzana, Leccino, Frantoio, S. Caterina
09/10	Ascoli Satriano (FG)	330	Moraiolo
09/10	Ascoli Satriano (FG)	370	Coratina
09/10	Manfredonia (FG)	65	Ogliarola Garganica
09/10	Manfredonia (FG)	50	Ogliarola Garganica
09/10	S. Giovanni Rotondo (FG)	105	Ogliarola Garganica
09/10	Cerignola (FG)	65	Coratina
09/10	Cerignola (FG)	77	Coratina
09/10	Monte S.Angelo (FG)	10	Ogliarola Garganica
09/10	Ascoli Satriano (FG)	311	Coratina
09/10	Castelluccio dei Sauri (FG)	270	Coratina
09/10	Statte (TA)	100	Frantoio
09/10	Palagianello (TA)	80	Nociara
09/10	Statte (TA)	115	Leccino, Ogliarola, Coratina
09/10	Crispiano (TA)	350	Coratina
09/10	Martina Franca (TA)	500	Leccino, Frantoiana
09/10	Statte (TA)	120	Frantoio
09/10	Castellaneta (TA)	120	Nociara, Leccino, Frantoio
09/10	Grottaglie (TA)	370	Picholine, Leccino, Nociara
09/10	Crispiano (TA)	200	Olearola, Cellina
09/10	Manduria (TA)	70	Cellina, Ogliarola, Leccino
09/10	Manduria (TA)	30	Cellina di Nando

09/10	Statte (TA)	100	Frantoio, Leccino
09/10	Palagianello (TA)	80	Leccino, Frantoio
09/10	Palagianello (TA)	80	Coratina
09/10	Palagianello (TA)	80	Leccino
10/11	Monte S. Angelo (FG)	20	Ogliarola Garganica
10/11	Cerignola (FG)	166	Coratina
10/11	Lucera (FG)	80	Frantoio, Leccino
10/11	Biccari (FG)	380	Coratina, Leccino, Pizzuta
10/11	Carpino (FG)	145	Frantoio, Ogliarola garganica
10/11	Cerignola (FG)	160	Coratina
10/11	Agro di Poggio Imperiale (FG)	70	Peranzana, S. Caterina, Leccino
10/11	Ascoli Satriano (FG)	400	Coratina
10/11	Apricerna - (FG)	60	Peranzana, Leccino, mixed
10/11	Lucera (FG)	180	Coratina
10/11	Monte S. Angelo (FG)	30	Ogliarola Garganica
10/11	Ascoli Satriano (FG)	400	Coratina
10/11	Cerignola (FG)	160	Coratina
10/11	Carpino (FG)	145	Frantoio
10/11	Crispiano (TA)	120	Nociara
10/11	Palagianello (TA)	200	Leccino, Frantoio, Coratina, Cima di Melfi
10/11	Statte (TA)	100	Coratina
10/11	Statte (TA)	200	Coratina
10/11	Statte (TA)	100	Coratina, Leccino, Frantoio
10/11	Crispiano (TA)	150	Ogliarola Garganic
10/11	Accetta Piccola (TA)	150	Leccino, Ogliarola, Frantoio
10/11	Palagianello (TA)	60	Leccino, Frantoio, Coratina
10/11	Manduria (TA)		Cellina di Nardò, Leccino
10/11	Crispiano (TA)	220	Ogliarola, Cima di Melfi
10/11	Statte (TA)	96	Coratina
10/11	Statte (TA)		Ogliarola Garganica
10/11	Taranto	85	Picholine (Maroccan origin), Leccino
10/11	Palmariggi (LE)	30	Ogliarola
10/11	Giug, Giur, Magl, Otra, Palm (LE)	30-40	Cellina di Nardò
10/11	Galatone - Martignano		Leccino

10/11	(LE) Caprarica (LE)	0	Cellina, Ogliarola, Leccino

5.5.1 Statistical analysis of the olive oils of the first year

The olive oils collected in this region for the first year are 32: 17 from the province of Foggia and 15 from the province of Taranto.
Performing the PCA on this samples labelled by the geographic origin no differentiation has been pointed out.
On the other hand, the analysis of the genetic factor, performed only on the monovarietal olive oils, allowed us to obtain the results showed in figure **5.33**: the two largest groups, containing the olive oils of Coratina and Ogliarola cultivars respectively, are well separated along the PC1 (53.5% of the total variance). The former group is characterized by a lower concentration of terpene 3, terpene 1, sitosterol and saturated fatty acids; the latter presents higher levels of hexanal, trans-2-hexenal, linoleic acid and wax and a lower content of squalene.
The four Ogliarola olive oils, marked with red symbols in the graph, come from the Foggia province and, in particular, they present the same characteristics of the cultivation soil.

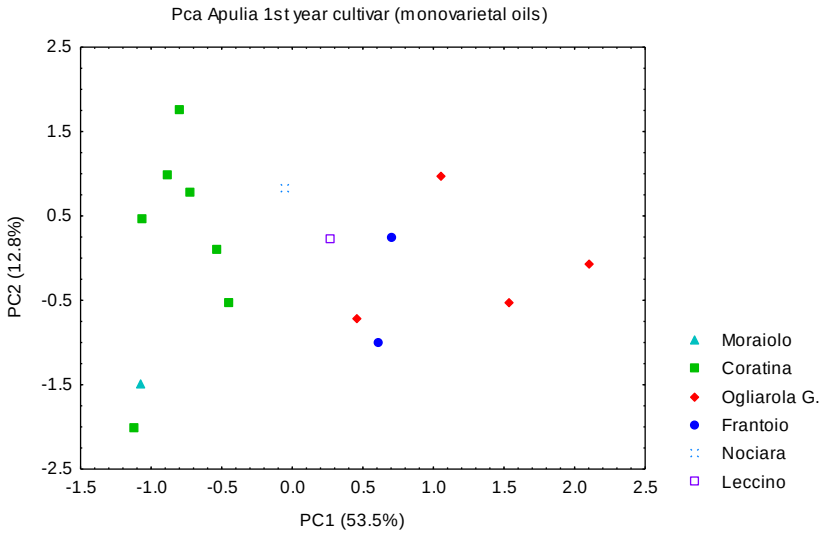


Figure 5.33: PCA of the olive oils from Apulia of the first harvesting year, labelled by the cultivar.

The nature of the soil again can explain the position of the olive oil of Moraiolo cultivar: it shares with 5 of the 7 olive oils of the Coratina cultivar the clay nature of the soil of provenance.
Performing the PCA on the 27 samples of which the soil information was available (table **5.6**) we obtained the graph reported in figure **5.34**.

Table 5.6: olive oils from Apulia (2009/10) with the respective additional information

HARVESTING YEAR	CULTIVATION SITE	SOIL NATURE
09/10	Poggio Imperiale (FG)	Middle texture
09/10	Apricena (FG)	Middle texture
09/10	Apricena (FG)	Middle texture
09/10	Ascoli Satriano (FG)	Clay soil
09/10	Ascoli Satriano (FG)	Clay soil
09/10	Manfredonia (FG)	Silty soil
09/10	Manfredonia (FG)	Silty soil
09/10	S. Giovanni Rotondo (FG)	Silty soil
09/10	Cerignola (FG)	Clay soil
09/10	Cerignola (FG)	Clay soil
09/10	Monte S.Angelo (FG)	Silty soil
09/10	Ascoli Satriano (FG)	Clay soil
09/10	Castelluccio dei Sauri (FG)	Clay soil
09/10	Statte (TA)	Calcareous soil
09/10	Palagianello (TA)	Middle texture
09/10	Statte (TA)	Middle texture
09/10	Crispiano (TA)	Middle texture
09/10	Martina Franca (TA)	Middle texture
09/10	Statte (TA)	Middle texture
09/10	Castellaneta (TA)	Middle texture
09/10	Grottaglie (TA)	Middle texture
09/10	Crispiano (TA)	Middle texture
09/10	Manduria (TA)	Middle texture
09/10	Statte (TA)	Calcareous soil
09/10	Palagianello (TA)	Middle texture
09/10	Palagianello (TA)	Middle texture
09/10	Palagianello (TA)	Middle texture

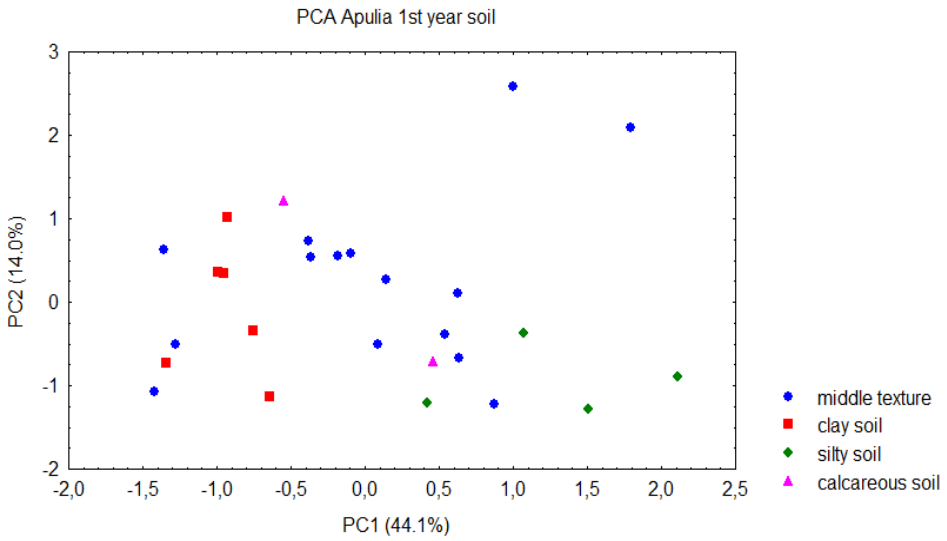


Figure 5.34: PCA of the oils from Apulia of the first harvesting year, labelled by the nature of the soil.

Clearly the nature of the soil gives an important contribute to the determination of the olive oil composition, additive to that of the cultivar.
The analysis of the variables pointed out the compounds most affected by this factor: trans-2-hexenal, terpenes 1 and 3, linoleic acid, *sn* 1,2-diglycerides, saturated fatty acids, wax and sitosterol, generally are present in low concentration in the clay soil oils and in high concentration in olive oils from silty soil. Olive oils from middle-texture soil generally have a more heterogeneous composition. These variables have been selected by the ANOVA having a lower p-level (≤0.05) and, consequently, a higher importance in the differentiation of the samples. (Table **5.7**).
We repeated the PCA, using only these variables, on the 28 samples whose soil information was available and the graph we obtained is shown in figure **5.35**.

Table 5.7: ANOVA results for the analysis concerning the nature of the soil

N.	Variable	*Fisher F*-value	p-level
1	Hexanal	1.177	0.340
2	**T-2hexenal**	3.970	**0.020**
3	Terpene 4	0.445	0.723
4	**Terpene 3**	4.047	**0.019**
5	Terpene 2	2.022	0.139
6	**Terpene 1**	12.030	**6.13E-05**
7	*sn* 1,3-Diglycerides	2.755	0.066
8	***sn* 1,2-Diglycerides**	7.391	**0.001**
9	Linolenic ac. (diall.)	0.223	0.879
10	**Linoleic ac. (dial.)**	6.392	**0.003**
11	Squalene	2.218	0.113
12	Unsaturated. F.A.	2.627	0.074
13	**Saturated. F.A.**	9.935	**2.16 E-04**
14	**Wax**	4.573	**0.0118**
15	Linolenic ac. (CH3)	0.132	0.940
16	**Linoleic ac. (CH3)**	6.901	**0.002**
17	**Sitosterol**	4.335	**0.015**

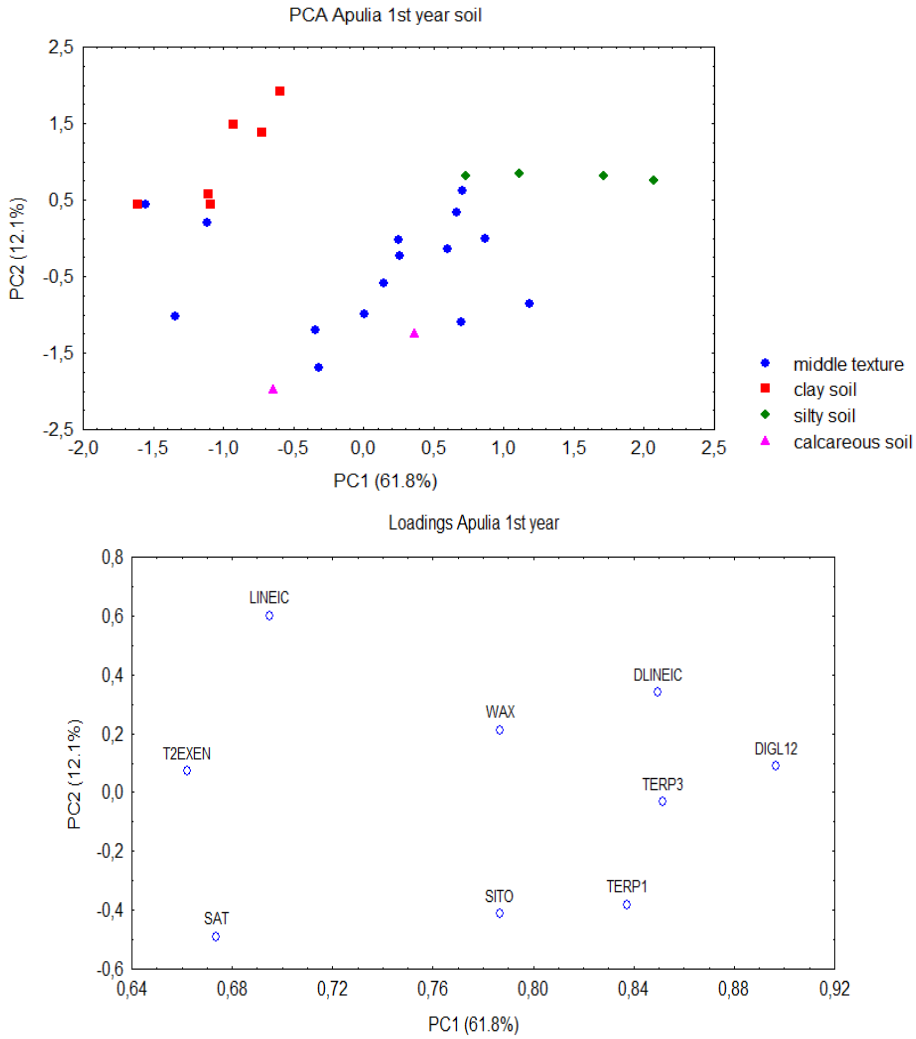


Figure 5.35: a) PCA of the oils from Apulia of the first harvesting year, labelled by the nature of the soil, after the ANOVA. b) Plot of loadings of the PCA after the ANOVA.

The PC1 component, explaining the 61.8% of the total variance, distinguishes quite clearly the groups of olive oils from a silty soil, placed on the right of the graph, the group of olive oils from a clay soil, placed on the left side, and the group of olive oils from middle-texture soil, that are more numerous but also more scattered than the others, having more heterogeneous characteristics.

5.5.2 Statistical analysis of the olive oils of the second year

The second harvesting year includes 31 samples, 14 from the province of Foggia, 13 from the province of Taranto and 4 from the province of Lecce.
The fact that the three province were well represented by the sampling allowed us to perform the LDA, shown in figure **5.36**. The three groups are perfectly separated with their confidential ellipses, suggesting the possibility to create a statistical model of each province.
The olive oils from the province of Taranto are characterized by a high value of squalene and low levels of terpenes 1, 2 and 3; on the other hand, olive oils from the province of Foggia are rich of terpenes 2 and 3 and wax. Finally the olive oils from the Lecce province present high levels of aldehydic compounds and terpene 4, but the lowest concentration of wax in this region.

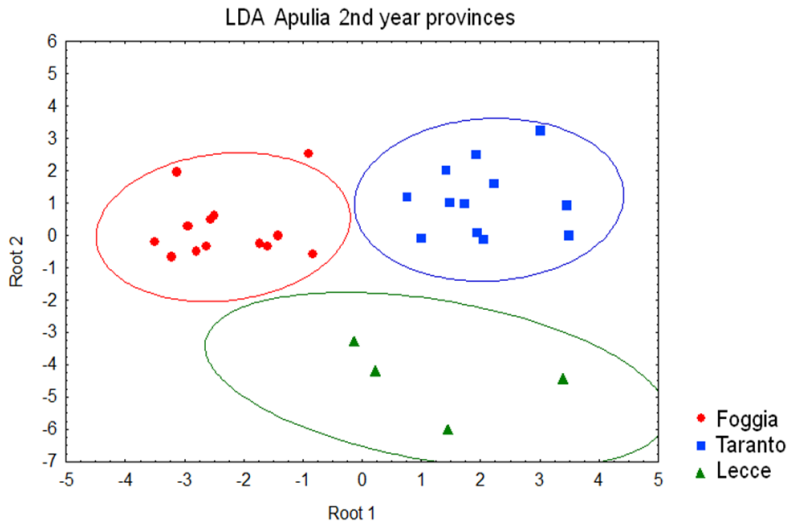


Figure 5.36: LDA of the olive oils from Apulia of the second harvesting year, grouped by province: Foggia, Taranto and Lecce.

Among the 31 samples from Apulia only one was produced at hill altitude, so we couldn't perform the analysis on the basis of this parameter.
On the other hand we can consider the genetic factor. We can see that 18 olive oil are monovarietal, among them 9 samples are obtained from the Coratina cultivar, 5 from the Ogliarola cultivar, and only 1 sample represents each of the following cultivars: Cellina, Leccino, Frantoio and Nociara. All the other samples are multivarietal oils, obtained by combining from 2 to 4

different cultivars, among those previously mentioned and others such as Cima di Melfi, Peranzana, Santa Caterina and Picholine.
So a PCA concerning the monovarietal olive oils, as shown in figure **5.37** was preformed. Samples from Coratina cultivar are placed at lower values of the PC1 (46.2% of the total variance), but they are less clearly separated from the other cultivars in comparison with the first harvesting year.
The olive oils obtained from the Ogliarola cultivar are richer of aldehydic compounds, terpene 1, sitosterol and linoleic acid, as seen in the previous harvesting year, and furthermore they are richer of terpene 2.

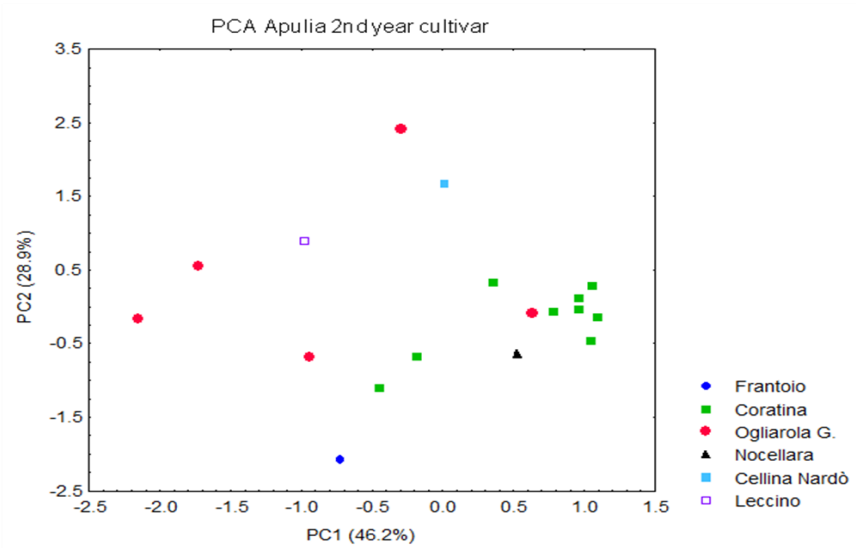


Figure 5.37: PCA of the oils from Apulia of the second harvesting year, labelled by the cultivar.

5.5.3 Comparison between olive oils of the two harvesting years

Figure **5.38** shows the PCA of the oils from Apulia whose sampling has been repeated in both the harvesting years: in this analysis we didn't take in consideration samples from the Lecce province, present only in the second year. The seasonal effect is not so clear: we can only see a partial separation along the PC2 (15.3% of the total variance), due to the higher concentration of squalene and unsaturated fatty acids in the second harvesting year samples, and to their lower concentration of aldehydic compounds, terpenes 4 and 1. (figure **5.39**)

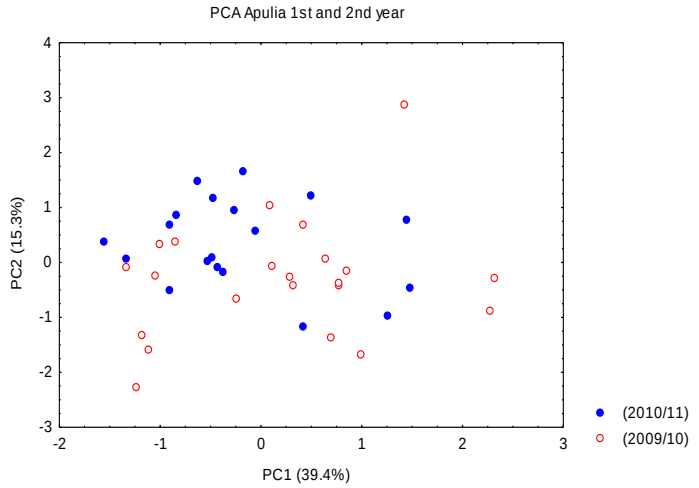


Figure 5.38: PCA of the olive oils from Apulia of both the harvesting years.

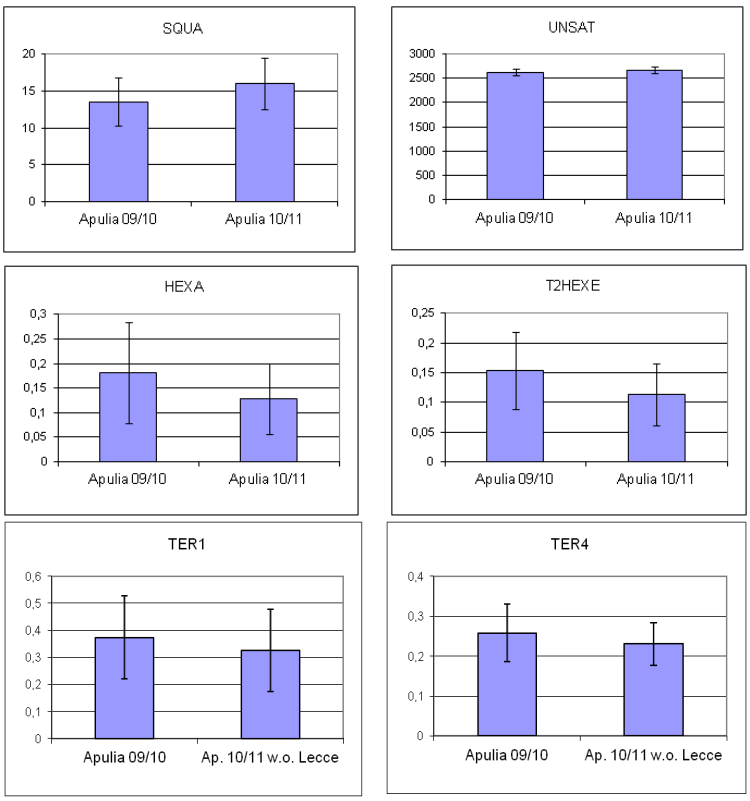
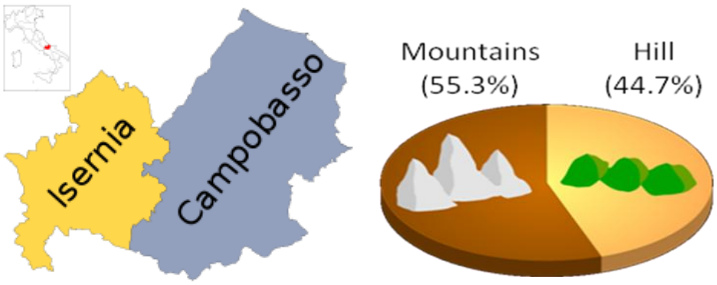


Figure 5.39: mean values of the signal of squalene, unsaturated F.A., hexanal, trans-2-hexenal, terpene 1 and 4, in the oils from Apulia of the two harvesting years.

5.6 Characterization of olive oils from Molise



Olive tree farming is a fundamental element of the culture, history and economy of this region from the Romans age. These trees are present on the whole territory of Molise, despite the climate differences, sometimes very strong . We can see plantations in the Campobasso province and in the Venafro plain, in province of Isernia, but we can also see them in the straight vicinity of the sea or on the mountainside of the western Molise.
From August 2003 the PDO “Molise” is officially recognized in the EU. This olive oil has a yellow-green colour, with a light to medium fruity aroma, a fruity flavour with delicate hints of spice and bitterness.[64]

Table 5.8: olive oils from Molise with the respective information

HARVESTING YEAR	CULTIVATION SITE	ALTITUDE (m above sea level)	CULTIVAR
09/10	S. Martino in P. (CB)	200	Gentile
09/10	S. Martino in P. (CB)	200	Gentile
09/10	Montecilfone (CB)	240	/
09/10	Montecilfone (CB)	240	/
09/10	Campobasso (CB)	650	Leccino
09/10	Campobasso (CB)	650	Leccino
09/10	Larino (CB)	280	Gentile
09/10	Larino (CB)	280	Gentile
09/10	Guglionesi (CB)	200	Leccino
09/10	Guglionesi (CB)	200	Leccino
09/10	Ururi (CB)	250	Gentile-Leccino
09/10	Ururi (CB)	250	Gentile-Leccino
09/10	S. Martino in P. (CB)	250	Gentile-Leccino
09/10	S. Martino in P. (CB)	250	Gentile-Leccino
09/10	Casacalenda (CB)	440	Ogliarola
09/10	Casacalenda (CB)	440	Ogliarola
09/10	Palata (CB)	280	Gentile-Leccino
09/10	Palata (CB)	280	Gentile-Leccino
10/11	Campobasso	750	Leccino
10/11	Piane di Larino (CB)	150	Cipressini
10/11	Camposarcone (CB)	550	Leccino 60%, Gentile 40%
10/11	Castellino del Belfino (CB)	380	Others 40%, Leccino 60%
10/11	Larino Carpineto (CB)	200	Gentile di Larino
10/11	Cocciolete Campomarino (CB)	40	Leccino 60%, Coratina 40%
10/11	Lupara (CB)	550	Others
10/11	Campo di Pietra (CB)	600	Frantoio 60%, Leccino 40%
10/11	S.Giacomo (CB)	150	Gentile, Leccino, Peranzana
10/11	Larino Bagnoli del Trigno (IS)	400-600	Larino, Olivone di Bagnoli del Trigno

5.6.1 Statistical analysis of the olive oils of the first year

All the 18 samples of this harvesting year come from the Campobasso province, produced two by two in the same farm. The couples of samples are clearly visible in the PCA of figure **5.40.** The concentration of *sn* 1,3 diglycerides, sitosterol, aldehydes and terpenes are high in the samples from Guglionesi and low in those from Larino. Olive oils from Montecifone, Ururi and S. Martino in Pensilis show low levels of linolenic acid, squalene and *sn* 1,3 diglycerides. In the samples from Campobasso and Casacalenda we find a content of squalene and terpene 4 higher than those from Palata, richer of saturated fatty acids and terpene 3 and 2.

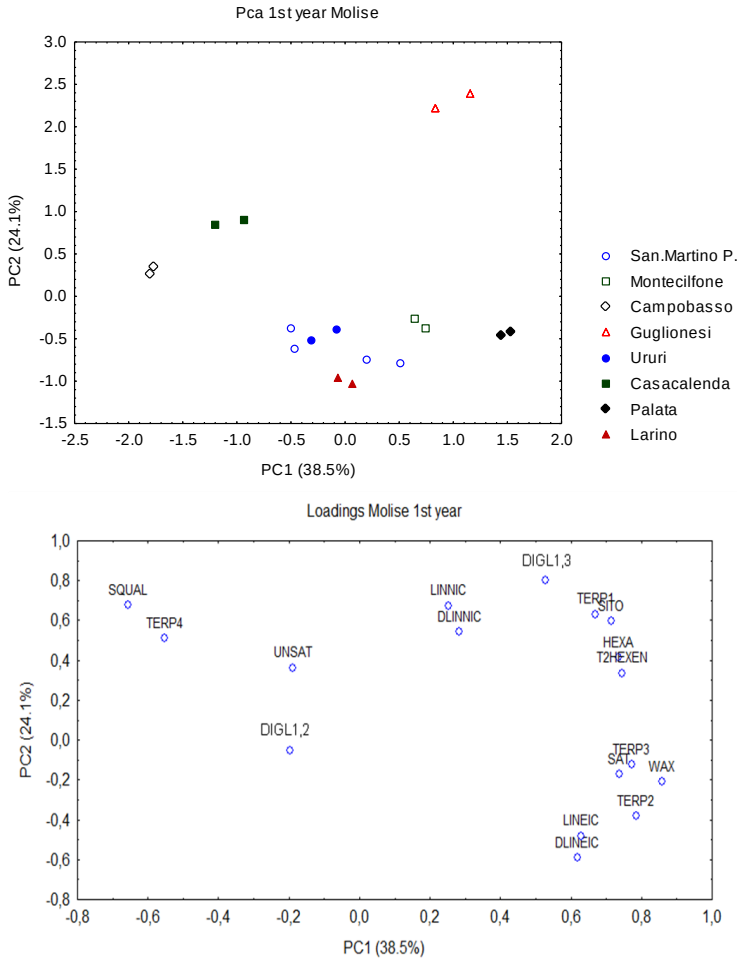


Figure 5.40: a) PCA of the olive oils from Molise of the first harvesting year labelled by geographic origin. b) Plot of loadings of the PCA.

The altimetric effect helps to explain the distribution of the samples in the plot (figure **5.41**): the four samples coming from altitudes higher than 350 m ASL are clearly separated in the left region of the graph along the PC1, explaining the 38.5% of the total variance.
These olive oils are characterized by high concentration of squalene and low concentrations of linoleic acids, saturated fatty acids, wax, terpene 2 and 3.
Among the olive oils produced at lower altitudes a couple of samples is placed at higher values of the PC2, explaining the 24.1% of the total variance: they come from Guglionesi, a place nearer to the coast than the others, and they have a high content of terpene 1, linolenic acid and sitosterol.

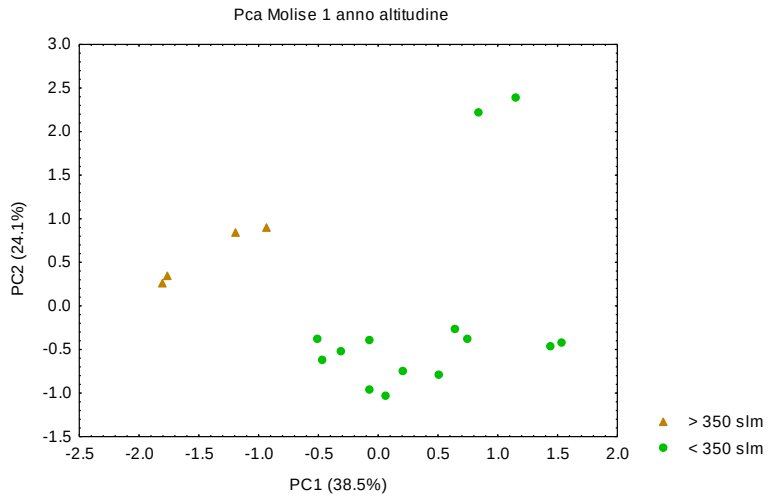


Figure 5.41: PCA of the oils from Molise of the first harvesting year, labelled by the altitude.

A further analysis pointed out the differences concerning the cultivars of the olive oils. This information was available only for 16 samples, that submitted to the PCA gave the graph of figure **5.42**. As we can see that samples from cultivar Gentile and mixed Gentile-Leccino form the group placed at the bottom right of the plot, separated from the others along the PC2 (39.2 % of the total variance) and characterized by a high level of linoleic acid and low concentration of squalene.
The six samples Gentile-Leccino present also higher concentrations of sn 1,2 diglycerides, excluding the samples from Palata, that have the highest level

of aldehydic compounds. These compounds are abundant also in the two oils from Guglionesi having the Leccino cultivar, placed in the highest side of the plot. The two remaining Leccino oils are closer to the two Ogliarola samples: they show a high concentration of squalene and lower level of linoleic acid, sitosterol and terpene 2.
The four samples with Leccino cultivar share higher level of terpene 4 and squalene and low concentration of linoleic acid, even if they are distant in the plot because of the different altitude of production.

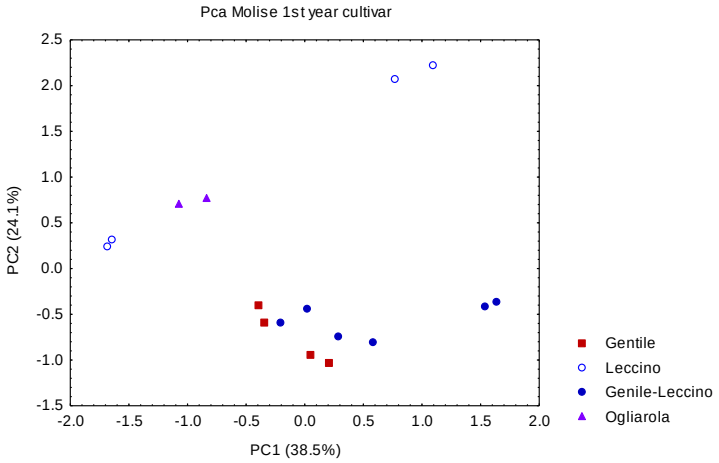


Figure 5.42: PCA of the oils from Molise of the fist harvesting year, labelled by the cultivar.

Table 5.9: oils from Molise (2009/10) with the respective additional information

HARVESTING YEAR	CULTIVATION SITE	SOIL NATURE	RIPENESS DEGREE
09/10	S. Martino in P. (CB)	calcareous	veraison
09/10	S. Martino in P. (CB)	Calcareous	veraison
09/10	Montecilfone (CB)	Clay soil	veraison
09/10	Montecilfone (CB)	Clay soil	veraison
09/10	Campobasso (CB)	/	veraison
09/10	Campobasso (CB)	/	veraison
09/10	Larino (CB)	/	Initial veraison
09/10	Larino (CB)	/	Initial veraison
09/10	Guglionesi (CB)	Clay soil	veraison
09/10	Guglionesi (CB)	Clay soil	veraison
09/10	Ururi (CB)	Middle-texture	Initial veraison
09/10	Ururi (CB)	Middle-texture	Initial veraison
09/10	S. Martino in P. (CB)	Middle-texture	Initial veraison
09/10	S. Martino in P. (CB)	Middle-texture	Initial veraison
09/10	Casacalenda (CB)	Middle-texture	veraison
09/10	Casacalenda (CB)	Middle-texture	veraison
09/10	Palata (CB)	Clay soil	veraison
09/10	Palata (CB)	Clay soil	veraison

Referring to the information reported in table **5.9** we performed some other analysis, concerning the effect on the oil composition of the nature of the soil where the trees are cultivated and the ripeness of the olives.

In figure **5.43** the PCA of all the 18 samples shows the effect of the different nature of the soil on the olive oil composition: analyzing the plot of loadings and the histograms we can see how olive oils from clay soils have a higher content of aldehydes, terpene 1 and sitosterol. The remaining samples, produced on calcareous or middle-texture soil, or whose soil information is not available, are not very differentiated among themselves.

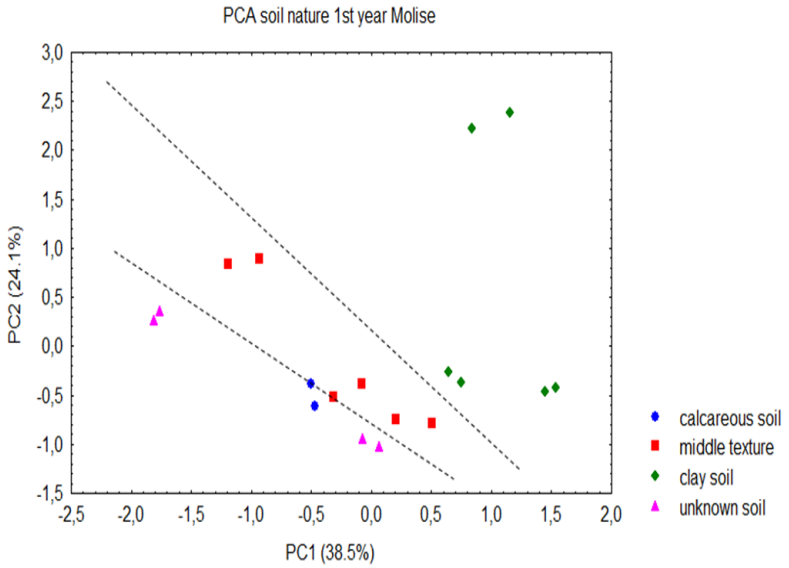


Figure 5.43: PCA of the oils from Molise of the first harvesting year, labelled by the nature of the soil.

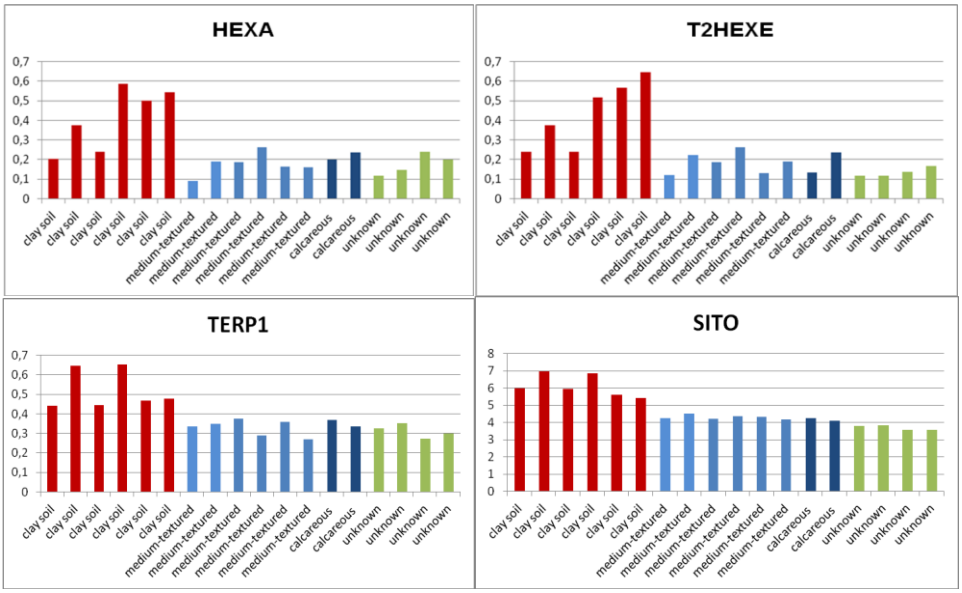


Figure 5.44: Hystograms of some variables of the olive oils from Molise of the first harvesting year. Samples are labelled by the nature of the soil (clay soil, calcareous soil, medium-texture soil).

One last analysis takes in consideration the degree of ripeness of the fruits: the olive oils of this year have been obtained from fruit collected at starting complete veraison. The veraison is a visual ripeness index and is based on the colour of the olive, going from green (start) to purple (complete).
The figure **5.45** shows the PCA of the samples labelled by veraison degree: only six samples have been produced from fruit at starting veraison and they are all grouped at the bottom of the graph.

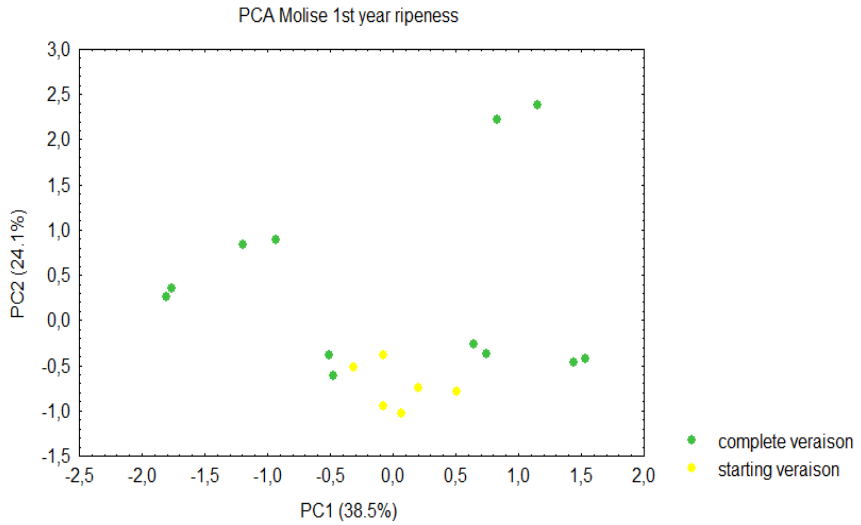


Figure 5.45: PCA of the olive oils from Molise of the first harvesting year, labelled by the veraison degree.

5.6.2 Statistical analysis of the olive oils of the second year

Ten olive oils have been collected all from the province of Campobasso, as the first year.
The first PCA is about the geographic origin and, as we can see in figure **5.46**, the sample from Larino di Carpineto is clearly isolated from the others: it is placed in the rightmost part of the graph and its composition has very high level of aldehydic compounds, terpenes 2 and 3, linoleic acid and wax and lower content of squalene. Notably all the remaining samples are grouped in pairs along the PC2 (17.0% of the total variance), although they have no elements in common, such as cultivar, geographic origin or altitude.

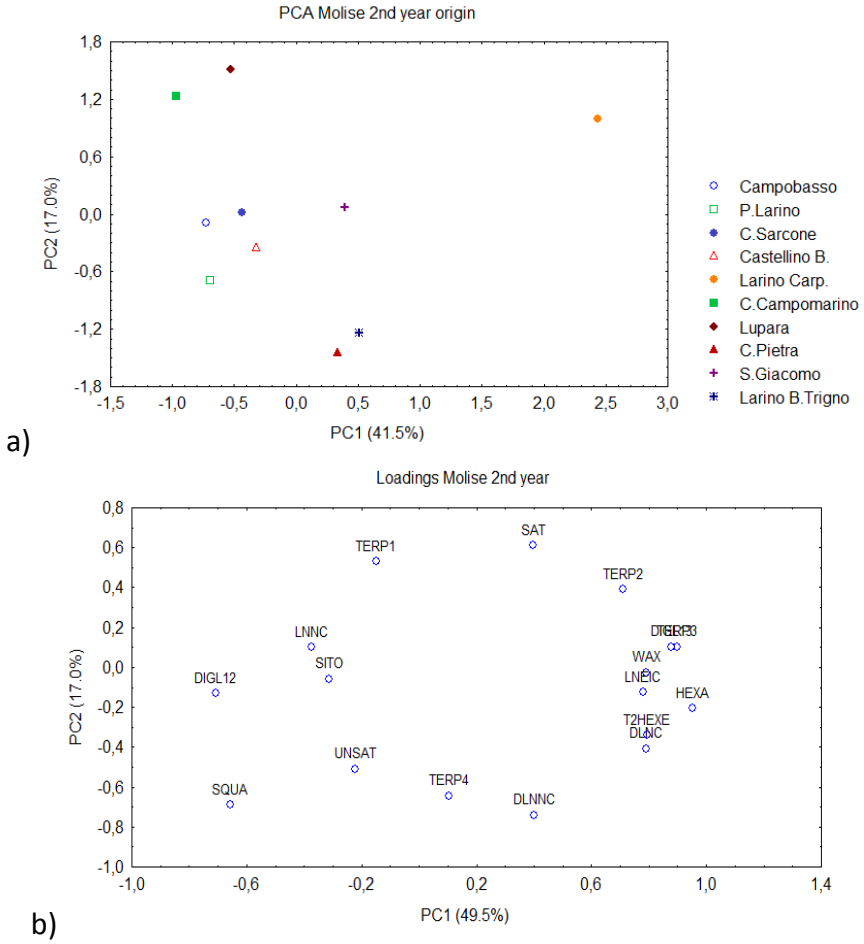


Figure 5.46: a) PCA of the oils from Molise of the second harvesting year labelled by geographic origin. B) Plot of loadings of the PCA.

All the samples are different for cultivar: three oils are monovarietal, respectively Leccino, Cipressini and Gentile di Larino cultivars; the others are samples obtained by mixing different cultivars, in different ratios.
Figure **5.47** shows the PCA of these samples labelled by cultivar: olive oils of Leccino cultivar are placed on the left side of the graph, while those containing the Gentile cultivar stay on the right side. The other criteria, such as geographic origin or altitude (figure **6.48**) don't point out a clear separation of the olive oils of Molise.

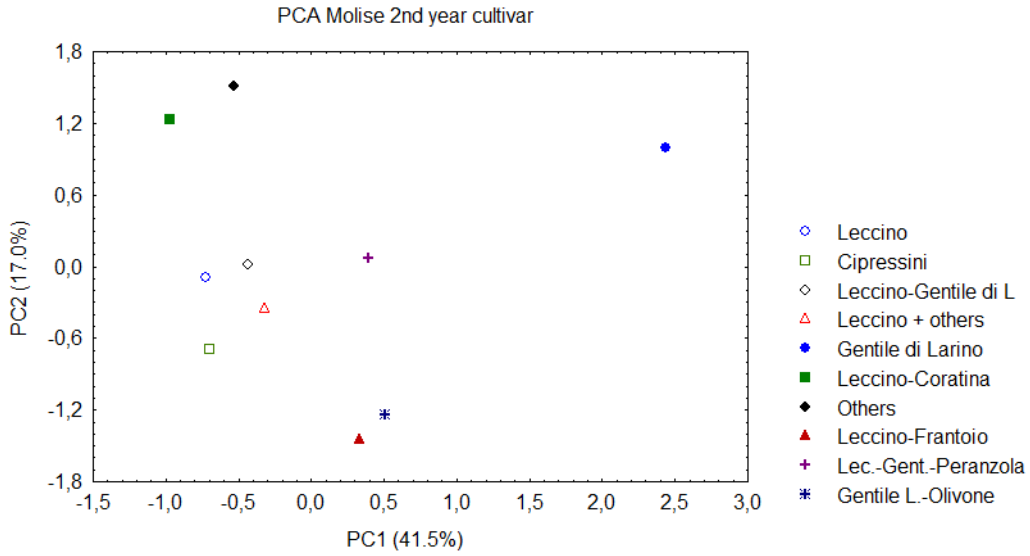


Figure 5.47: PCA of the olive oils from Molise of the second harvesting year, labelled by the cultivar.

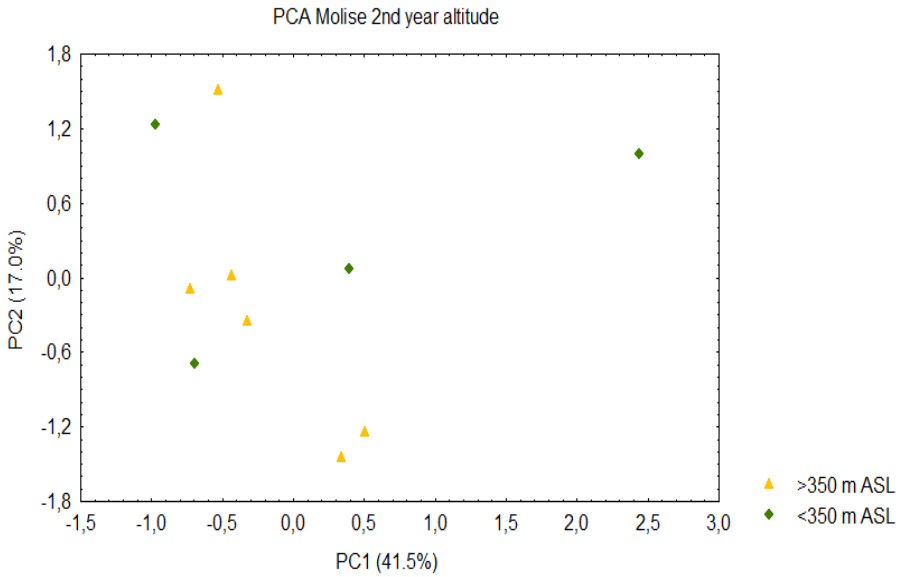


Figure 5.48: PCA of the olive oils from Molise of the second harvesting year, labelled by the altitude.

5.6.3 Comparison between olive oils of the two harvesting years

In the PCA concerning the seasonal effect on the olive oil composition, shown in figure **5.49**, we can see that only one sample from the second harvesting year is apart from the rest of the olive oils of the same season: it is the sample from Larino Bagni del Trigno. For the other samples the separation between the two harvesting years is quite clear along the PC1, explaining the 40.8% of the total variance. Samples from the first year present higher concentration of the aldehydic compounds, saturated fatty acids and terpenes 1, 3 and linoleic acid, while the samples from the second year, beside having very low concentration of these compounds, are characterized by a higher content of squalene and terpene 4.

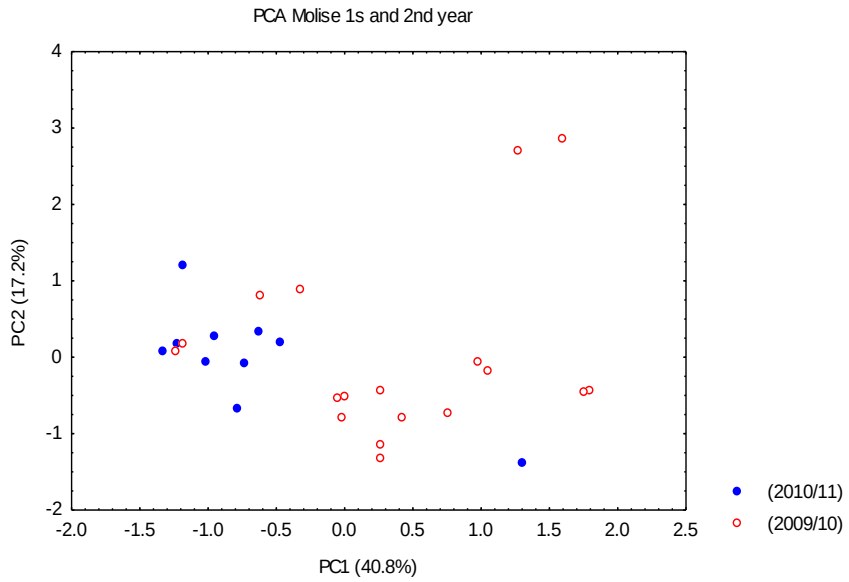
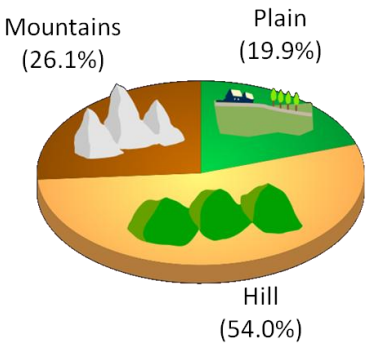


Figure 5.49: PCA of the olive oils from Molise of both the harvesting years.

Comparing the olive oils from Molise with those from the rest of Italy, it is possible to observe in the former group a high concentration of terpene 2 and 3 and wax.

5.7 Characterization of olive oils from Latium



The olive tree farming is very widespread in this region, and the olive oil production is the fifth in the whole nation.
Different cultivars are cultivated in different areas of the region, with particular soil and climate. The European Union recognized two POD for the Latium (Sabina and Canino, vide infra) and are currently in examination other request for POD recognition: Tuscia, Soratte, Colline Pontine, Terre Tiburtine, Terre di Ciociaria e Castelli Romani.
Culstivars such as Frantoio, Leccino, Moraiolo are farmed on the whole region and present in almost all the POD specification, approved or under examination.
The POD Canino includes cultivar such as Cassinese, Leccino, Pendolino, Maurino e Frantoio (pure or mixed with each other). The colour of this oil is emerald green with golden highlights, the aroma is fruity and the flavour decisive, with bitter and spicy aftertaste.
The other POD, Sabina, includes cultivars such as Carboncella, Leccino, Raja, Frantoio, Moraiolo, Olivastrone, Salviana, Olivago and Rosciola (min. 75%). The oil is golden, with green highlights if very fresh, the aroma is fruity and the flavour fruity, velvety, smooth, aromatic, sweet, bitter for very fresh olive oils.[62] Claudio Galeno (129 a.C. – 216 a.C.), the creator of the modern pharmacopeia, defined the oil of Sabina as “the best in the known world”.

Table 5.10: olive oils from Latium, with the respective information

HARVESTING YEAR	CULTIVATION SITE	ALTITUDE (m above sea level)	CULTIVAR
09/10	Rocca Massima (LT)	500	Itrana
09/10	Itri (LT)	300	Itrana
09/10	Itri (LT)	300	Itrana
09/10	Sonnino (LT)	200	Itrana
09/10	Sonnino (LT)	200	Itrana
09/10	Rocca Massima (LT)	500	Itrana
09/10	Palombara (RM)	300	Carboncella Salviana Frantoio
09/10	Nerola (RM)	350	Carboncella Frantoio
09/10	Castelmadama (RM)	450	Rosciola
09/10	Tivoli (RM)	500	Itrana Leccino
09/10	Nerola (RM)	350	Carboncella Frantoio
09/10	S.Vito Romano (RM)	500	Rosciola
09/10	Tivoli (RM)	500	Itrana Leccino
09/10	Ciampino (RM)	300	Frantoio
09/10	Ciampino (RM)	300	Frantoio
09/10	Frascati (RM)	400	Frantoio
09/10	Frascati (RM)	400	Frantoio
09/10	S.Vito Romano (RM)	500	Rosciola
09/10	Palombara (RM)	300	Carboncella Salviana Frantoio
09/10	Arpino (FR)	450	Moraiolo Leccino (various mixed)
09/10	Anagni (FR)	280	Moraiolo and mixed CV
09/10	Arpino (FR)	450	Moraiolo Leccino (various mixed)
09/10	Anagni (FR)	280	Moraiolo and mixed CV
09/10	Poggio Nativo (RI)	400	Frantoio Moraiolo
09/10	Fara Sabina (RI)	200	Frantoio Raja Leccino
09/10	Corese Terra (RI)	200	Frantoio Leccino
09/10	Poggio Nativo (RI)	400	Frantoio Moraiolo
09/10	Corese Terra (RI)	200	Frantoio Leccino
09/10	Fara Sabina (RI)	200	Frantoio Raja Leccino

09/10	Vetralla (VT)	250	Frantoio Leccino
09/10	Viterbo (VT)	450	Canino
09/10	Viterbo (VT)	450	Canino
09/10	Vetralla (VT)	250	Frantoio Leccino
10/11	San Vito Romano (RM)	600	Rosciola 100%
10/11	Tivoli (RM)	300	Itrana 80%, Leccino 20%
10/11	CastelMadama (RM)	400	Rosciola 90%, Pendolino 10%
10/11	Subiaco (RM)	500	Pendoli, Frantoio, Leccino, Rosciol
10/11	Canino (VT)	200	Canino 50%, Frantoio 50%
10/11	Viterbo	300	Canino, Frantoio, Leccino
10/11	Montefiascone (VT)	400	Frantoio 60%, Canino 40%
10/11	Aprilia (LT)	200	Itrana 90%
10/11	Corese Terra (RI)	500	Carboncella 50%, Leccino 50%
10/11	Poggio Mirteto (RI)	550	Leccino 50%, Frantoio 50%

5.7.1 Statistical analysis of the olive oils of the first year

Samples of the first harvesting year are 33: 13 from the province of Rome, 6 from the province of Rieti, 4 from the province of Viterbo and 4 from that of Frosinone.

In the PCA of figure **5.50** the olive oils are labelled by province or place of origin, while in figure **5.51** they are labelled by cultivar.

Olive oils from the same zone generally share also some characteristics concerning the cultivars. In particular when the olive oils from the province of Rome and Rieti are multivarietal, more or less all of them present the Frantoio cultivar in their composition; only 4 olive oils from Rome are monocultivar Frantoio, 3 are monocultivar Rosciola and 2 are bivarietal Itrana-Leccino.

So we can assume that for the olive oils coming from the province of Rome and Rieti the presence of cultivar Frantoio is definitely responsible for the general similarity, in addition to the common site of origin. These olive oils show a high concentration of sitosterol, wax, terpene 2 and 3, that are less abundant in olive oils from the northern provinces of Frosinone and Latina. Analogous results for olive oils from Rome and Rieti concerning wax, terpene 2 and 3 has already been seen in a previous publication.[34]
As we can see, all the olive oils from the province of Latina are monocultivar Itrana and in the PCA (figure **5.50** and **5.51**) they are close to each other and grouped in the left part of the graph, separated from the others along the PC1, that explain the 38.1% of the total variance. They present a very low level of wax.
All the olive oils from the province of Frosinone, still in the southern Latium, even if multivarietal, contain the cultivar Moraiolo as the principal one, and in the PCA graph they are close to the samples from Latina along the PC1, but separated along the PC2, explaining the 14.5% of the total variance. Common elements in the composition of the olive oils from the southern Latium (Frosinone and Latina), differentiating them from the others, are a low content of sitosterol, wax and terpenes, but also a higher concentration of squalene and a lower level of linoleic acid and saturated fatty acids. Those characteristics are also shared by the 2 samples from the province of Rieti containing the cultivar Moraiolo in their multivarietal composition and placed in the graph further down than the other samples from the same province.
The 4 olive oils from the northern province of Viterbo are placed in the central part of the graph, and they have a low concentration of wax, terpene 2 and 3 and high level of terpene 1. Two of them are monocultivar Canino, but they are not so different from the other 2 samples, that are multivarietal Frantoio-Leccino. These two olive oils present some similarities with the olive oils produced in the central and southern Latium containing the Leccino cultivar, and so they are close in the PCA graph.

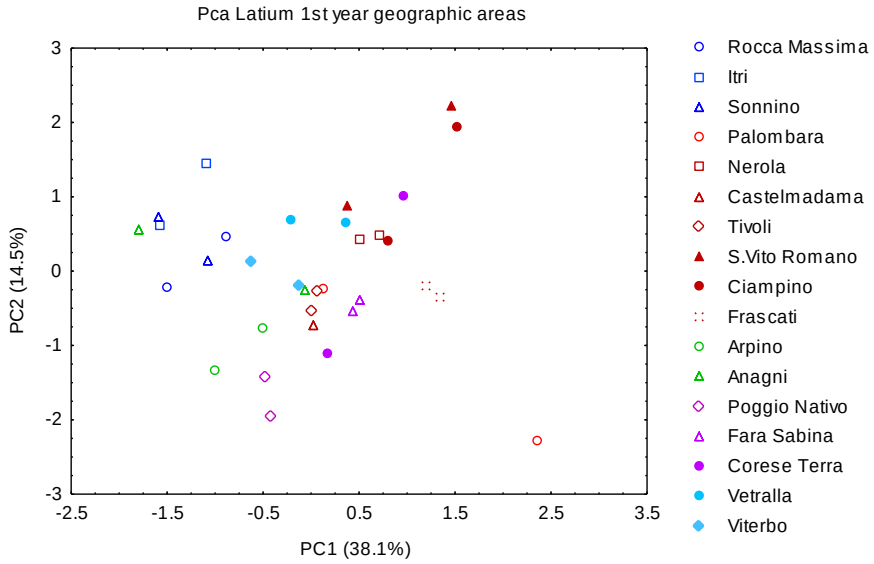


Figure 5.50: PCA of the olive oils from Latium of the first harvesting year, labelled by geographic areas: the symbols of the olive oils from the province of Latina are blue, those for Rome are red, those from Frosinone are green, those for Rieti are purple and those for Viterbo are azure.

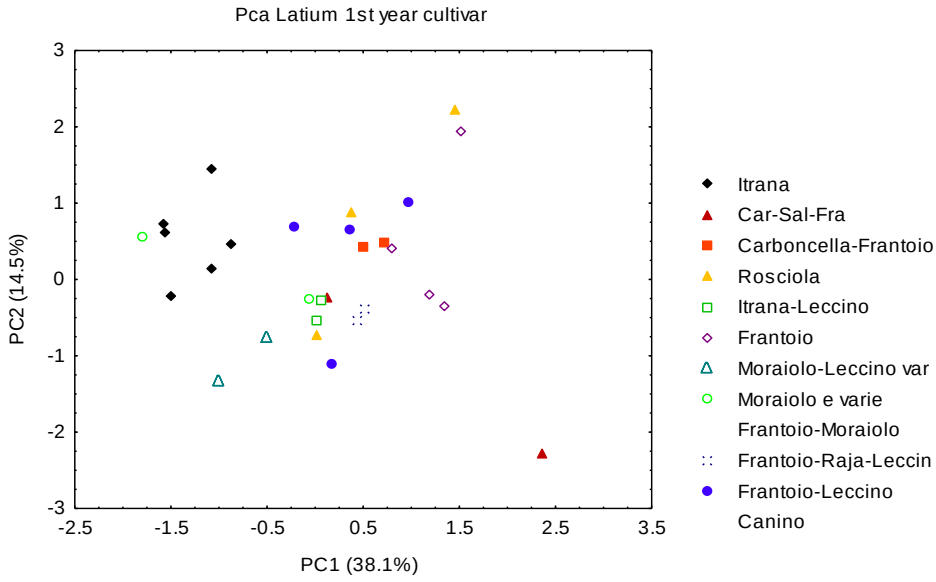


Figure 5.51: PCA of the olive oils from Latium of the first harvesting year, labelled by the cultivar.

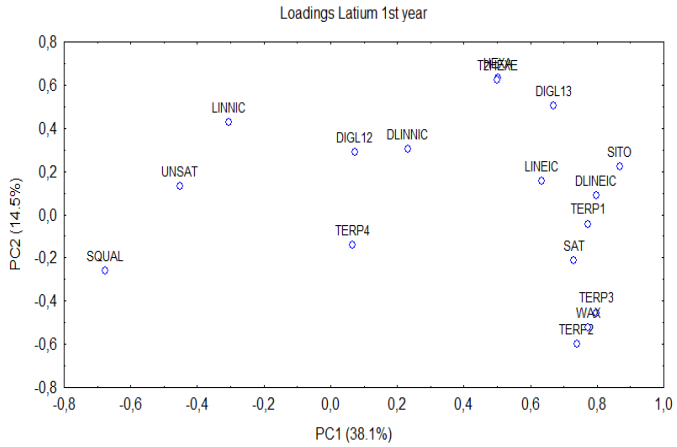


Figure 5.52: Plot of loadings for the PCA of the olive oils from Latium of the first harvesting year.

Performing on the olive oils of Latium the tree clustering analysis (TCA) we obtain a further demonstration of the distinction by areas. This analysis points out the real grade of similarity and consequently of differentiation among the different samples. The ramifications for subsequent levels of the tree represent the gradual increase of the differences between the samples, that, in the end, will be represented with single branches. Groups of samples collected on the same group of branches are similar to each other and different from those of the other branches.
In the TCA of figure **5.53**, the samples seem to be arranged in the order North, Centre, South, even if it is not possible to point out a clear separation of the different groups. As we proceed through the levels of separation, the samples of the north area are placed on the left side of the graph maintaining a greater degree of similarity with some samples of the central provinces; in the same way the samples of the southern provinces, Frosinone and Latina, are placed on the right side maintaining a greater degree of similarity with other samples of the central provinces.
So what we can observe is the distinction of the northern samples from the southern samples, while the samples from the centre of the region are similar to both the other groups.

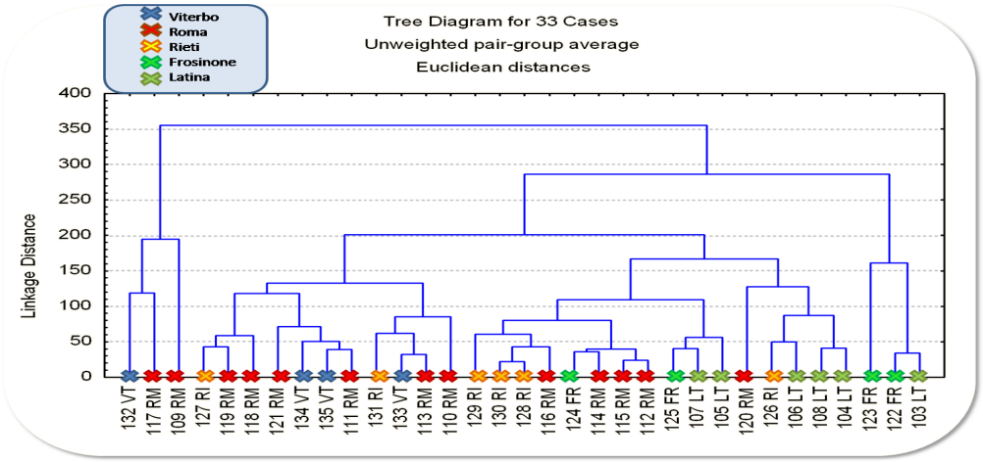


Figure 5.53: TCA of the olive oils from Latium of the first harvesting year, labelled by the province. (VT, RI, RM, FR, LT)

Because of the differences among the olive oils from the three areas, we also performer a LDA to discriminate the three hypothesized groups. In the graph of figure **5.54** we can see how the samples of Latium are pretty separated in three groups unequivocally enclosed in the three ellipses.
The olive oils of the northern group (province of Viterbo) are well separated from the other two groups, that are closer to each other and share a sample in the area of interception of their ellipses of confidence. Anyway the sensitivity and specificity of this analysis are high, as evident from the information matrix. (Table **5.11**)

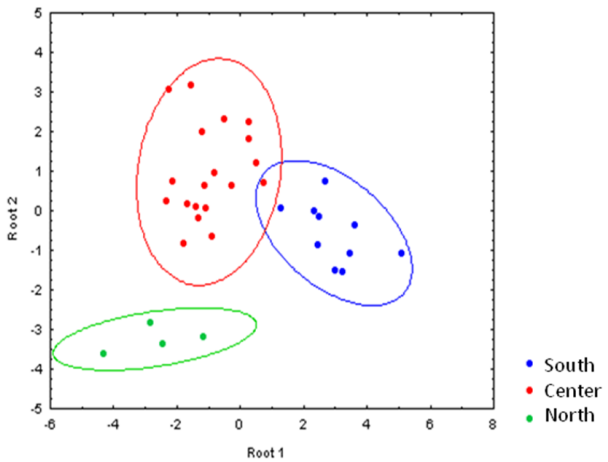


Figure 5.54: LDA of the olive oils from Latium of the first harvesting year, grouped by geographic areas (North, Centre and South).

Table 5.11: Information matrix for the LDA of the olive oils from Latium of the first harvesting year, grouped by geographic areas

	Predicted classes			
Observed classes		South	Centre	North
	South (10)	10	0	0
	Centre (19)	1	19	0
	North (4)	0	0	4

To get more information about the most discriminating variables we can evaluate their Wilk's λ values (Table **5.12**): the lower Wilk's λ value, the higher discriminating power. In this case the most important variables in the LDA of figure **5.54** are terpene 4, squalene, *sn* 1,2 diglycerides and linoleic and linolenic acids.

Table 5.12: Linear Discriminant Analysis of the olive oils from Latium concerning the geographical areas

N	**Variable**	**Wilks' Lambda**	F-remove (2,14)	p-level	Toler.
1	Hexanal	0,06124	0,348995	0,711362	0,100163
2	T-2-hexenal	0,080579	2,669699	0,104183	0,08651
3	**Terpene 4**	**0,058518**	0,022368	0,977915	0,15172
4	Terpene 3	0,061556	0,386946	0,68617	0,041226
5	Terpene 2	0,062624	0,515071	0,608352	0,044036
6	Terpene 1	0,083591	3,031176	0,080579	0,160223
7	*sn1,3*-Diglycerides	0,069009	1,281243	0,308331	0,106254
8	***sn1,2*-Diglycerides**	**0,059108**	0,093123	0,911641	0,61061
9	**Linolenic ac. (diall.)**	**0,058741**	0,049077	0,952271	0,24876
10	Linoleic ac. (diall.)	0,064911	0,789482	0,473289	0,048757
11	**Squalene**	**0,059552**	0,146433	0,86509	0,185617
12	Unsaturated. F.A.	0,06201	0,441356	0,651811	0,186137
13	Saturated. F.A.	0,062083	0,450196	0,646416	0,221358
14	Wax	0,060551	0,266281	0,770018	0,044125
15	**Linolenic ac. (CH3)**	**0,059357**	0,123035	0,885178	0,199493
16	**Linoleic ac. (CH3)**	**0,059495**	0,139566	0,870931	0,064502
17	Sitosterol	0,065718	0,886314	0,434078	0,160723

The PCA in figure **5.55** doesn't show a very clear differentiation of the olive oils from Latium on the basis of the altimetric criterion.

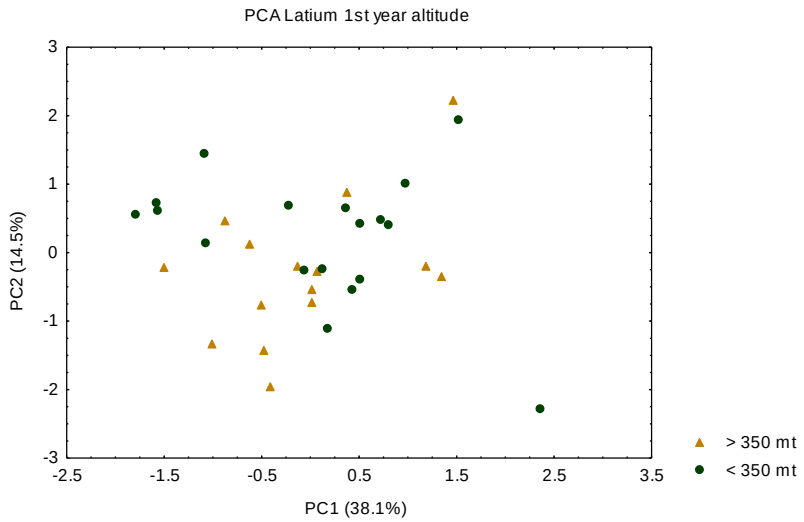
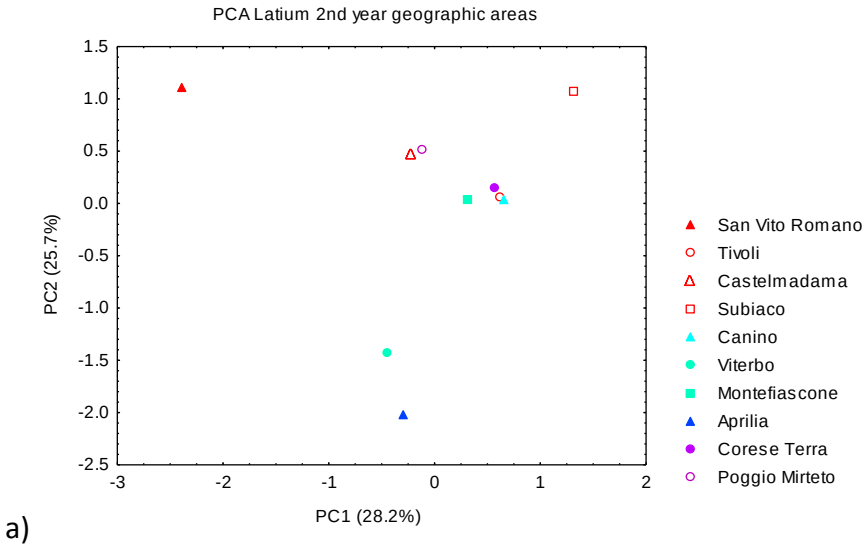


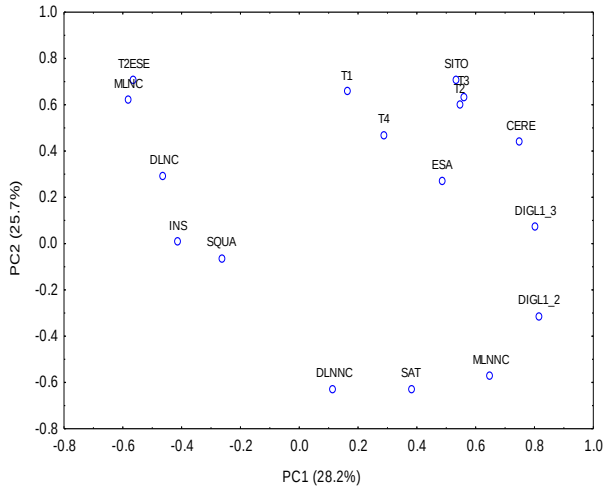
Figure 5.55: PCA of the olive oils from Latium of the first harvesting year, labelled by the altitude.

5.7.2 Statistical analysis of the olive oils of the second year

Ten olive oils from the second harvesting year were analyzed: 4 samples coming from the province of Rome, 3 samples from that of Viterbo, 2 samples from that of Rieti and 1 sample from the province of Latina.
A first PCA is shown in figure **5.56**, with the samples labelled by the geographic origin: the olive oil from Latina, with the blue symbol, is separated from the others along the PC2, that explain the 25.7% of the total variance, at negative values, while all the others, excepted a sample from the province of Viterbo, are grouped in correspondence of its positive values. The olive oil from Latina shows a high concentration of linolenic acid and low levels of aldehydic compounds, sitosterol and terpene 1.



a)



b)

Figure 5.56: a) PCA of the oils from Latium of the second harvesting year, labelled by geographic areas: the symbols of the oils from the province of Latina are blue, those for Rome are red, those for Rieti are purple and those for Viterbo are azure. b)Loadings of the PCA.

Labelling the samples by the altitude to which they are produced (figure **5.57**) we can see a clear effect, more than the first year, when the altitude was not so important. It seems that the altitude is the discriminating factor: samples from altitude higher than 350 m ASL are placed at PC2 values higher than the others. The olive oils from plain present a higher concentration of

linolenic acid and saturated fatty acids, but lower level of trans-2-hexenal and terpene 3 and 4.

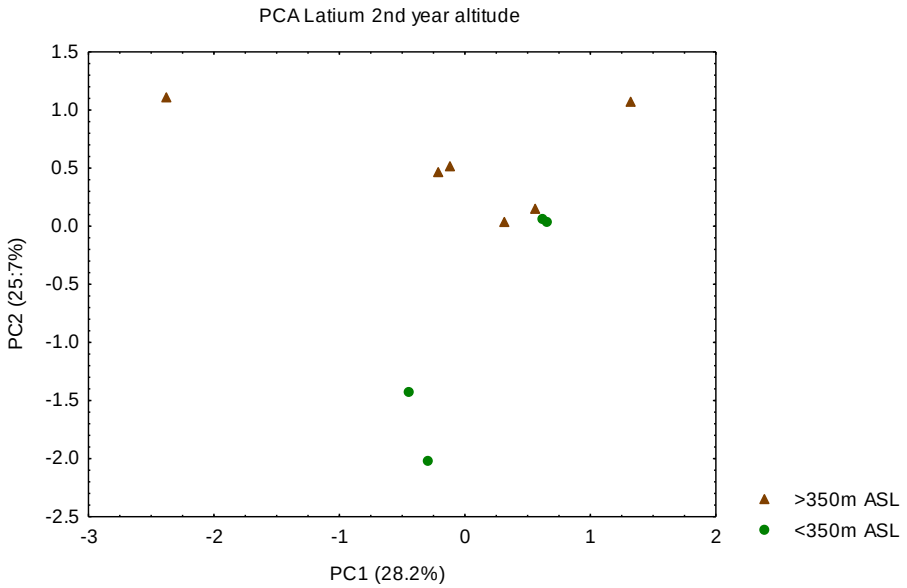


Figure 5.57: PCA of the oils from Latium of the second harvesting year, labelled by the altitude.

The figure **5.58** shows PCA concerning the cultivars of the olive oils from Latium. The 4 samples from the province of Rome are the most differentiated. The monovarietal sample Rosciola is separated along the PC1, that explain the 28.2% of the total variance, from the other multivarietal samples from the same province. We can also see how the only multivarietal sample from the province of Rome, that doesn't contain the cultivar Rosciola, is placed on the opposite side of the graph, along the PC1.

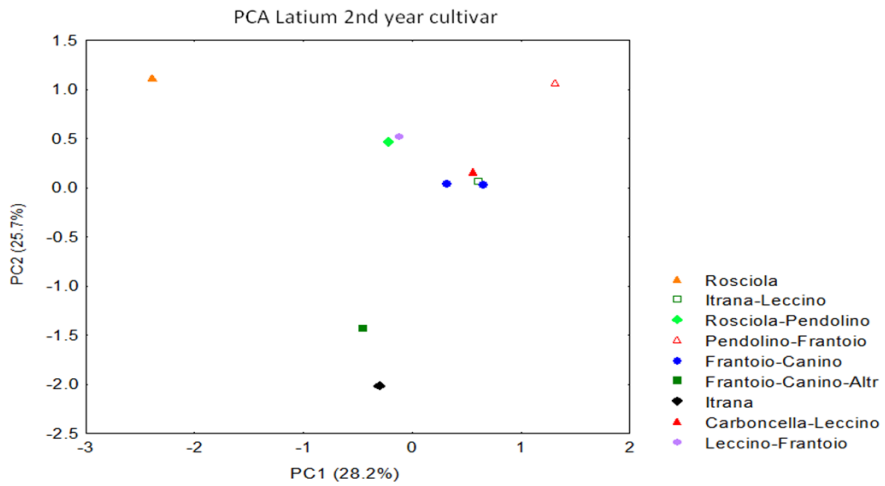


Figure 5.58: PCA of the oils from Latium of the second harvesting year, labelled by the cultivar.

5.7.3 Comparison between olive oils of the two harvesting years

Performing the analysis on both the harvesting years we didn't recognize clear seasonal effect (see figure **5.59**), probably due to a different sampling performed during the two harvesting years.

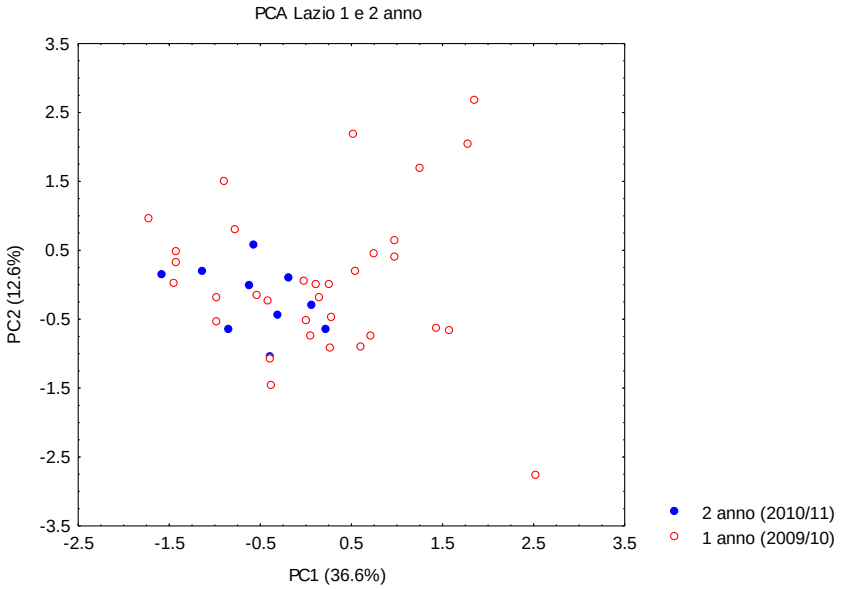


Figure 5.59: PCA of the olive oils from Latium of both the harvesting years.

Limiting the analysis to the olive oils whose sampling has been repeated (Rome, Rieti and Viterbo) we obtain the graph of figure **5.60**. All these provinces show a partial seasonal effect, moving to more negative values of the PC1, that explain the 31.8% of the total variance.
In the samples from Rieti and Rome we have a decrease of the level of wax, aldehydic compounds and terpene 3, while the content of squalene increases. The same trend, excluding the decrease of the aldehydic compounds, is also visible in the samples from Viterbo.

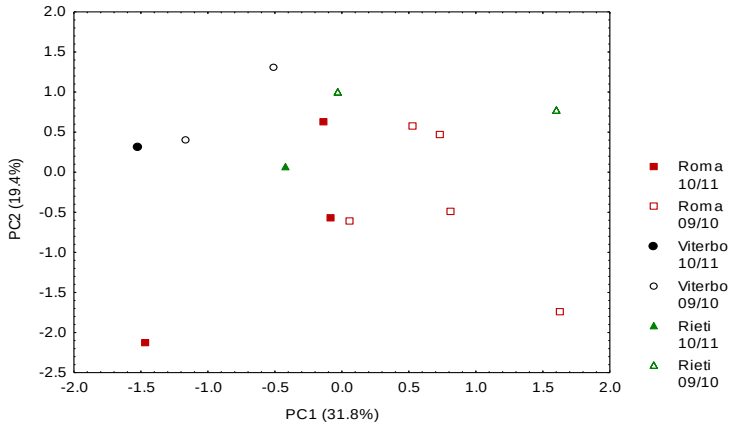


Figure 5.60: PCA of the oils from Latium collected in both the harvesting years.

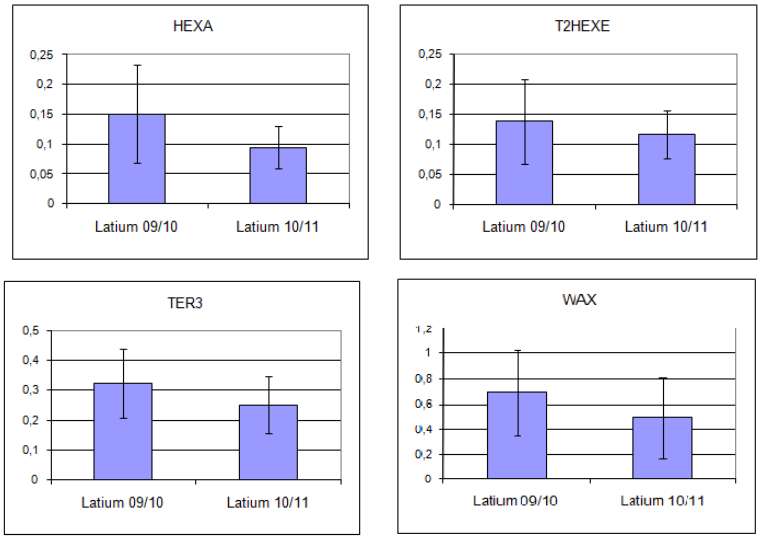
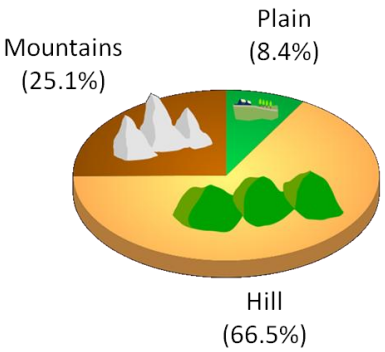


Figure 5.61: mean values of the signal of hexanal, trans-2-hexenal, terpene 3 and wax, in the oils from Latium of both the harvesting years.

5.8 Characterization of olive oils from Tuscany



The olive trees were harvested in this fertile region from the second half of the VII cent. b.C. and they always played a key role for the ecological, social, cultural and economic life of this region.
This oil has the only Protected Geographical Indication (PGI) in Italy and it is obtained from the following cultivars (alone or combined): Americano, Arancino, Ciliegino, Frantoio, Grappolo, Gremignolo, Grossolana, Larcianese, Lazzero, Leccino, Leccio del Corno, Leccione, Madonna dell'Impruneta, Marzio, Maurino, Melaiolo, Mignolo, Moraiolo, Morchiaio, Olivastra Seggianese, Pendolino, Pesciatino, Piangente, Punteruolo, Razzaio, Rossellino, Rossello, San Francesco, Santa Caterina, Scarlinese, Tondello (and their synonyms).
The colour of this oil ranges from green to gold, depending on the time; the aroma is fruity with a hint of almonds, artichoke, ripe fruit and green leaf; the flavour is strongly fruity.[62]

Table 5.13: oils from Tuscany, with the respective information

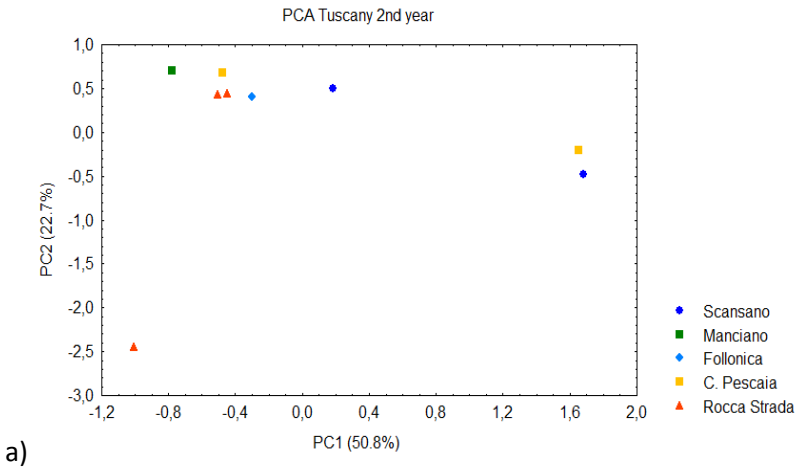
HARVESTING YEAR	CULTIVATION SITE	ALTITUDE (m above sea level)	CULTIVAR
09/10	Roccastrada (GR)	300	Frantoio, Leccino, Pendolino
09/10	Caldana di Gavorrano (GR)	300	Frantoio, Leccino, Pendolino
09/10	Roselle di Grosseto (GR)	300	Frantoio, Leccino, Pendolino
10/11	Scansano (GR)	100	Moraiolo, Leccino, Pendolino
10/11	Castiglion della Pescaia (GR)	100	Moraiolo, Leccino, Pendolino
10/11	Castiglion della Pescaia (GR)	100	Moraiolo, Leccino, Pendolino
10/11	Manciano (GR)	100	Moraiolo, Leccino, Pendolino
10/11	Scansano (GR)	100	Moraiolo, Leccino, Pendolino
10/11	Roccastrada (GR)	100	Moraiolo, Leccino, Pendolino
10/11	Roccastrada (GR)	100	Moraiolo, Leccino, Pendolino
10/11	Follonica (GR)	100	Moraiolo, Leccino, Pendolino
10/11	Roccastrada (GR)	500	Moraiolo, Leccino, Pendolino

5.8.1 Statistical analysis of the olive oils of the first year

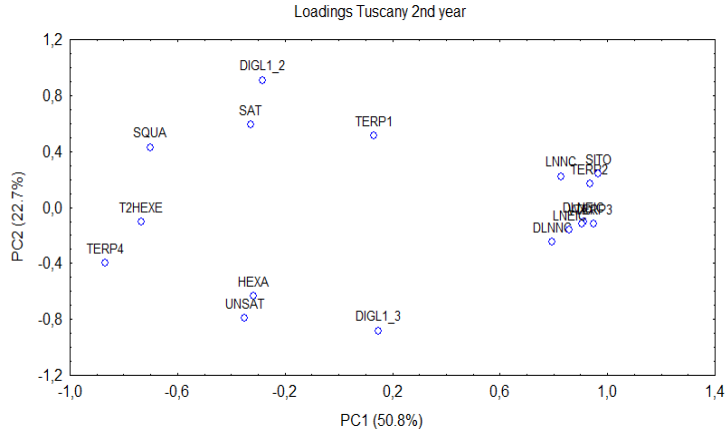
During the first year we had at our disposal only 5 samples, 4 from the Grosseto province and 1 from the Livorno province. The small number of samples restrained us from doing the statistical analysis.

5.8.2 Statistical analysis of the olive oils of the second year

The 9 samples, all produced in the Grosseto province, the most southern one on the borderline with Latium, are characterized by the same genetic composition: they present the same cultivars, Moraiolo, Leccino and Pendolino. Figure **5.62** shows a PCA with samples labelled by the production site: there is a strong similarity between 2 samples from Rocca Strada, while the third one from this place is strongly differentiated from them by the high level of the aldehydic compounds and terpene 4 and the low content of sitosterol, linolenic acid and terpenes 1 and 2. This sample is the only one coming from a harvesting site at 500 m ASL, while all the others come from places not higher than 100 m ASL.



a)



b)

Figure 5.62: a) PCA of the Tuscan oils of the second harvesting year, labelled by geographic origin. b)Plot of Loadings of the PCA.

5.8.3 Comparison between olive oils of the two harvesting years

Comparing the two harvesting years, we can easily see, in the PCA of figure **5.63**, how the samples of the two seasons are clearly separated from each other along the PC2, that explain the 19.0% of the total variance: olive oils produced in the second year are richer of squalene and poorer of trans-2-hexenal, linoleic acid and saturated fatty acids.

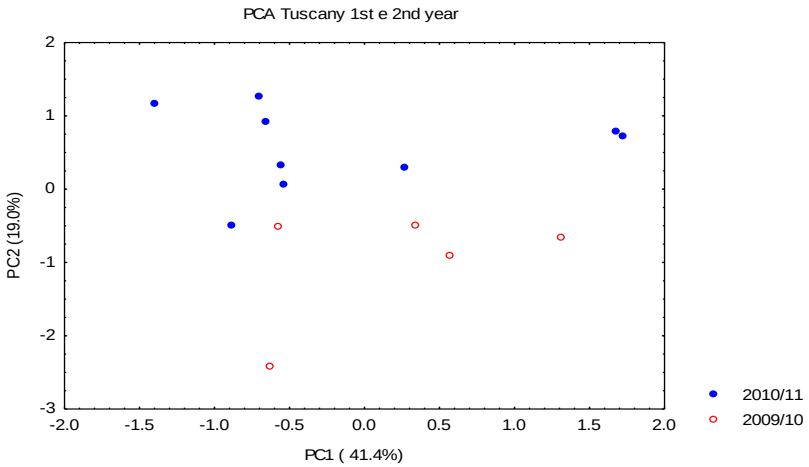


Figure 5.63: PCA of the Tuscan olive oils of both the harvesting years.

The comparison among the Tuscan olive oils, especially for the second year, with a higher number of samples, and those from the other regions brings out the fact that the former are definitely rich of terpenes 1, 2 and 3, wax (only in the oils from Liguria we have an higher level of this compound) and sitosterol and poor of squalene. (Figure **5.64**)

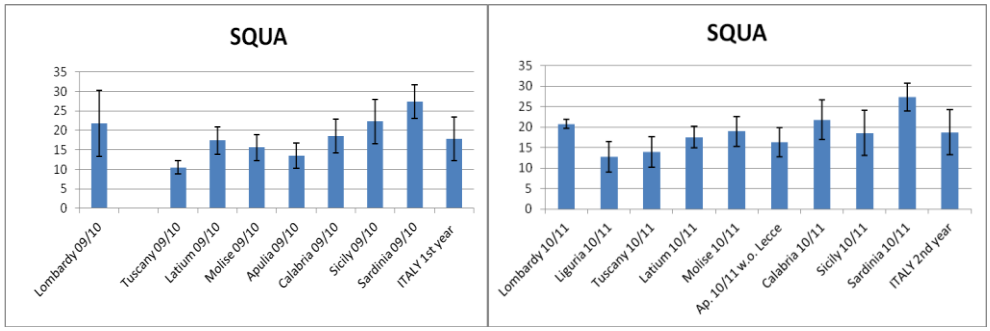


Figure 5.64: mean values of the signal of squalene in the olive oils from the different regions of Italy of both the harvesting years.

5.9 Characterization of olive oils from Liguria

Olive tree are one of the most important elements in the Ligurian landscape: introduced in this region from 3000 b.C., only at the end of the XVIII cent. they became widespread on the hills and mountains of this region.
Ligurian olive oil is perfect as a mild condiment not changing the taste of delicate foods.
The PDO is associated to one of the following geographic references: from Imperia "Riviera dei Fiori" (cultivar: Taggiasca, max 90%; colour: yellow; aroma: ripe fruity; flavour: fruity, with strong sweetness); from Savona "Riviera di Ponente Savonese" (cultivar: Taggiasca, max 60%; colour: yellow-green; aroma: ripe fruity; flavour: fruity, with strong sweetness); from Genova and La Spezia "Riviera del Levante" (cultivar: Lavagnina, Razzola, Pignola, min. 65%; colour: green-yellow; aroma: ripe fruity; flavour: fruity with a medium taste of sweetness and eventually a light sensation of bitter and spicy).[62]

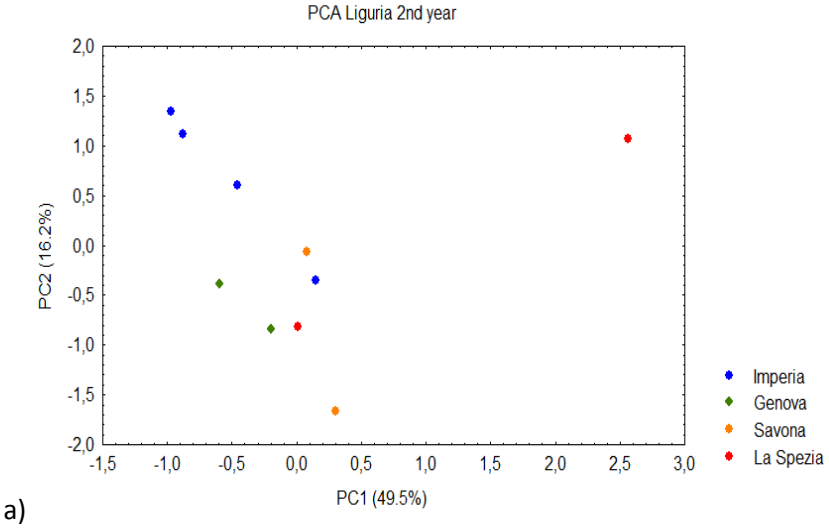
Table 5.14: olive oils from Liguria, with the respective information

HARVESTING YEAR	CULTIVATION SITE	ALTITUDE (m above sea level)	CULTIVAR
10-nov	Borgomaro IM	300	Faggiasca
10/11	Pontedassio (IM)	220	Faggiasca
10/11	Diano Castello (IM)	100	Faggiasca
10/11	Diano Marina (IM)	100	Faggiasca
10/11	Loc S.Giulia - Lavagna (GE)	100 - 200	Lavagnina, Razzola, Leccino, Pimola
10/11	Loc Lemeggio - Moneggia (GE)	40	Faggiasca
10/11	Fraz.Castellaro - Stellanello (SV)	250	Fraggiasca
10/11	Fraz.Marmoreo - Casanova (SV)	400	Faggiasca
10/11	Loc. Masignano - Arcola (SP)	250	Frantoio 60%, Leccino 40%
10/11	Loc. Casano - Ortonovo (SP)	80	Razzola 50%, Leccino 30%

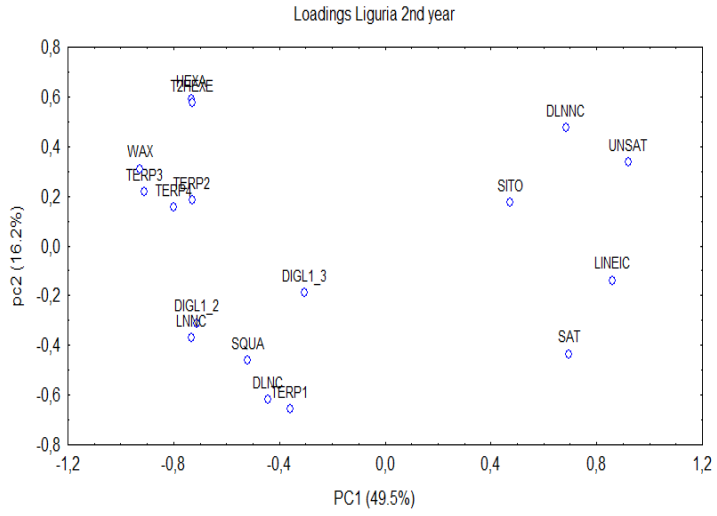
5.9.1 Statistical analysis of the olive oils of the second year

The 10 samples from Liguria, collected only in the second year, are divided as follows: 4 samples from the Imperia province, 2 samples from the Genova province, 2 samples from the Savona province and 2 samples from the La Spezia province.

In the PCA shown in figure **5.65a** it is possible to see how the samples from Genova are close to each other: they present the lowest concentration of unsaturated fatty acids and a highest concentration of sitosterol. Also the samples coming from Imperia are placed in a restricted zone of the graph: they are characterized by highest levels of the aldehydic compounds in this region, with the exception of the sample from Diano Marina.



a)



b)

Figure 5.65: a) PCA of the olive oils from Liguria of the second harvesting year labelled by geographic origin. b)Plot of Loadings of the PCA.

Also in this case the genetic factor seems to explain the differences highlighted by the PCA (figure **5.66**): labelling all the samples by cultivar it is possible to see that olive oils from Imperia and Savona (Riviera di Ponente), and one from the Genova province, are monovarietal olive oils, Taggiasca 100%. The other sample from Genova is composed mainly by cultivar Lavagnina, while a sample from the province of La Spezia contains mostly the cultivar Razzola. Notably, all these olive oils present typical Ligurian

cultivars (Taggiasca, Lavagnina e Razzola) and they gather in the left area of the graph. The only sample containing only the cultivars Frantoio and Leccino, is clearly separated along the PC1, that explain the 49.5% of the total variance. This sample comes from the La Spezia province and its composition is totally different from the other Ligurian olive oils, because it is rich of sitosterol, saturated and unsaturated fatty acids and poor of wax, squalene and all the terpenic and aldehydic compounds.

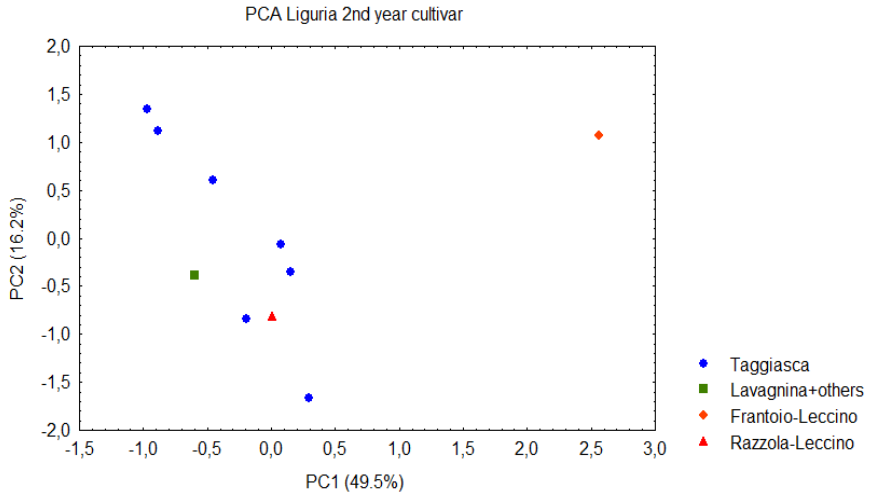
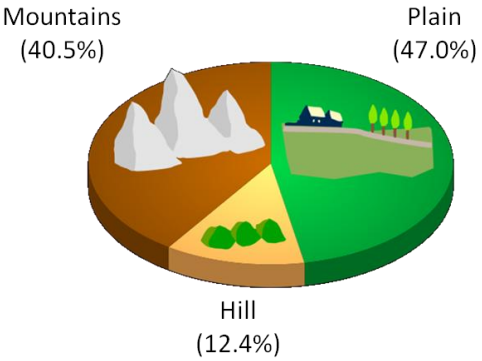


Figure 5.66: PCA of the olive oils from Liguria of the second harvesting year labelled by cultivar

Although Ligurian olive oils have been collected only during the second harvesting year, it is possible to highlight some characteristics, distinguishing them from the other Italian samples. In fact their composition is pretty different from the olive oils of the second harvesting year coming from all the other regions: in fact the have the highest content of wax, aldehydes, terpene 1, 2, 3 and unsaturated fatty acids, and the lowest level of squalene and saturated fatty acids, and also a low concentration of linoleic acid.

5.10 Characterization of olive oils from Lombardy



Introduced in the Roman period by settlers from Magna Greece, the olive tree became an important element of the local landscape thanks to the ideal, mild climate of the neighbourhood of the major Lombard lakes.
The definition of “Lombard Lakes Extra Virgin Olive Oil” includes the olive production from Lake Como and Iseo, located in the so called oil valley: this is a place between Griante and Sala Comancina, along the western coast of the Lario.[65]

Table 5.15: oils from Lombardy, with the respective information

HARVESTING YEAR	CULTIVATION SITE	ALTITUDE (m above sea level)	CULTIVAR
09/10	Rodengo Saiano (BS)	330-350	-
09/10	Rodengo Saiano (BS)	330-350	-
09/10	Padenghe del Garda (BS)	330-350	Leccino Frantoio
09/10	Padenghe del Garda (BS)	330-350	Leccino Frantoio
10/11	Moniga - BS	150	Casaliva 100%
10/11	Alto Garda	200	Casaliva 100%
10/11	Cazzago San Martino (BS)	150	Leccino, Frantoio, Pendolino
10/11	Predore (BG)	-	Leccino, Frantoio, Casaliva, Pendolino
10/11	Musso (varie località) (CO)	250-350	Frantoio, Leccino, Pendolino
10/11	Perledo (LC)	250-400	Leccino, Frantoio, Pendolino

5.10.1 Statistical analysis of the olive oils of the first year

We only have four samples, all coming from the Brescia district, so we couldn't perform the statistical analysis for this harvesting year

5.10.2 Statistical analysis of the olive oils of the second year

Six samples were collected during the second year, 3 from the Brescia province (one from lake Garda zone), 1 from the Bergamo province, 1 from the Lecco province and 1 from the Como province. Figure **5.67** shows the PCA of the samples labelled by the harvesting area: they are rather different among themselves; even within the group of samples coming from the Brescia province the only common element is the high value of terpene 3. The sample from Bergamo and the sample from Alto Garda are placed in two regions of the graph well separated from each other and from the other samples: in fact these two samples present a higher content of sitosterol and saturated fatty acid compared to the other samples of this region.

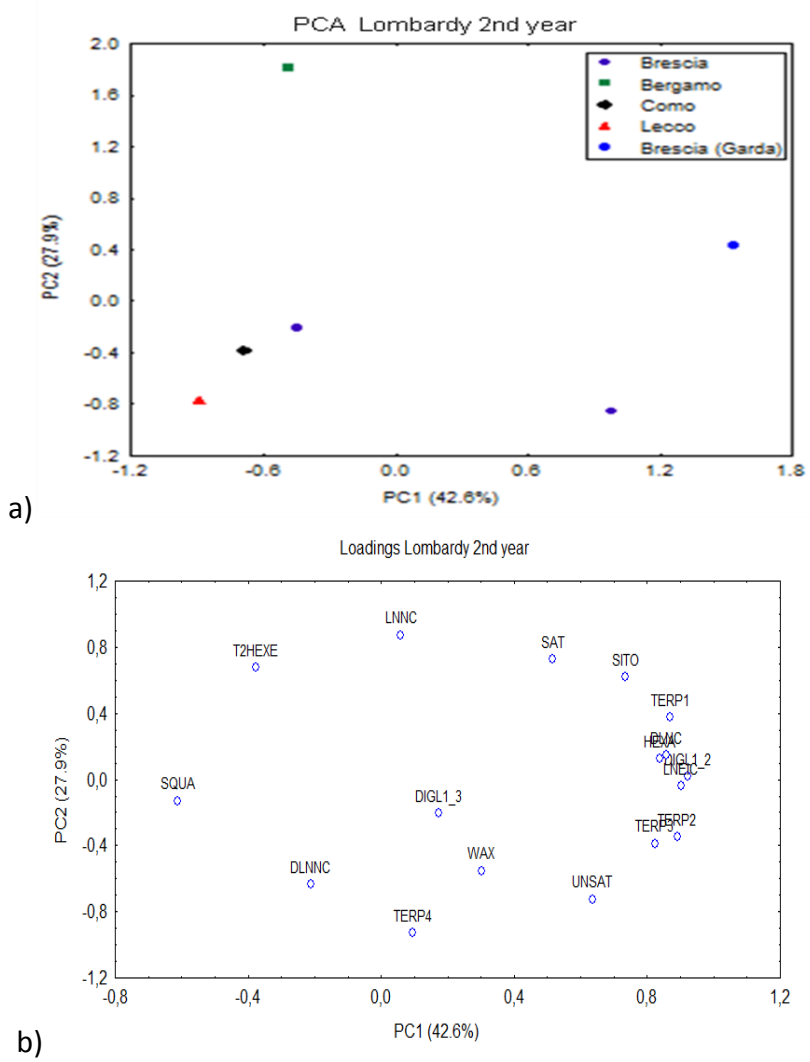


Figure 5.67: a) PCA of the olive oils from Lombardy of the second harvesting year, labelled by geographic origin. b)Plot of loadings of the PCA.

The distribution of samples in PCA is better explained by the cultivar composition: relabeling the samples by this criterion we can see that the three oils grouped in the bottom left of the graph are obtained by combining three cultivars: Leccino, Pendolino and Frantoio. The two samples separated along the PC1, explaining 42.2% of the total variance, are monovarietal Casaliva and present high level of hexanal, terpenes 1, 2 and 3.

The small number of samples from this region didn't allow an accurate characterization of the oils, but we can see a clear separation between the samples 100% Casaliva and all the others. The red sample of figure **5.68**, coming from the Bergamo province, is composed by the same cultivars (Leccino, Pendolino and Frantoio) of the samples indicated in green, but it includes also a 20% of Casaliva cultivar, that makes it different from the other one.

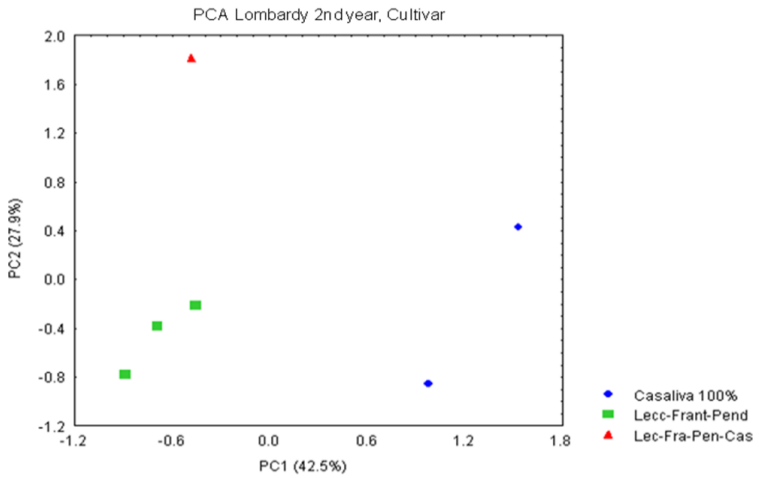


Figure 5.68: PCA of the Lombard oils of the second harvesting year, labelled by the cultivar.

5.10.3 Comparison between olive oils of the two harvesting years

Comparing the two harvesting years we can see a clear seasonal effect, shown in the PCA of figure **5.69**, where the olive oils are separated along the PC1, that explains the 37.4% of the total variance. In particular samples of the second year are richer of unsaturated fatty acid, but poorer of aldehydic compounds, linoleic and linolenic acid and saturated fatty acids, as we can see in figure **5.70**.

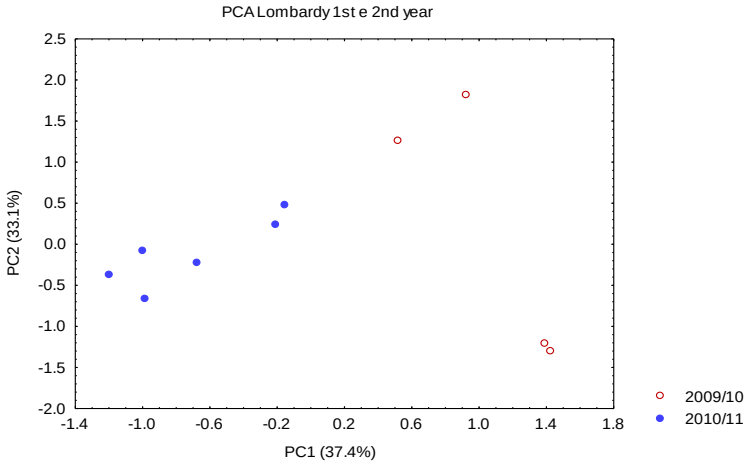


Figure 5.69: PCA of the Lombard oils of both the harvesting years.

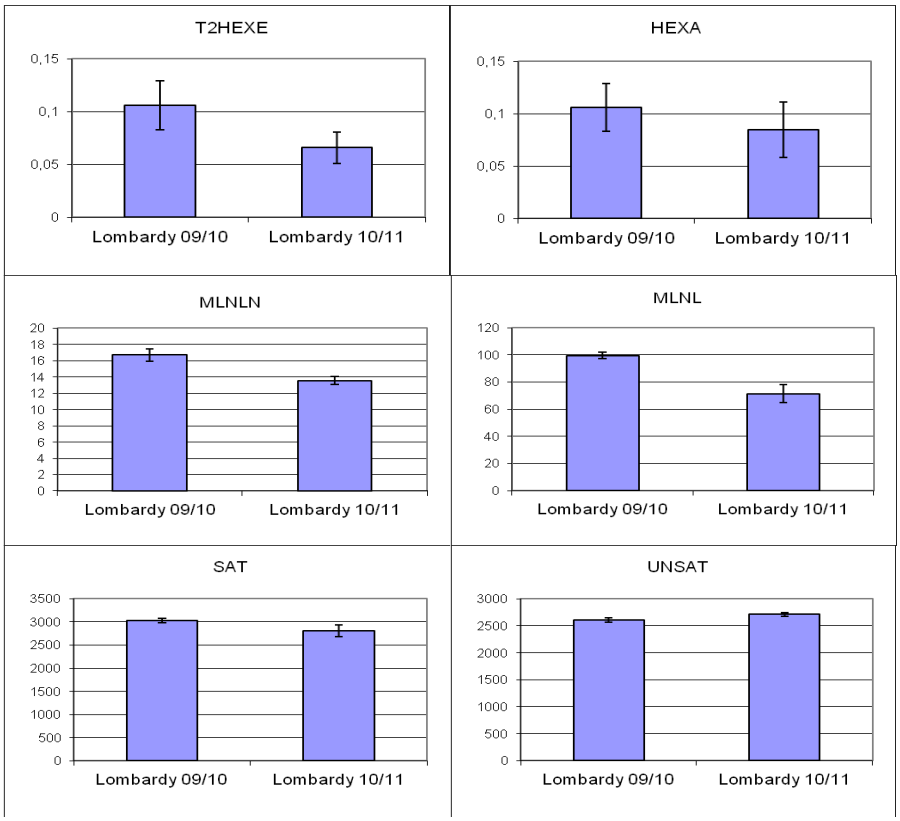


Figure 5.70: mean values of the signal of wax, saturated and unsaturated fatty acids, aldehydic compounds and linoleic acid in the Lombard olive oils of the two harvesting years.

6. Conclusions

This work was performed within the agreement between the Institute of Chemical Methodologies of CNR and UNAPROL (Italian Olive Oil Consortium) with the aim to develop an analytical control system able to support the chain traceability, for the improvement and the certification of the high quality virgin Italian olive oils.

The NMR technique has confirmed its potential in the analysis and the study of olive oils. In particular the olive oil has been studied successfully for several years with this technique. In fact the NMR spectroscopy allows to reach information about all the compounds contained in the olive oil, both the most abundant and the minor ones, with the same experimental error, that, in addition, is very small. Its ability to give the complete metabolic fingerprint of the olive oil allows us to step into the metabolomics.

Combining the NMR analysis with the chemometric statistical analysis interesting results have been obtained about the characterization of the olive oils according to their genetic and geographic origin and also to ecological, pedoclimatic and agronomic factors.[33,34,36]

During this work we followed the NMR and statistical analysis protocol developed and used during the past years in the study of olive oil at the CNR laboratory, for the analysis of the Italian olive oils sampled by UNAPROL in Lombardy, Tuscany, Latium, Molise, Calabria, Apulia, Sicily and Sardinia during the harvesting years 2009-10 and 2010-11, and in Liguria during 2010-11.

The data of the 1H-NMR experiments, put through the statistical analysis, provided the characterization profiles of the olive oils for each region and allowed the classification and the comparison of the oils, within the whole Country and within each region and between the two harvesting year.

For the single region the characterization concerned the genetic factor, taking in consideration the effect of the cultivar from which the oils were obtained, and the geographic origin, considering the province or the area (North, Centre, South) from which the oils were sampled and, when it was possible, considering the nature of the soil and the ripeness degree.

Thanks to the sampling carried out evenly during the two harvesting seasons, it has been possible to highlight the effect of the seasonal factor,

but also to point out the existence of some variables not affected by this factor, that could be considered characteristic elements of each region.
Concerning the variables less affected by the seasonal factor we noticed that:
The oils from Sardinia have a content of squalene and linoleic acid higher than those from the rest of Italy, but they are poorer, as the Sicilian oils, of terpene 1.
The oils from Sicily are rich of linoleic acid and poor of terpene 3 and 1.
The oils from Tuscany are rich of terpenes 1 and 2, sitosterol and wax, but poor of squalene.
The oils from Calabria show constant values of compounds such as aldheydes, terpenes 2 and 3, variables often affected by the seasonal effect in the other regions, and present also constant values of linoleic acid and unsaturated fatty acids. In particular the concentration of the terpenes 2 and 3 is very low among the peninsular regions.

The seasonal effect is visible as a decrease of the aldheydic compounds from the first to the second harvesting year in all the regions examined excepted Sicily, where the trend is inverted. Also the terpenes 2 and 3, along with linoleic acid and wax reach a lower level in the second year. The squalene concentration, instead, increases from the first to the second year, but also in this case the Sicily behaves differently from the rest of the Italy, together with Lombardy.
Variables such as unsaturated and saturated fatty acids and sitosterol do not show a clear seasonal effect. Generally the difference between the harvesting years is more evident in Tuscany, Lombardy, Molise, Sicily and Sardinia, while it is less definite in Latium and Apulia. The only region not affected by the seasonal factor is Calabria.
Notably the linoleic acid in the second harvesting year shows an interesting trend, with a clear increasing gradient, particularly in the in the olive oils from the second harvesting year, from Northern to Southern regions, reaching maximum levels in the two Islands. Anyway it must be taken in consideration that the samples from Liguria are available only for the second harvesting year.

Considering each region we highlighted the effect of the genetic factor, the geographic origin and the altitude of production.

The first one, for example, allows in the Sicilian olive oils a clear differentiation, mostly in the first harvesting year, among the oils of Nocellara, Biancolilla and Cerasuola cultivar: the last present a low content of squalene and saturated fatty acids and high level of terpene 3 in both the harvesting year.
The genetic factor is often linked to the geographic factor: some cultivar are typically farmed in a specific zone of the region, as we see in Sardinia, where the cultivar Tonda di Cagliari is typically found in the Cagliari province while the cultivar Bosana is more common in the Sassari province and in both the harvesting years the two groups are separated in PCA.
The geographic origin is really important also in the olive oils from Latium, especially in the first harvesting year, when we can easily distinguish the oils coming from the North, the Centre or the South of this region. This differentiation due to the areas of origin was already observed in a previous paper.[34]
The altitude is not an a factor with univocal consequences: for the regions whose data were available for both the harvesting years, the different altitude shows its influence only in one of the two years, as Molise and Calabria, or it doesn't shows it at all, as in Latium.
For some of the samples of the first harvesting year we analyzed the influence of the soil where the olive tree was farmed: this can be seen in both the regions that had this kind of information, Molise and Apulia.
Finally the role of the ripeness degree has been considered for Sicily and Molise, and both the regions present their oils grouped in accordance with the ripeness.
In conclusion we can say that this work confirmed the potential of the NMR analysis as fundamental instrument in the characterization of the Italian olive oil.
Moreover, for the first time it has been possible to carry out a study involving olive oils collected by a national scale sampling, allowing a comparison among different regions, but this sampling was also repeated evenly for two consecutive years, allowing the study of the seasonal effect on the olive oil composition.

4. References

1. Mannina, L.; Luchinat, C.; Emanuele, M.C.; Segre, A. “Acyl Position distribution of glycerol tri-esters in vegetable oils: a 13C NMR study.” *Chem. and Phys. of Lipids.* **1999**, 103, 47-55.

2. Emanuele, M.C. et al. “L'olio di oliva e la risonanza magnetica nucleare protonica ad alto campo.” *Olivo&Olio*. **1998**, 2/3, 53-57.

3. Moreno, J.J.; Mitjavila, M.T. “The degree of unsauration of dietary fatty acids and the development of atherosclerosi (review).” *J Nutr Biochem.* **2003**, 14(4), 182-95.

4. Nicolosi, R.J.; Woollfrey, B.; Wilson, T.A.; Scollin, P.; Handelman, G.; Fisher, R. “Decreased aortic early atherosclerosis and associated risk factors in hypercholesterolemic hamsters fed a high- or mid-oleic acid oil compared to a high linoleic acid oil.” *J Nutr Biochem.* **2004**, 15 (9), 540-7.

5. Reaven, P.; Parthasarathy, S.; Grasse, B.J.; Miller, E.; Steinberg, D.; Witztum, J.L. “Effects of oleate-rich and linoleate-rich diets on the susceptibility of low density lipoprotein to oxidative modification in mildly hypercholesterolemic subjects.” *J Clin Invest.* **1993**, 91 (2), 668-76.

6. Owen, R.W.; Mier, W.; Giacosa, A.; Hull, W.E.; Spiegelhalder, B.; Bartsch, H. “Phenolic compounds and squalene in olive oils: the concentration and antioxidant potential of total phenols, simple phenols, secoiridoids, lignans and squalene.” *Food Chem Toxicol.* **2000**, 38, 647-59.

7. Chimi, H.; Sadik, A.; Le Tutour, B.; Rahamani, M. “Contribution à l’ètude comparative des pouvoirs antioxydants dans l’huile d’olive du tyrosol, de l’hydroxytyrosol, de l’acide cafeique, de l’oleuropeine et du BHT.” *Rev Franc Corps Gras.* **1988**, 35 (8/9), 339-344.

8. Servili, M.; Baldiolo M.; Miniati E.; Montedoro G.F. “Antioxidant activity of new phenolic compounds extracted from virgin olive oil and their interaction with α-tocopherol and β- carotene.” *Riv Ital. Sostanze Grasse.* **1996,** 73 (2), 55-59.

9. Reaven, P.D.; Khouw, A.; Beltz, W.F.; Parthasarathy, S.; Witztum, J. “Effect of dietary antioxidant combinations in humans. Protection of LDL by vitamin E but not by b -carotene.” *Arterioscler Thromb. Vasc. Biol.* **1993**, 13, 590.

10. Rahamani, M.; Saari Csallany, A. “Role of minor constituents in the photooxidation of virgin olive oil.” *JAOCS.* **1998**, 75 (7), 837-843.

11. Gimeno, E. et al**.** “The effects of harvest and extraction methods on the antioxidant content (phenolics, alpha-tocopherol, and beta-carotene) in virgin olive oil.” *Food Chemistry* **2002**, 78(2), 207-211.

12. Arrigo, L.; Rondinone, R. “I micronutrienti “eu-ossidanti” nell’olio di oliva.” *Riv Ital Sostanze Grasse* **1995**, 72, 11-14.

13. Ruggiero, N.; Meloni M. “*La valorizzazione degli oli a denominazione di origine.”* **2000,** p. 1-145.

14. Balestrieri, F.; Marini, D. *Metodi di analisi chimica dei prodotti alimentari.* s.l. : ed. M editrice, **1996**, 1, 1-344.

15. De Leonardis, A. et al. *La produzione dell'olio di oliva vergine di qualità.* Campobasso : ERSAM Molise. 1-43.

16. Olìas, J.M. et al. “Aroma of virgin olive oil: biogenesis of the green odor notes.” *J. Agric. Food Chem.* **1993**, 41, 2368-2373.

17. Millqvist Fureby, A. et al. “Preparation of diglycerides by lipase-catlyzed alcoholysis of triglycerides.” *Enzyme and Microbiological Technology* **1997**, 20, 198-206.

18. Mèndez, A.I.; Falquè E. “Influence of container type and storage tine on olive marc oil quality.” *Electronic J. of Environmental Agricultural and Food Chemistry* **2002**, 1 (2), 71-95.

19. Taner Baysal, Aslıhan Demirdöven *Enzyme and Microbial Technology* **2007**, 40, 491–496

20. Ravalli, M., “La qualità nel frantoio.” s.l. : Roma: ARM, Azienda Romana Mercati, **2004**, 1-103.

21. Corradetti, R.; Di Muzio, R.; Molfese M. “L'olio extravergine di oliva - la filiera della qualità: dall'oliveto al mercato.” Agenzia regionale per i servizi di sviluppo agricolo Abruzzo, **2003**, 1-63.

22. Lercker, G.; Rodriguez-Estrada M.T. “Chromatographic analysis of unsaponifiable compounds of olive oils and fat-containing foods.” *J. of Chromatography A* **2000**, 881(1-2), 105-129.

23. Cercaci, L.; Rodriguez-Estrada M.T.; Lercker G. “Solid-phase extraction–thin-layer chromatography–gas chromatography method for the detection of hazelnut oil in olive oils by determination of esterified sterols.” *J. of Chromatography A* **2003**, 985, 211-220.

24. Aparicio, R.; Aparicio Ruìz R. “Authentication of vegetable oils by chromatographic techniques.” *J. of Chromatography A* **2000**, 881(1-2): 93-104.

25**.** Gordon, M. H.; Firman, C. “Effects of heating and bleaching on formation of stigmastadienes in olive oil.” *Journal of the Science of Food and Agriculture,* **2001**, 81 (15), 1530–1532.

26. Fiehn, O. *Plant Molecular Biology* **2002**, 48, 155-171.

27. Mannina, L.; D'Imperio, M.; Lava, R.; Schievano, E.; Mammi, S. “Caratterizzazione NMR e analisi statistica di oli di oliva DOP veneti = NMR characterization and statistical analysis of PDO in olive oils from Veneto” ***La Rivista Italiana delle Sostanze Grasse*** 2005**, 81, 59-63 .**

28. Mannina, L.; Patumi, M.; Fiordiponti, P.; Emanuele, M.C.; Segre, A.L. *Italian Journal of Food Science* **1999**, 11, 139-149.

29. Mannina, L.; Patumi, M.; Proietti, N.; Bassi, D.; Segre, A.L. *J. Agric. Food Chem.* **2001**, 49, 2687-2696.

30. *“Development and Assessment of Mehods for the Detection of Adulterazion of Olive with Hazelnut oil”,* **MEDEO progetto Europeo n° GRD1-2000-25011.**

31. Mannina, L.; D’Iimperio, M.; Capitani, D.; Rezzi, S.; Guillou, C.; Mavromoustokos, T. et al. “^{1}H NMR-Based Protocol for the Detection of

Adulterations of Refined Olive Oil with Refined Hazelnut Oil" *J. Agric. Food Chem.* **2009**, 57, 11550–11556

32. Garzia-Gonzales, D.L.; Mannina, L.; D'Imperio, M.; Segre, A.L.; Aparicio, R. *Eur. Food Res. Technol.* **2004**, 219, 545-548 .

33. Mannina, L.; Patumi, M.; Proietti, N.; Segre, A.L. "P.D.O (Protected designation of origin): Geographical characterization of Tuscan extra virgin olive oils using High-field 1H-NMR Spectroscopy" . *Ital. J. Food Sci.* **2001**, 13, 1.

34. D'Imperio, M.; Mannina, L.; Capitani, D.; Bidet, O.; Rossi, E.; Bucarelli, F.M.; Quaglia, G.B.; Segre, A.L. *Food Chemistry* **2007**, 105 (3), 1256-1267 .

35. Mannina, L.; Marini, F.; Gobbino, M.; Sobolev, A.P.; Capitani, D. "NMR and chemometrics in tracing European olive oils: The case study of Ligurian samples". *Talanta* **2010**, 80, 2141-2148.

36. Mannina, M.; Dugo, G.; Salvo, F.; Cicero, L.; Ansanelli, G.; Calcagni, C.; Segre, A.L. "Study of the cultivar-composition relationship in Sicilian olive oils by GC, NMR and statistical methods" *J. Agric. Food Chem.* **2003**, 51, 120-127.

37. Duarte, I.F.; Barros, A.; Belton, P.S.; Righelato, R. et al. *J. Agric. Food Chem.* **2002**, 50, 2475-2481.

38. Duarte, I.F.; Barros, A.; Almeida, C.; Spraul, M.; Gil, A.M. *J. Agric. Food Chem.* **2004**, 52, 1031-1038.

39. Mannina, L.; Cristinzio, M.; Sobolev, A.P.; Ragni, P.; Segre, A.L. "High-field Nuclear Magnetic Resonance (NMR) study of truffles (tuber aestivum vittadini)." *J. Agric. Food Chem.* **2004**, 52, 7988-7996.

40. Moing, A.; Maucourt, M.; Renaud, C.; Gaudillere, M. et al. **"**Quantitative metabolic profiling by 1-dimensional ^{1}H-NMR analyses: application to plant genetics and functional genomics" *Funct Plant Bio.* **2004**, 31, 889-902.

41. Belton, P.S.; Colquhoun, I.J.; Kemsley, E.K.; Delgadillo, I. et al. *Food Chem.* **1998**, 61, 207-213 .

42. Le Gall, G.; Puaud, M.; Colquhoun, I.J. *J. Agric. Food Chem.* **2001**, 49, 580-588.

43. Sobolev, A.P.; Segre, A.L. Lamanna, R. "Proton high-field NMR study of tomato juice. *Magnetic Resonance in Chemistry."* **2003**, 41, 237-245.

44. Holmes, E.; Tang, H.; Wang, Y.; Seger, C. *Planta Med.* **2006**, 72, 771-785.

45. Krishnan, P. Kruger, N.J.; Ratcliffe, R.G. *Journal Exper. Botany.* **2005**, 56, 255-265 .

46. Le Gall, G.; Colquhoun, I.J.; Davis, A.L. et al. *J. Agric. Food Chem.* **2003**, 51, 2447-2456.

47. Choi, H.K.; Choi, Y.H. Verberne, M.; Lefeber, A.W.M.; Erkelens, C.; Verpoorte, R. *Phytochemistry* **2004**, 65, 857-864.

48. Manetti, C.; Bianchetti, C.; Bizzarri, M.; Casciani, L.; Castro, C.; D'Ascenzo, G. et al. *Phytochemistry* **2004**, 65, 3187-3198.

49. Baker, J.M.; Hawkins, N.D.; Ward, J.L.; Lovegrove, A. et al. *Plant Biotech. J.* **2006**, 4, 381-392 .

50. Charlton, A.; Allnutt, T.; Holmes, S.; Chisholm, J.; Bean, S.; Ellis, N. et al. *Plant Biotech J.* **2004**, 2, 27-35.

51. Defernez, M.; Gunning, Y.M.; Parr A.J.; Shepherd, L.V.T.; Davies, H.V.; Colquhoun, I.J. "NMR and HPLC- UV profiling of Potatoes with Genetic Modifications to Metabolic Pathways." *Journal of Agricultural and Food Chemistry.* **2004**, 52, 6075-6085.

52. Mood, A.M.; Graybill, F.A.; Boes, D.C. **"**Introduzione alla statistica" Milano : McGraw Hill, **1997**.

53. Martens, H.; Martens, M. "Multivariate Analysis of Quality - An Introduction." s.l. : Wiley Interscience Publication, **2001**.

54. Beebe, K.R.; Pell, R.J.; Seasholtz, M.B. “Chemometrics - A Practical Guide“ s.l. : Wiley Interscience Publication, **1998**.

55. Del Nobile, M.A. et al. “Influence of packaging geometry and material properties on the oxidation kinetic of bottled virgin olive oil.” *J. Food Engineering.* 2003, 57, 189-197.

56. Mattei, A.; Burattini, M.; Zanoni, B. “Effetto della tipologia di confezionamento primario sulla conservabilità di oli extravergini di oliva.“ *La Rivista Italiana delle Sostanze Grasse.* **2005**, 82, 65-70.

57. Angerosa, F.; Basti, C.; Vito, R. “Virgin olive oil volatile compounds from Lipoxygenase pathway and characterization of some italian cultivars.” *J. Agric. Food Chem.* **1999**, 47, 836-839.

58. Williams, M. et al**.** “Analysis of volatiles from callus cultures of olive Olea europea.” *Phytochemistry* **1998**, 47, 1253-1259.

59. Morales, M.T.; Rios, J.J.; Aparicio, R. “Changes in the volatile composition of virgin olive oil during oxidation: flavors and off-flavors.” *J. Agric. Food Chem.* **1997**, 45, 2666-2673.

60. Mannina, L; Segre, A.L. “High resolution nuclear magnetic resonance: from chemical structure to food authenticity.” *Grasas y Aceites**.*** **2002**, 53, 22-23.

61. Zerilli, V. *Teatro Naturale.* Dip. Colture Arboree Univ. Studi Palermo, 20 Settembre 2003, n. 3 Anno 1.

62. http://www.italystore.com/olio/caratteristiche.asp

63.http://www.sardegnaagricoltura.it/index.php?xsl=443&c=3592&s=45303&v=2

64. http://www.qualigeo.eu/prodotti/d26d8827-70f5-439e-b8f1-5c652d3a292b/OlioextraverginediolivaMolise.aspx

65. http://www.buonalombardia.it/browse.asp?goto=11733

Part III

NMR structural analysis of polysaccharides

1. Introduction

1.1 Carbohydrates structure determination[1]

Carbohydrates play key roles in biological recognition processes,[2,3] in development of diseases,[4] and in many important areas of food and technical industry. Nevertheless medicinal chemists have doomed complex carbohydrates as uninteresting in drug development: they are difficult to synthesize and manipulate,[5] and the limited availability makes it difficult to study their biological functions in detail; carbohydrates larger than a disaccharide are generally too complex to be prepared in sufficient amounts for testing and too hydrophilic to have good bioavailability, and furthermore, they are generally orally inactive and easily degraded in the digestive tract.

It is important, however, to encourage and continue the study of carbohydrate-mediated biological processes and to investigate their behaviour, as understanding the mechanism of carbohydrate function may lead to the development of carbohydrate-based therapeutics. Recent advances in glycochemistry have helped to solve some of the problems associated with studies of carbohydrates, and methods for the large-scale synthesis of complex carbohydrates for drug development have been documented.

The building blocks in oligosaccharides are more diverse in nature than in proteins or nucleic acids. Carbohydrates often differ from each other only in the stereochemistry, and the pattern of interresidue linkages can be very heterogeneous.[6] The information capacity of the carbohydrates is much larger than in proteins, particularly due to branched structures. It has been claimed that carbohydrates contain the hidden code to biological recognition.[6-8] Consequently, the structure determination of complex oligosaccharides is a challenging but difficult problem.

1.2 Structural information from 1D NMR spectra

A wealth of techniques[9-19] are currently applied in the identification of known oligosaccharides or the determination of new structures. The identification of known structures can be carried out on relatively small amounts of material using, e.g., capillary electrophoresis[10] or fluorescence detected HPLC.[9] Both techniques can be combined with mass spectrometry to give important structural information.[11,12] An interesting technique has been developed for structure elucidation of glycoprotein-derived oligosaccharides based on the combined use of specific hydrolases and HPLC.[20] Mass spectrometry of oligosaccharides[11,16,21,22] is rapidly developing following improvements in instrumentation. However, the crucial issue of sugar stereochemistry cannot be solved by current methodology used routinely in mass spectroscopy.
Recently the NMR analysis has shown its potential and utility for the structure elucidation, especially when combined with data from mass spectrometry or chemical information, such as monosaccharide composition or methylation analysis.[23]

1.2.1 Number of sugar residues

A good starting point for a structural analysis is the anomeric proton chemical shift. Integration of the anomeric resonances offers an initial estimate on the number of different monosaccharide residues present. The anomeric proton resonances are found in the shift range 4.4-5.5 ppm. The remaining ring proton resonances are found in the range 3.0-4.2 ppm in unprotected oligosaccharides. Additionally, the number of anomeric C1 resonances present in a 1D ^{13}C NMR spectrum will confirm the number of different residues. Such results can also be obtained from 2D ^{1}H- ^{13}C HSQC,[24] HMQC,[25.26] or HMBC[27]spectra, which in many cases are more sensitive than a 1D ^{13}C spectrum.

1.2.2 Anomeric configuration

Normally the α-anomer resonates downfield compared to the β-anomer in D-pyranoses in $^{4}C_{1}$ conformation. The vicinal coupling constant between the anomeric H1 and the H2 indicates the relative orientation of the two protons: if they are both in an axial configuration in pyranose structures, a large coupling constant (7-8 Hz) is observed, whereas if they are equatorial-

axial, this is smaller ($J_{1,2}$ ~4 Hz), and for axial-equatorial or equatorial-equatorial oriented protons, even smaller coupling constants are observed (<2 Hz).[28] This principle can be used when assigning the relative orientation of protons in a hexopyranose ring as first demonstrated by Lemieux et al. in 1958.[29] The ^{13}C chemical shift reveals the anomeric configuration in a manner similar to the proton chemical shifts, but most importantly the one bond ^{1}H-^{13}C coupling constants in pyranoses can be used to determine the anomeric configuration unequivocally.[30]

1.2.3 Linkage and sequence

The multiplicity of possible linkages between monomeric units adds a further level of structural complexity to carbohydrates because they may be linked between the anomeric hydroxyl group of one monosaccharide and any other hydroxyl-bearing carbon in another monosaccharide, yielding for aldohexoses a total of four possible linkages (excluding links between an anomeric hydroxyl and another anomeric carbon, that can only lead to the formation of disaccharides).[31]

The determination of linkages is often the primary objective of an NMR study of a carbohydrate obtained from natural sources. Even with this apparently simple objective, the perennial problem of limited chemical shift dispersion may require supplementary data from enzymatic and chemical degradation and methylation analysis. In the latter, all free hydroxyl groups of the native saccharide are tagged with a methyl group[32] prior to hydrolysis and reductive acetylation; this yields a series of partially methylated alditol acetates that are identified by gas chromatography or gas chromatography–mass spectrometry. The non-methylated positions in the characterized alditol acetates correspond to positions that were involved in linkages in the native saccharide. The number of possible linkages to be considered in the NMR study is thereby clearly reduced.

Both the ^{1}H and the ^{13}C chemical shift may give an indication of the linkage type if the chemical shifts for the specific linkage have been reported previously. The effect of glycosylation depends on the linkage type, and the changes in the chemical shift are in general larger at the glycosylation site than at neighbouring positions. Interresidue NOEs may give information about the glycosidic linkage, but it should be kept in mind that the strongest NOE might not be between the protons across the glycosidic linkage.[33,34] A

HMBC experiment can also give linkage information, keeping in mind that both intra- and interresidue correlations are seen.

1.2.4 Position of appended groups

The proton and carbon chemical shifts are sensitive to the attachment of a non-carbohydrate group like a methyl, acetyl, sulphate, or a phosphate group. Attachment of such a group will affect the proton and carbon resonances where the group is located. Normally downfield shifts ~0.2-0.5 ppm are observed[35] for protons and higher Δδ values for ^{13}C. This places these resonances in a less crowded area of the spectra and helps the identification of modified residues. Such appended groups may also contain NMR-active nuclei, which may give rise to additional splitting due to couplings (i.e. ^{31}P-^{1}H long-range couplings). The use of other homo- or heteronuclear correlations may help in the determination of their position.

1.2.5 Structural information from ^{13}C NMR data[31]

To obtain a reliable assignment of the ^{13}C spectra of oligosaccharides, [36,37,38] there are few guidelines that can be kept in mind:

1. In the absence of structural modification, a sugar residue in an oligosaccharide has ^{13}C chemical shifts that are usually within ±0.3 ppm of the corresponding resonances in the free monosaccharide.
2. Glycosylation leads to high frequency shifts of 4–10 ppm for the carbons at the anomeric and linked positions.
3. Resonances of carbon atoms that are adjacent to each carbon in the glycosyl linkage are in general shifted to slightly lower frequency.
4. The chemical shift of an anomeric carbon in an α-linkage (δ 97–101 ppm) is *in general* less than that of an anomeric carbon in a β-linkage (δ 103–105 ppm), but the exceptions to this rule recommend determination of $^{1}J_{C\text{-}1,H\text{-}1}$ as the most reliable indication of anomeric configuration.
5. ^{13}C nuclei in furanose sugars are less shielded than in their configurationally related pyranose counterparts.

1.2.6 Classical NMR methods

Carbohydrate samples are normally dissolved in D_2O, and when working with carbohydrates containing charged groups, like phosphates, we should care to report the pH of the sample.[39,40] This is particularly relevant when more samples or structures should be compared. If the molecule is not water soluble different solvents may be used, such as DMSO, methanol-d4,

or $CDCl_3$. It can be an advantage to use a mixture of H_2O/D_2O (i.e. 9:1) to be able to observe exchangeable protons as amide H in the NAc group.
Spectra are often acquired at close to room temperature, but raising the temperature can give better resolution by sharpening of the resonances, lowering the viscosity of the sample, and increasing the tumbling rate.
^{13}C chemical shifts are often obtained from multidimensional heteronuclear proton observe experiments using an inverse detection probe because more sensitive and requires less compound.

Although existent 2-D NMR techniques can yield unequivocal assignments, these empirically derived rules still provide useful starting points and confirmatory checks of 2-D NMR data.

1.3 Structural information from 2D NMR spectra

The 2D experiments are based on the submission to the sample of vary pulse sequences, structured in four fundamental phases: preparation, evolution, mixing time and detection. Among these, the evolution time is progressively increased of a constant Δ, and the acquired series of monodimensional spectra, once transformed, gives a two-dimensional plot.
The base of the 2D-NMR techniques is the scalar coupling, that results in the *J* constant concerning the homonuclear (^1H-^1H) or heteronuclear (^1H-^{13}C) interaction.
Here we report a brief description of the most used 2D experiment.

- COSY (**CO**rrelation **S**pectroscop**Y**): this 2D experiment is a homonuclear 2D technique that is used to correlate the chemical shifts of ^1H nuclei J-coupled to another one.

The different pulse sequences can be explained as follows: the first pulse creates transverse magnetization components which generate chemical shift and homonuclear J-coupling during the evolution period t1. Following this variable time period, a second pulse mixes the spin states, transferring magnetization between coupled spins. The spectra is then acquired during t2 (detection time). The COSY spectrum is processed by a 2D Fourier transform with respect to t1 and t2; the monodimensional protonic

spectrum lays along the diagonal of the 2D spectrum and the cross peaks, symmetric with respect to this, indicate which ^{1}H nuclei are J-coupled.

- TOCSY (**TO**tal **C**orrelation **S**pectroscop**Y**): in TOCSY, cross peaks are generated between all members of a coupled spin network.

In traditional COSY, the coherence transfer is restricted to directly spincoupled nuclei. In TOCSY, an oscillatory exchange is established, which proceeds through the entire coupling network so that there can be a net magnetization transfer from one spin to another even without direct coupling.
The coherence transfer period of the TOCSY sequence occurs during a multiple-pulse spin-lock period. The length of the spin-lock period determines how far the spin coupling network will be probed: five transfer steps are typically desired for the TOCSY spectrum.
The monodimensional protonic spectrum lays again along the diagonal of the 2D spectrum and there are more cross peaks for each proton, symmetrically disposed on respect to this, indicating which ^{1}H nuclei belong to the same spin system.

- HSQC (**H**eteronuclear **S**ingle **Q**uantum **C**orrelation): this 2D experiment correlates the chemical shift of a proton with the chemical shift of the directly bonded heteronucleus, in our case carbon. Cross peaks give the shift of the corresponding proton and carbon. This experiment utilizes the one-bond coupling between carbon and proton (J=120-215 Hz). On the horizontal axis is placed a proton spectrum and on the other one is placed a carbon spectrum.

- HMBC (**H**eteronuclear **M**ultiple **B**ond **C**orrelation): this experiment utilizes multiple-bond couplings over two or three bonds (J=2-15Hz). Cross peaks are generated between protons and carbons that are two or three bonds away (and sometimes up to four or five bonds away). Direct one-bond cross-peaks are suppressed. This experiment provides connectivity information over several bonds and is very valuable to detect indirectly quaternary carbons coupled to protons - especially useful if direct ^{13}C spectrum is impossible to obtain due to low amount of material available.

This very useful experiment provides information about the skeleton of a molecule. It is also very useful in the carbohydrates study as a sequence

analysis tool that provides unique information concerning connectivity across glycosidic linkages.

2. Structural investigation of Scleroglucan and its derivative Scl-CM-80

2.1 Introduction[41]

Scleroglucan (Scl) is a water-soluble extracellular polysaccharide produced by fungi of the genus Sclerotium. Its structure presents a backbone of (1→3)-β-linked glucose residues, substituted with single (1→6)-β-D-glucopyranosyl residues on every third backbone unit. [42]

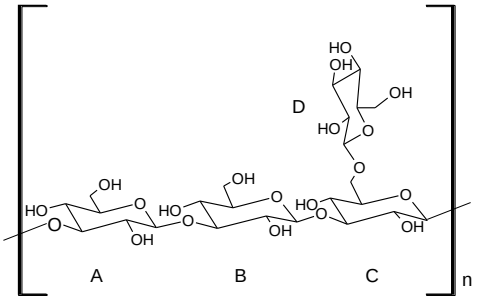


Figure 2.1: Representation of scleroglucan repetitive unit.

The X-ray fibre diffraction data indicate that scleroglucan adopts a triple helix structure (figure **2.2**), that is destroyed in solutions at pH higher than 12 or in dimethylsulfoxide (DMSO).[43,44]

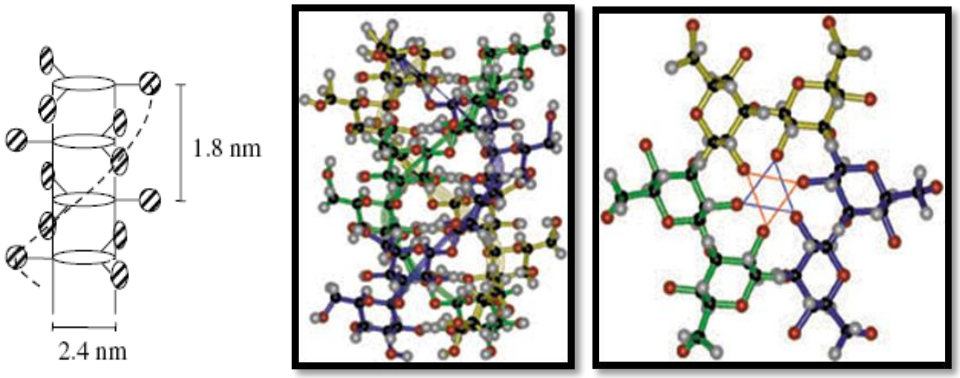


Figure 2.2: scleroglucan triple helix structure

Scleroglucan and some of its derivatives at different oxidation degrees have been employed as starting materials for the synthesis of chemical and physical hydrogels suitable as matrices for the controlled release of drugs.[45-50] Recently the synthesis of chemical gels of scleroglucan obtained by reaction of the native polymer with 1,ω-dicarboxylic acids with a different number of carbon atoms in the chain has been reported.[51] Other chemical gels have been prepared with different derivatives of scleroglucan, obtained by the controlled oxidation of the polymer. These reactions yield the opening of the glucose of the side chain, giving rise, respectively, to scleraldehyde and carboxylated scleroglucan (sclerox), employed for the synthesis of chemical gels.[52-54] As it happens for many other charged polysaccharides, physical gels of scleroglucan have been prepared starting from sclerox in the presence of cations.[55] Moreover, scleroglucan is able to form physical gels with the addition of borax; the structural characteristics and the employment of these hydrogels for the controlled delivery of drugs of different steric hindrance have been extensively investigated.[56,57]
Recently, in the laboratories of Prof. M. A. Casadei, Sapienza University of Rome, a carboxymethyl derivative (Scl-CM) of scleroglucan has been synthesized, with the aim to obtain a polymer able to afford a controlled drug release, depending on the salt concentration.[41]
This polymer has been obtained by the reaction of scleroglucan and chloroacetic acid in alkaline water solution, and it has been characterized using FT-IR and potentiometric titration. The latter provided the derivatization degree as % of repetitive units in the total scleroglucan mass that underwent the linkage with the carboxymethyl group. The amount of acid in mmol (equal to the mmol of NaOH employed during the titration) was inserted into the equation:

$$\text{amount of polymer (mg)} = X \bullet PM_{(\text{repetitive unit})} + (\text{mmol of acid}) \bullet PM_{(\text{acid group})}$$

where X are the mmol of repetitive units of the polymer.
The ratio between the mmol of acid and the mmol of the repetitive units of the polymer multiplied by 100 gave the degree of derivatization (DD, number of carboxylic groups for 100 repetitive units).
However this method doesn't give any structural information about the derivatized product, i.e. which is the glucose unit, among the four of each repetitive unit, and which glucose position is submitted to the derivatization.

The minor steric hindrance of the position 6 on the D component of the repetitive unit suggested that this could be the most favorable target for the derivatization, but it was not clear whether it was the only derivatized position. To get these information the NMR characterization has been carried out for both the unmodified scleroglucan and its derivative Scl-CM-80, in order to understand the position of the carboxymethyl group into the scleroglucan backbone.

2.2 Materials and methods

2.2.1 Preparation of the samples

Scl and Scl-CM derivative samples were both characterized by NMR spectra recorded on a Bruker AVANCE AQS600, operating at the protonic frequency of 600.13 MHz and at ^{13}C frequency of 150.92 MHz. The instrument was equipped with a 5 mm inverse probe BBI, properly tuned on the characteristic protonic frequency in the applied field B_0 of 14.3 T.
First of all, the samples were dissolved in D_2O, and then freeze-dried according to a protocol already reported in literature for the treatment of other (1→3)-β-glucans prior of NMR analysis.[58] The protocol was repeated twice. The obtained polymers were dissolved in DMSO-d_6 (4mg of polymer/1 ml DMSO-d_6) and the ^1H and ^{13}C assignments were obtained with 2D ^1H-^1H COSY (COrrelation SpectroscopY), ^1H-^1H TOCSY (TOtal Correlation SpectroscopY), ^1H-^{13}C HSQC (Heteronuclear Single Quantum Coherences) and ^1H-^{13}C HMBC (Heteronuclear Multiple Quantum Coherences) at the temperature of 348 K (75°C).

2.2.2 NMR Spectra acquisition

^1H-NMR:
Time Domain (TD): 32K;
Sweep Width (SW): 6.74 ppm;
O1: 1891.14 Hz;
Receiver Gain: 256;
Delay (D1): 2.5 sec;
Number of Scans (NS): 256;
Dummy Scans (DS): 4;
Temperature: 348 K.

^1H-^1H-COSY:
Pulse program: cosyqf45;
Time Domain (TD): 1024;

Sweep Width (SW): 6.74 ppm;
O1: 1891.14 Hz;
Receiver Gain: 256;
Delay (D1): 2,5 sec;
Acquisition Time (AQ): 0.127 sec;
Number of Scans (NS): 48;
Dummy Scans (DS): 16;
Temperature: 348 K.

^{1}H-^{1}H-TOCSY:
Pulse program: mlevph;
Time Domain (TD): 1024;
Sweep Width (SW): 6.74 ppm;
O1: 1891.14 Hz;
Receiver Gain: 256;
90 degrees pulse;
Mixing time: 90 ms;
Delay (D1): 2,4 sec;
Acquisition Time (AQ): F_1=0.063 sec; F_2=0.127 sec;
Number of Scans (NS): 48;
Dummy Scans (DS): 16;
Temperature: 348 K.

^{1}H-^{13}C HSQC:
Pulse program: hsqcph;
Time Domain (TD): F_2=1024, F_1=512;
Sweep Width (SW): 6.74 ppm -130.00 ppm;
O1: 1891.14 Hz; O2: 9054.17 Hz;
Receiver Gain: 456;
90 degrees pulse;
CNST2: 140 Hz;
Delay (D1): 2,4 sec;
Acquisition Time (AQ): 0.127 sec;
Number of Scans (NS): 92;
Dummy Scans (DS): 16;
Temperature: 348 K.

^{1}H-^{13}C HMBC:
Pulse program: hmbclpndqf;
Time Domain (TD): F_2=1024, F_1=512;
Sweep Width (SW): 6.74 ppm -130.00 ppm;
O1: 1891.14 Hz; O2: 15090.28 Hz;
Receiver Gain: 512;
90 degrees pulse;
CNST2: 140 Hz;
CNST13: 5 Hz;
Delay (D1): 2,0 sec;
Acquisition Time (AQ): 0.127 sec;
Number of Scans (NS): 384;
Dummy Scans (DS): 16;
Temperature: 348 K.

2.2.3 NMR Spectra processing

After Fourier transformation, the ^{1}H spectra were processed performing a baseline correction and manual phase correction.
The chemical shifts are given in parts per million, using the internal DMSO residual signals as reference (^{1}H = 2.500 ppm and ^{13}C = 40.40 ppm).

2.3 Results and discussion

2.3.1 NMR characterization of the polysaccharides

The NMR spectra of polysaccharides are generally difficult to interpreter. The ^{1}H-NMR peaks are usually broad because of the low mobility of their chain in solution and the dissolution of these macromolecules is often difficult. When the spectra are not performed in D_2O, but in other solvents, such as DMSO-d_6, it is necessary to remove the signals of the hydroxyl groups of the polysaccharide and water by dissolving the sample in D_2O and lyophilizing it, repeating the procedure several times. However, even in this way, some water traces can still remain in the sample, and, working at room temperature, they can cover a region of the spectra that can be very important for the determination of the polysaccharide structure.
For all these reasons the NMR spectra of the polysaccharides are preferentially performed at high temperature. At high temperature, the solubility of the polymer and the mobility of its chains increase and the peaks become more sharp and resolved. All the following spectra have been performed at 348 K.
As previously said, all the techniques used for the Scl-CM-80 characterization didn't allow the identification of the position of the Scl backbone structure where the carboxymethyl group is linked: to get this information the NMR characterization of both the Scl and its derivative, Scl-CM-80, has been carried out, using DMSO-d_6 as solvent at 348 K.
Both the ^{1}H-NMR spectra (figures **2.3** and **2.4**) show two signals in the anomeric region: the one at 4.544 ppm is due to the β-anomeric protons of the three molecules of β-(1→3)glucose of the glucan backbone, namely A, B and C (figure **2.5**). They are practically equivalent, so their signals are not distinguishable. The other signal in the anomeric region is at 4.246 ppm and is generated by the β-anomeric proton of the β-(1→6)glucose branching unit.
The signal at 4.094 ppm is due to one of the protons of the branched C6 position, while the signal at 3.041 ppm belongs to the C2 proton of the branching unit.

Because of the overlapping of the other signals, to make a reliable attribution of the signals we had to perform some 2D-NMR experiments: first we characterized the Scl polysaccharide, by the ^{1}H-^{1}H COSY, ^{1}H-^{1}H TOCSY and the ^{1}H-^{13}C HSQC; then by the comparison of these spectra with those of the Scl-CM-80, we noticed, over the presence of the peaks of the units of the unmodified backbone, a set of new signals whose appearance was a direct consequence of the derivatization, and that could be used, therefore, to determine its position on the glucose units.

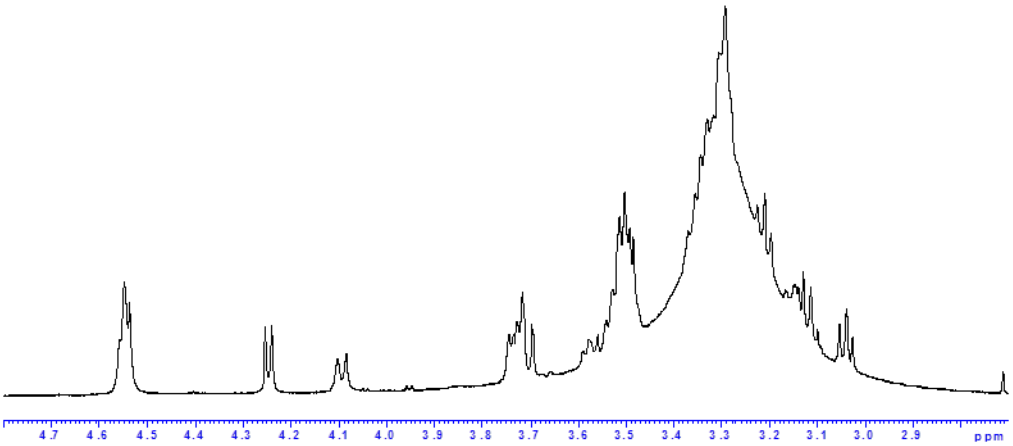


Figure 2.3: ^{1}H-NMR spectrum of Scl in DMSO-d_6 (T=348 K).

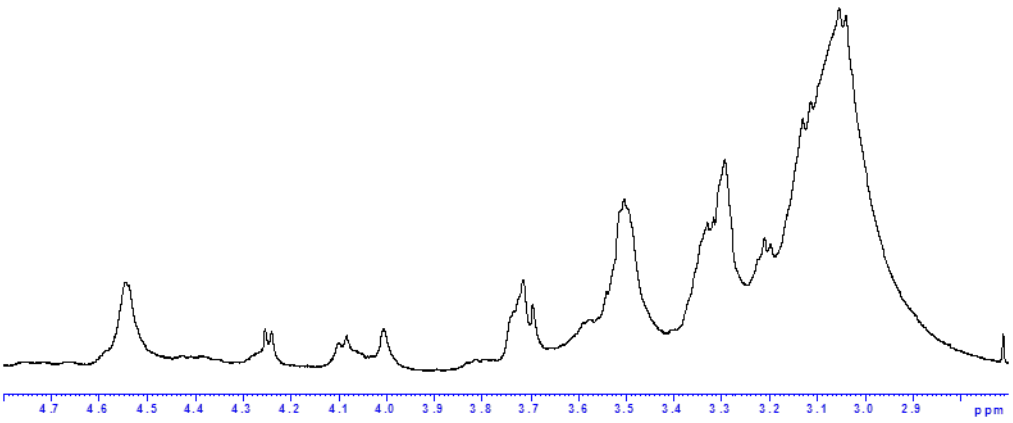


Figure 2.4: ^{1}H-NMR spectrum of Scl-CM-80 in DMSO-d_6 (T=348 K).

2.3.2 NMR characterization of Scleroglucan

By the combination of the information obtained from the 2D NMR spectra, it was possible to assign completely the ^{1}H spectrum of scleroglucan.
The assignment (table **2.1**) was consistent with the literature data.[59]

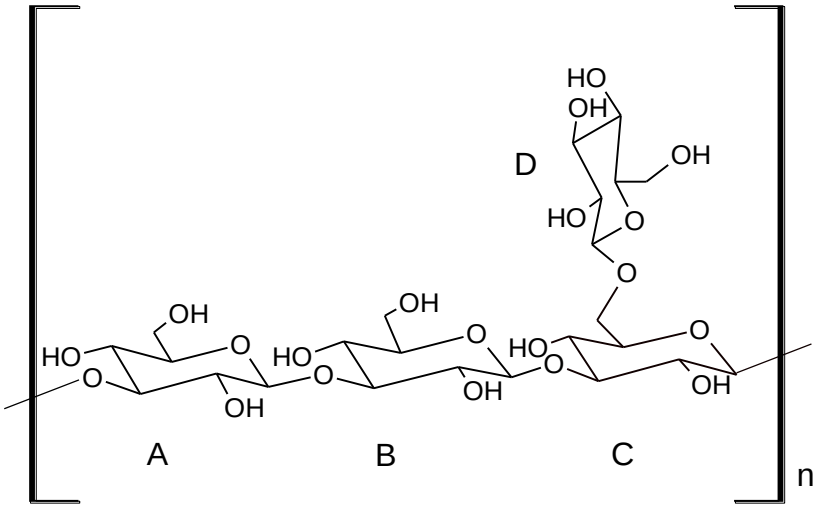


Figure 2.5: Representation of scleroglucan repetitive unit.

Table 2.1: ^{1}H and ^{13}C NMR assignment of Scl sample in DMSO at 348 K.

Residue	Type	^{1}H δ (ppm)	Multiplicity (*J* in Hz)	^{13}C δ (ppm)
1A	CH	4.544		103.3
2A	CH	3.334		73.0
3A	CH	3.502		86.7
4A	CH	3.293		68.9
5A	CH	3.293		76.6
6A	CH_2	**6A** 3.496	dd (11.3; 5.5)	61.3
		6'A 3.731	dd (11.3; 5.5)	
1B	CH	4.544		103.3
2B	CH	3.334		73.0
3B	CH	3.502		86.7
4B	CH	3.293		68.9
5B	CH	3.293		76.6
6B	CH_2	**6B** 3.496	dd (11.3; 5.5)	61.3
		6'B 3.731	dd (11.3; 5.5)	
1C	CH	4.544		103.3
2C	CH	3.334		73.0
3C	CH	3.502		86.7
4C	CH	3.293		68.9
5C	CH	3.528		75.2
6C	CH_2	**6C** 3.571	d (10.3; 8.5)	68.8
		6'C 4.094	d (10.3)	
1D	CH	4.246	d (7.9)	103.4
2D	CH	3.041	dd (8.4; 8.3)	74.0
3D	CH	3.212	t (8.76)	76.7
4D	CH	3.121	m	70.7
5D	CH	3.141	m	77.0
6D	CH_2	**6D** 3.497	dd (11.5; 5.5)	61.5
		6'D 3.704	dd (11.5; 1.9)	

In the map reported in figure all cross peaks are labeled according to the corresponding nucleus. (figure **2.6**, **2.7** and **2.8**)

In the ^{1}H-^{1}H-COSY two cross-peaks due to the coupling between the anomeric protons and the protons in position 2 of glucose rings, namely A, B, C and D were identified. The 1D proton give a cross-peak with the 2D at 4.246-3.041 ppm, and in the region of the spectrum between 3.00 and 3.22

ppm we can see the cross peaks due to the coupling between the proton 2D and the 3D, between the 3D and 4D and between the 4D and the 5D. We can also recognize the cross-peak of 5D with 6D.
The other three glucose units, A, B and C, share the same cross-peaks for the coupling between 1ABC and 2ABC, between 2ABC and 3ABC and between 3ABC and 4ABC. The brunched unit (C) presents specific cross-peaks for the coupling between the nuclei 4C and 5C, 5C and 6C and between the nuclei 6C-6'C. The A and B units share the same cross-peaks for the coupling between the nuclei 4 and 5 and between 5 and 6 and share also with the D unit the peaks of the coupling between the protons 6 and 6'.

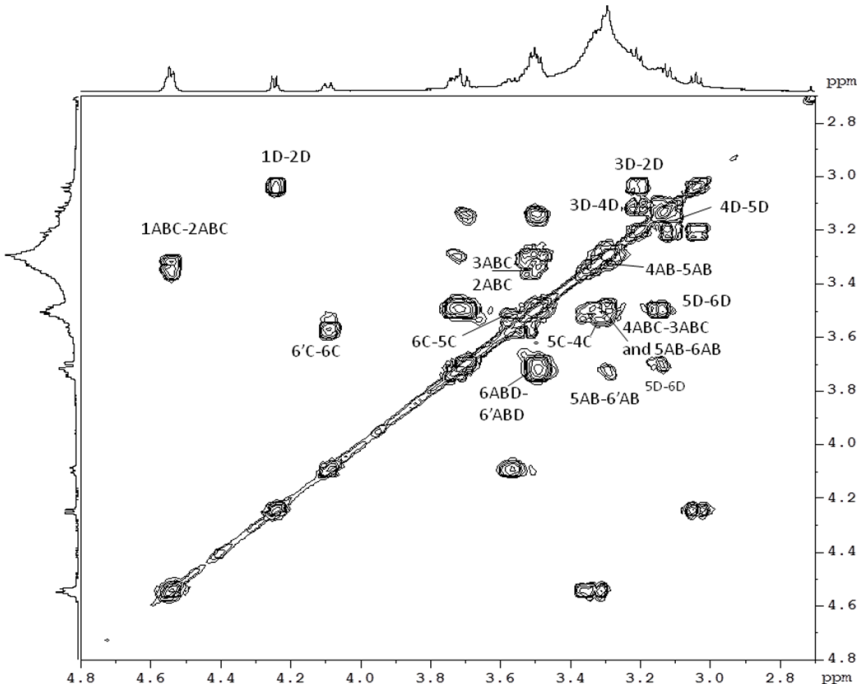


Figure 2.6: ^{1}H-^{1}H COSY NMR map of Scl in DMSO-d_6.

Similarly, in the TOCSY map (figure **2.7**) the A and B units are totally equivalent, so their cross peaks are coincident. This is true also for the first four nuclei of the C unit, but the signals of the proton in the fifth and sixth position differ from those of A and B. The D unit presents almost all the cross-peaks distinguished from those of the other units, but the cross-peak between the protons 6 and 6' is in common with the A and B nuclei.

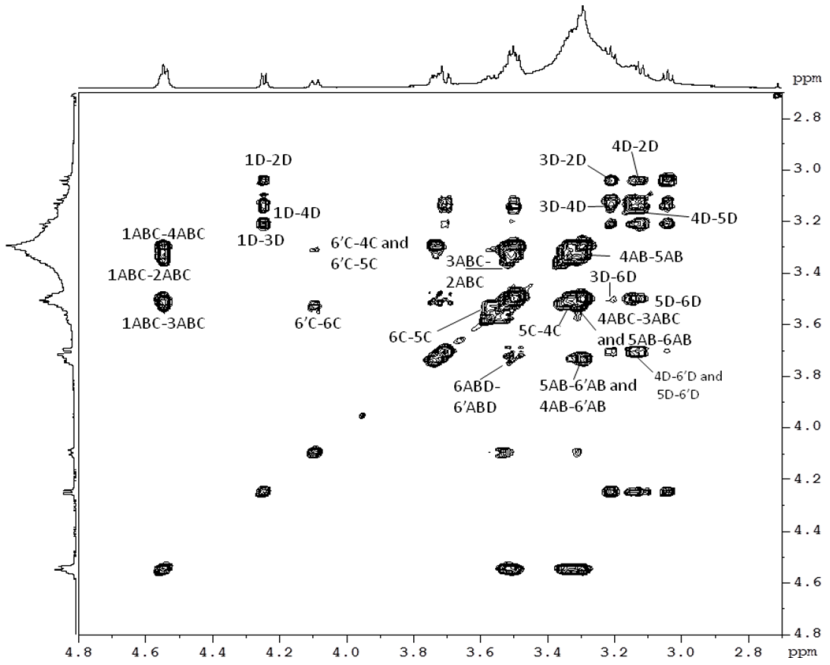


Figure 2.7: ^{1}H-^{1}H TOCSY NMR map of Scl in DMSO-d_6.

The ^{1}H-^{13}C HSQC map allowed the correlation of each proton with the respective carbon atom, confirming the assignment above described. (Figure **2.8**)

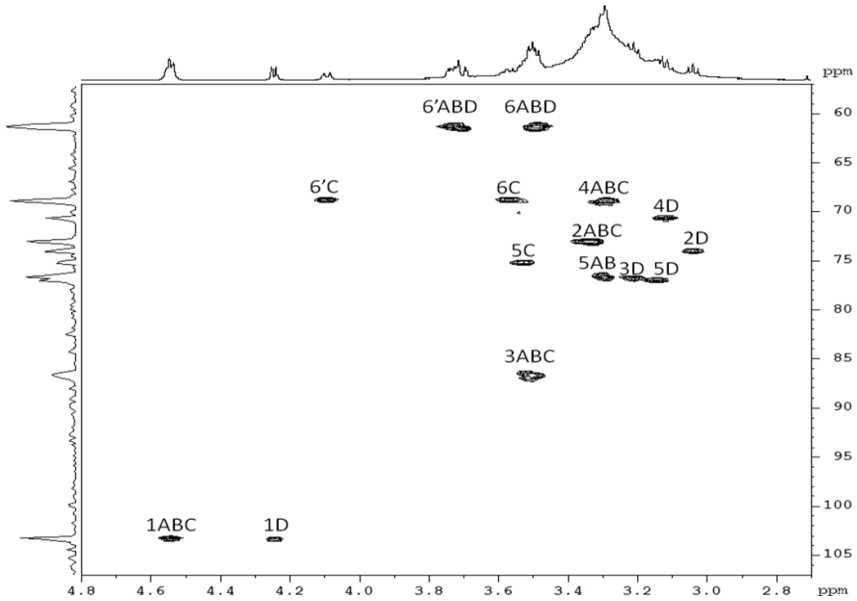


Figure 2.8: ^{1}H-^{13}C HSQC NMR map of Scl in DMSO-d_6.

2.3.3 NMR characterization of Scleroglucan-CM-80

The spectrum of the Scl-CM-80 shows the presence of all the signals observed in spectrum of Scl, because the derivatization doesn't involve all the glucose units of the polymer, so we can still find some unmodified A, B, C or D units. (figure **2.9** and table **2.2**)

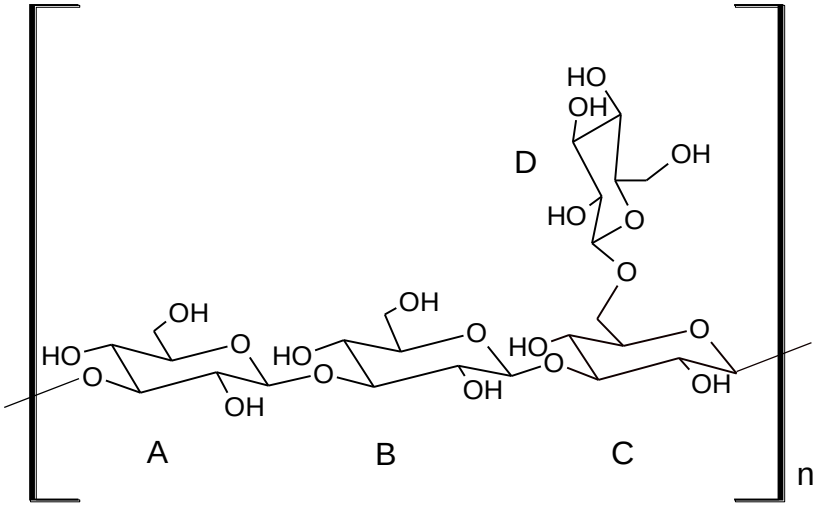


Figure 2.9: Representation of unmodified repetitive unit present in Scl-CM-80.

Table 2.2: ^{1}H and ^{13}C NMR assignment of Scl-CM-80 sample in DMSO at 348 K.

Residue	Type	^{1}H δ (ppm)	Multiplicity (*J* in Hz)	^{13}C δ (ppm)
1A	CH	4.544		103.3
2A	CH	3.334		73.0
3A	CH	3.502		86.7
4A	CH	3.293		68.9
5A	CH	3.293		76.6
6A	CH_2	**6A** 3.496	dd (11.3; 5.5)	61.3
		6'A 3.731	dd (11.3; 5.5)	
1B	CH	4.544		103.3
2B	CH	3.334		73.0
3B	CH	3.502		86.7
4B	CH	3.293		68.9
5B	CH	3.293		76.6
6B	CH_2	**6B** 3.496	dd (11.3; 5.5)	61.3
		6'B 3.731	dd (11.3; 5.5)	
1C	CH	4.544		103.3
2C	CH	3.334		73.0
3C	CH	3.502		86.7
4C	CH	3.293		68.9
5C	CH	3.528		75.2
6C	CH_2	**6C** 3.571	d (10.3; 8.5)	68.8
		6'C 4.094	d (10.3)	
1D	CH	4.246	d (7.9)	103.4
2D	CH	3.041	dd (8.4; 8.3)	74.0
3D	CH	3.212	t (8.76)	76.7
4D	CH	3.121	m	70.7
5D	CH	3.141	m	77.0
6D	CH_2	**6D** 3.497	dd (11.5; 5.5)	61.5
		6'D 3.704	dd (11.5; 1.9)	

For the unmodified units all the considerations performed in the case of Scl structure are valid.

Comparing the two-dimensional spectra of Scl and Scl-CM-80, a set of new signals (underlined by the circlets) whose appearance is due to the derivatization of the different glucose units were also observed. Using the

information obtained by these experiments, it was possible to identify the position of the different units that underwent the functionalization.
In both the COSY and the TOCSY spectra (figure **2.12** and **2.13**) we can observe the presence of the cross-peaks due to the carboxymethyl group, whose two diastereotopic protons resonate respectively at 4.004 and 4.052 ppm. The corresponding carbon atom resonates at 172.2 ppm, as seen in the HMBC spectrum of figure **2.15**.
Combining the information of the COSY and TOCSY spectra it was possible to point out the presence of signals due to the modification of the D ring. In particular we can see in the COSY spectrum the appearance of the cross-peaks at 3.138-3.333 ppm, due to the coupling between the 4D and 5D nuclei, a cross-peak at 3.333-3.615 ppm due to the coupling of 5D and 6D and another cross-peak at 3.615-4.027 ppm corresponding to the coupling of the 6D and 6'D nuclei of the D ring derivatized in correspondence of the position 6. All these proton coupling are confirmed by the TOCSY spectrum and the respective ^{13}C signals, obtained by the HSQC spectrum, confirm this hypothesis of derivatization.
Over the set of the above-mentioned new signals, we can see also another set, that is in accordance with the derivatization of the A and B rings: in fact in the ^{1}H-^{1}H 2D-NMR spectra we see the cross-peaks due to the coupling of the 4AB nucleus with the 5AB (TOCSY), those of the coupling of 5AB with 6AB and those concerning the coupling of 6AB and 6'AB of the AB ring derivatized again in correspondence of the position 6. Also in this case the HSQC spectrum confirms the presence of the derivatization on the position 6 of these glucose units, identifying the carbon corresponding to each new protonic signal. (Figure **2.14**)
Hereafter we report the assignment of the nuclei of the derivatized glucose units (table **2.3** and **2.4**) and of the carboxy-methylic group (table **2.5**). We also report the 2D spectra of the Scl-CM-80, highlighting in the circlets the cross-peaks assigned to the nuclei of the derivatized units.

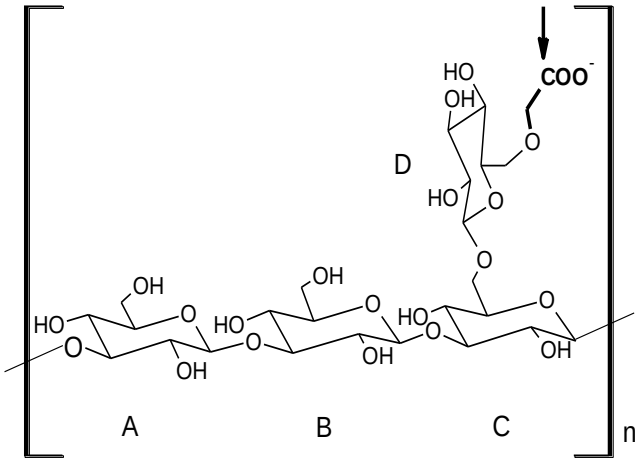


Figure 2.10: Representation of C_6 derivatized D unit present in Scl-CM-80.

Table 2.3: ^{1}H and ^{13}C assignment of Scl-CM-80 (derivatized D unit).

Residue	Type	^{1}H δ (ppm)	Multiplicity (*J* in Hz)	^{13}C δ (ppm)
1D	CH	4.246	d (7.9)	103.4
2D	CH	3.041	dd (8.4; 8.3)	74.0
3D	CH	3.212	t (8.76)	76.7
4D	CH	3.138		70.6
5D	CH	3.333		75.7
6D*	CH_2	**6D*** 3.615		68.8
		6D'* 4.027		

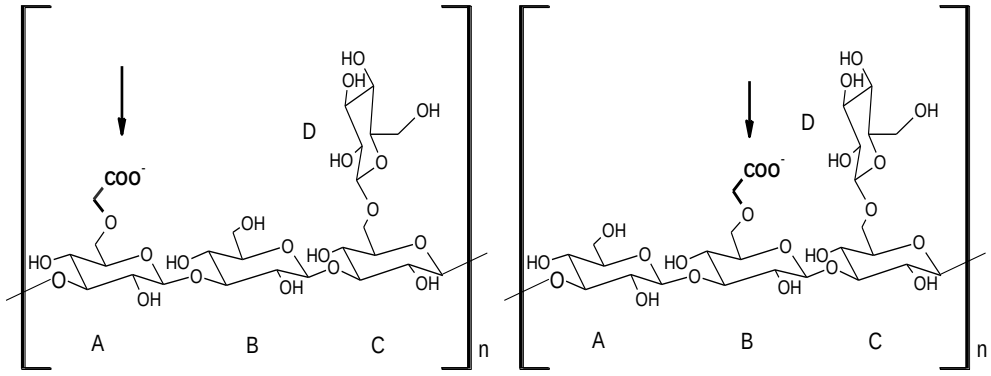


Figure 2.11: Representation of C_6 derivatized A and B units present in Scl-CM-80.

Table 2.4: ^{1}H and ^{13}C assignment of Scl-CM-80 (derivatized AB units).

Residue	Type	^{1}H δ (ppm)	^{13}C δ (ppm)
1AB	CH	4.544	103.3
2AB	CH	3.334	73.0
3AB	CH	3.502	86.7
4AB	CH	3.294	68.9
5AB	CH	3.455	75.2
6AB	CH_2	**6AB*** 3.591	70.7
		6AB'* 3.815	

Table 2.5: ^{1}H and ^{13}C assignment of carboxy-methyl group of Scl-CM-80.

Residue	Type	^{1}H δ (ppm)	^{13}C δ (ppm)
-OCH_2CO_2H	CH_2	4.004 4.052	68.7
	CO_2H	-	172.2

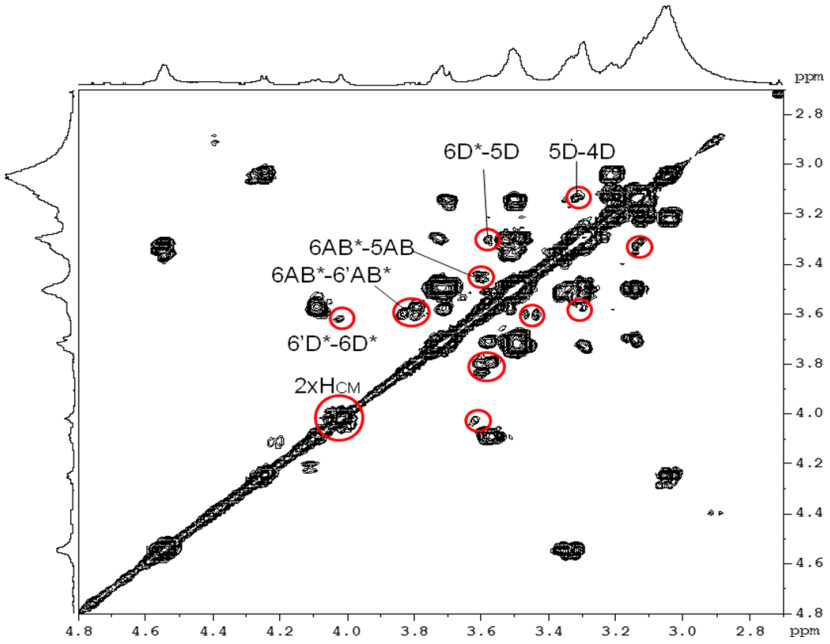


Figure 2.12: ^{1}H-^{1}H COSY map of Scl-CM-80 sample in DMSO-d_6.

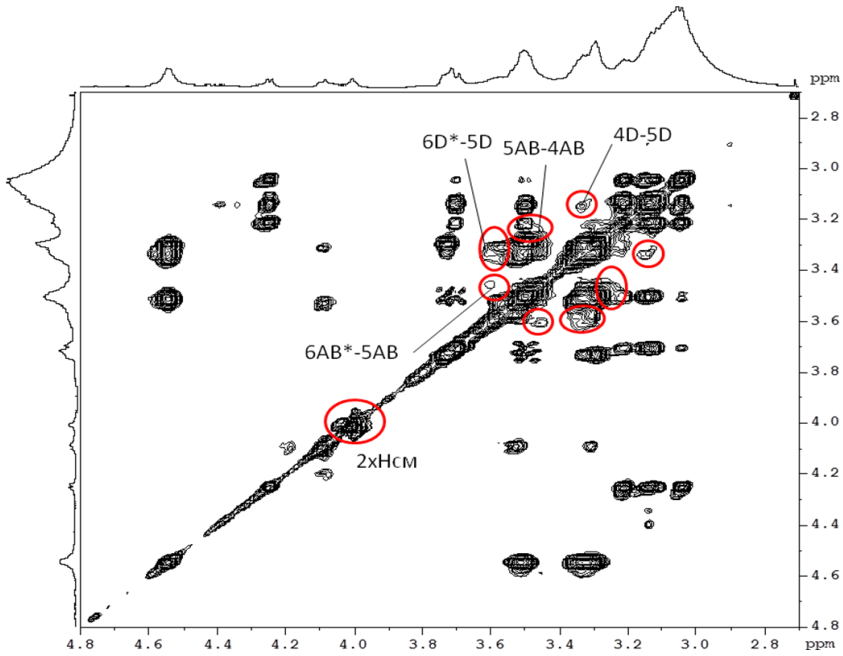


Figure 2.13: ^{1}H-^{1}H TOCSY map of Scl-CM-80 sample in DMSO-d_6.

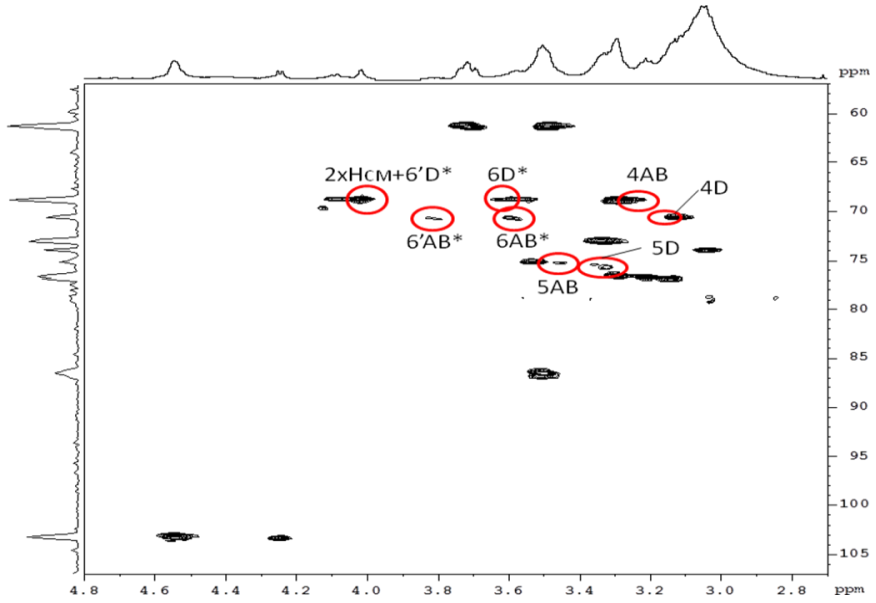


Figure 2.14: ^{1}H-^{13}C HSQC map of Scl-CM-80 in DMSO-d_6.

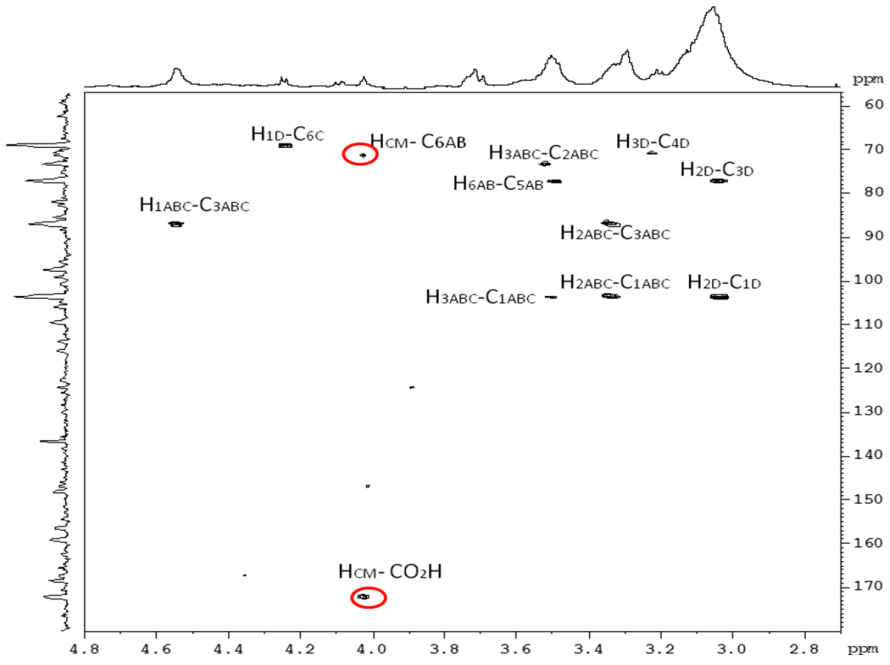


Figure 2.15: ^{1}H-^{13}C HMBC map of Scl-CM-80 in DMSO-d_6.

2.4 Conclusions

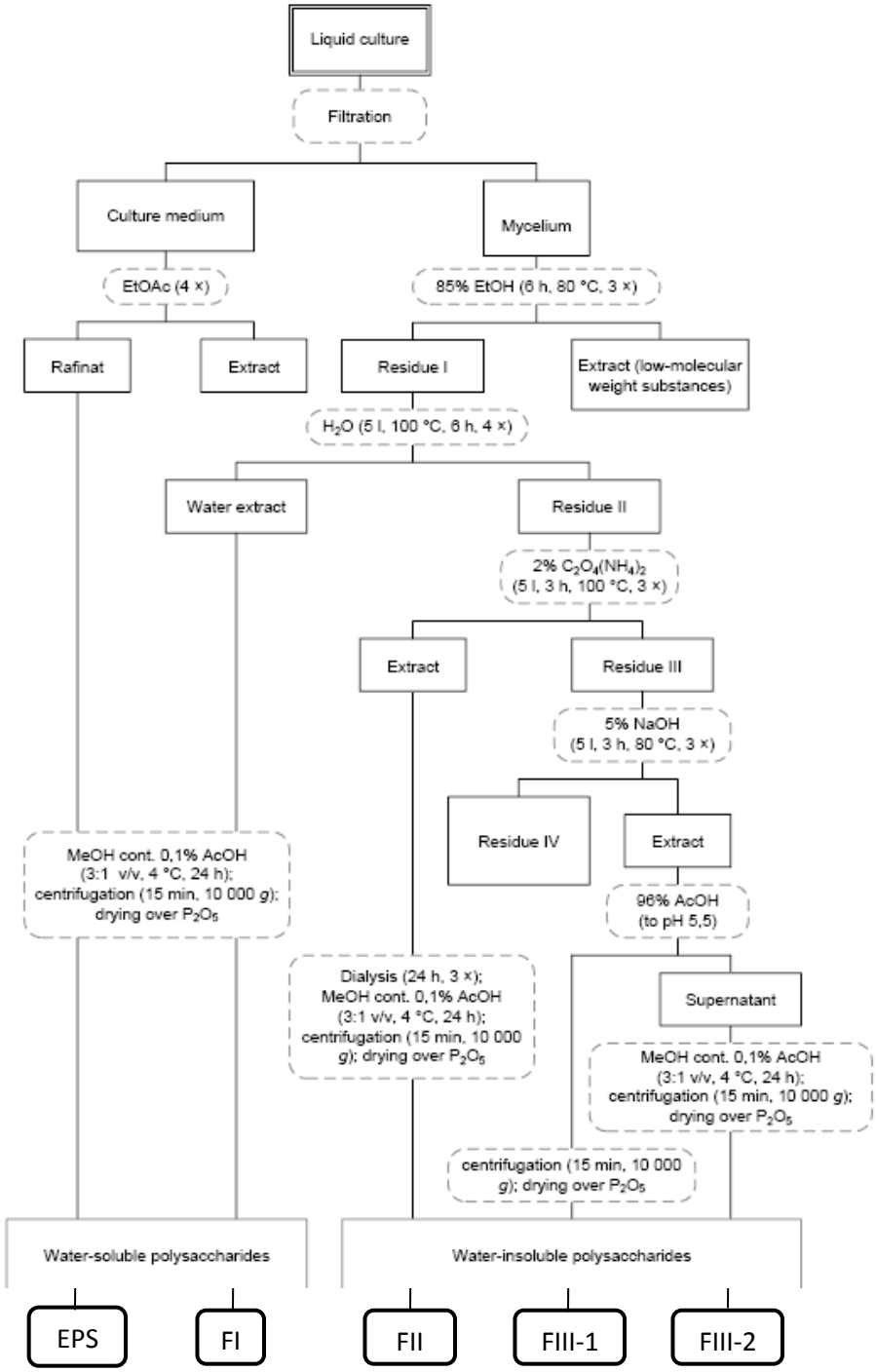
The NMR characterization of the Scl and its derivative Scl-CM-80 was carried out with the aim to identify the linkage positions of the different glucose units.
The total assignment of the spectrum of scleroglucan was performed. The spectrum of Scl-CM-80 showed the same peaks found in the scleroglucan, indicating the presence of a certain amount of unmodified glucose molecules, and also new signals. These signals visible only in the Scl-CM-80 spectra confirm the hypothesis of the D unit derivatization, the most probable derivatization due to the lower steric hindrance of the branch position. The ^{1}H spectra suggested also the existence of some A and B derivatized units.
Carbon 6 of the glucose ring is the carbon involved in the linkage with carboxymethyl group: it is in fact the most accessible one.
Due to the overlapping between the signals corresponding to the proton 6D'* and the protons of the carboxymethyl group, it was not possible to obtain the degree of derivatization by the integration of carboxymethyl signals. This information will be obtained using the potentiometric method.

3. Structural investigation of polysaccharides from *Hericium erinaceum*

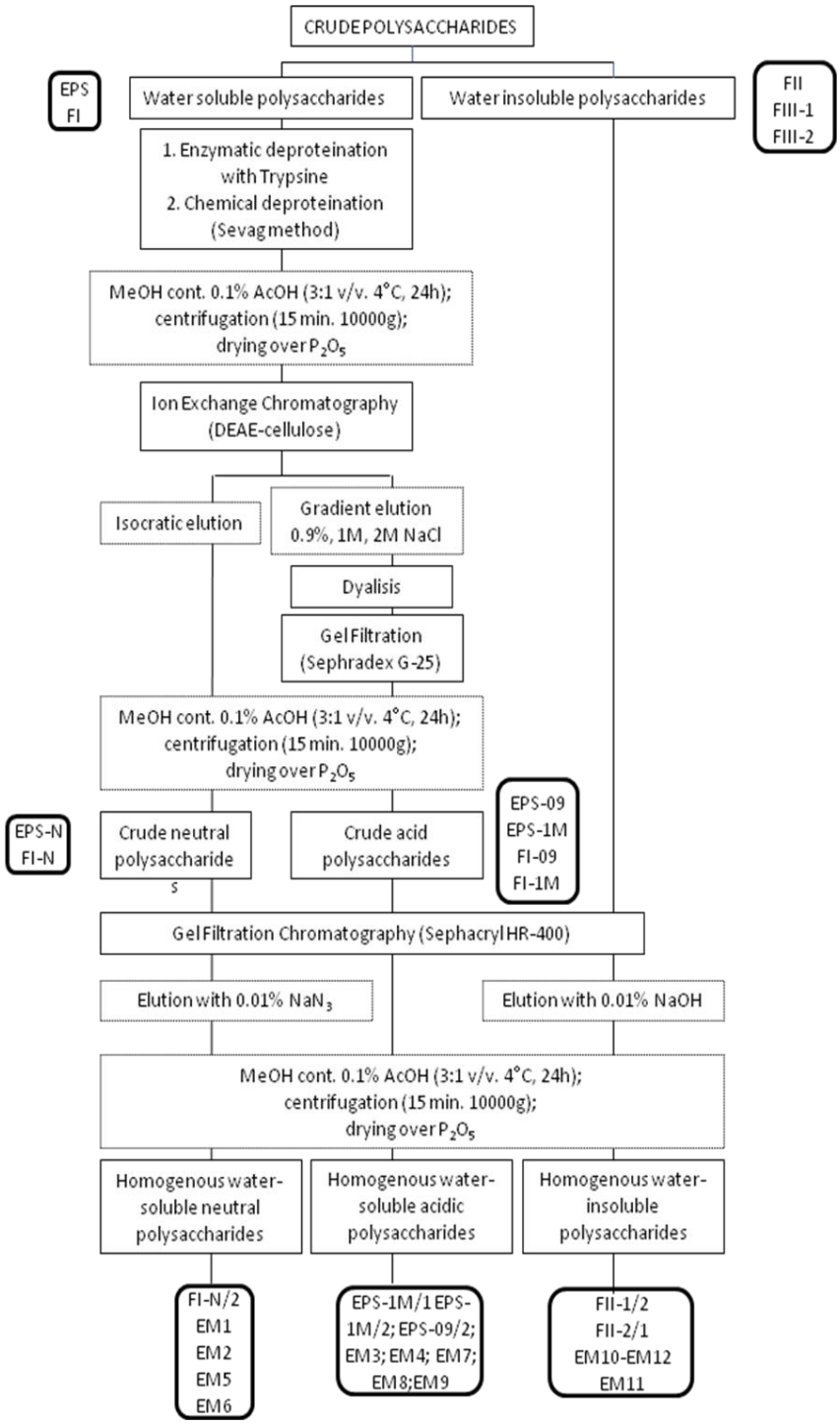
3.1 Introduction[60]

Recent studies showed that polysaccharides isolated from the mycelium of the edible and medicinal mushroom Hericium erinaceum,[61,62] exhibit immunostimulating and anticancer activity like observed for polysaccharides from other edible mushrooms (e.g. lentinan from Lentinus edodes or grifolan from Grifola frondosa).[63-70] Another mode of action of H. erinaceum polysaccharides is cholesterol-lowering and neurite outgrowth stimulating activity.[71,72]

This work has been performed in collaboration with the laboratory of Prof. Harold, Faculty of Pharmacy, The Medical University of Warsaw. In order to investigate the possible structure of some active polysaccharides produced by this mushroom, an extraction and purification procedure (scheme **3.1** and **3.2**) has been performed in the laboratory of Prof. Harold obtaining different compounds (EM1-12). These compounds showed a variable immunostimulatory activity. In particular compound EM10 showed a significant induction of mRNA of IL1, IL6, IL12 and TNF-α expression in raw 264.7 cells.



Scheme 3.1: procedure to obtain the analyzed polysaccharides (part 1)



Scheme 3.2: procedure to obtain the analyzed polysaccharides (part 2)

In table **3.1** we report the monosaccharide composition of the polysaccharides samples, determined by GC-MS analysis (according to the Jones and Albersheim method[73] and by HPLC analysis (values out of parentheses, according to the Lv method).[74]
Molecular weight and homogeneity were determined by densitometric method. The molecular weight of the obtained novel macromolecule was read from the calibration curve prepared with standards of dextrans.[60]
Among the obtained polysaccharides, the compounds EM1-EM9 were water soluble, while EM10-EM12 were soluble in diluted alkali, DMSO or in hot water. EM12 was obtained from a mycelium cultivated in presence of $Selol_{5\%}$.[75]

Table 3.1. The monosaccharides composition of *H. erinaceum* polysaccharides (%mol). Numbers out of parentheses are obtained from the HPLC analysis (performed according to the method of Lv eta al.).

	MW (kDa)	Carbohyd. (%)	Prot. (%)	Ara	Fuc	Gal	GalA	Glc	GlcA	GlcN	Man	Rha	Xyl
EM1	221	70.6	29.4	–	1,0	22,0	–	13,6	–	6,4	36,5	2,8	17,7
				(–)	(1,6)	(14,0)	(–)	(18,0)	(–)	(–)	(54,6)	(2,0)	(9,8)
EM2	62	78.7	21.3	3,7	4,3	16,8	–	13,9	–	–	44,6	3,9	12,9
				(1,1)	(6,3)	(12,3)	(2,8)	(12,6)	(–)	(–)	(53,2)	(2,3)	(9,2)
EM3	30	100.00	0.0	5,9	13,2	11,4	21,5	4,8	5,2	–	16,4	7,8	13,7
				(4,1)	(15,2)	(10,6)	(24,8)	(4,5)	(6,8)	(–)	(14,7)	(5,7)	(13,7)
EM4	151	71.6	28.4	8,1	8,7	16,9	–	17,4	–	–	21,0	15,4	12,6
				(1,1)	(6,6)	(7,1)	(–)	(19,1)	(–)	(–)	(57,6)	(4,9)	(3,5)
EM5	1980	100.0	0.0	0,7	–	1,1	–	94,2	–	–	4,0	–	–
				(–)	(–)	(–)	(–)	(96,4)	(–)	(–)	(3,6)	(–)	(–)
EM6	22	100.0	0.0	–	0,6	–	–	51,5	–	–	3,0	–	44,9
				–	–	4,3	–	87,7	1,2	–	4,5	–	2,5
EM7	2000	100.0	0.0	(–)	(–)	(1,8)	(–)	(92,5)	(0,7)	(–)	(5,0)	(–)	(–)
				–	–	2,0	–	88,9	5,8	–	3,3	–	–
EM8	30	100.0	0.0	(–)	(–)	(1,5)	(–)	(89,5)	(2,7)	(–)	(6,3)	(–)	(–)
				0,9	–	1,3	–	72,5	11,7	–	11,1	2,1	0,5
EM9	34.3	93.6	6.4	(–)	(–)	(1,0)	(–)	(71,4)	(13,5)	(–)	(10,5)	(3,6)	(–)
				–	–	–	–	66,1	12,9	–	13,0	–	7,9
EM10	37.5	100.0	0.0	(–)	(–)	(–)	(–)	(63,8)	(14,1)	(–)	(12,0)	(–)	(10,1)
				0,2	–	–	–	93,9	3,8	–	1,1	–	1,0
EM11	119	85.3	14.7	(–)	(–)	(–)	(–)	(92,6)	(4,3)	(–)	(3,1)	(–)	(–)

To obtain information about the linkages and the connectivity of the monosaccharide units inside each polysaccharide, various 1D and 2D-NMR experiments have been performed.
Hereafter we show the results of our preliminary studies, aimed at the structural elucidation of these samples. Unfortunately, due to the complex composition of some of them, it has not been possible to obtain complete information for all the samples. Further characterization studies are still ongoing.

3.2 Materials and methods

3.2.1 Preparation of the samples

The NMR spectra were recorded on a Bruker AVANCE AQS600, operating at the protonic frequency of 600.13 MHz and at ^{13}C frequency of 150.92 MHz. The instrument was equipped with a 5 mm inverse probe BBI, properly tuned on the characteristic protonic frequency in the applied field B_0 of 14.3 T.
The samples were dissolved in D_2O, and totally dried over P_2O_5 in a desiccators and the procedure has been repeated two times.
The obtained water soluble polymers EM1-9 were dissolved in phosphate buffer in D_2O containing TSP as internal standard (2 mM), with a final polymer concentration 4mg/mL, while the less water-soluble samples EM10-12 were dissolved in DMSO-d_6 (5mg/ml)
The ^{1}H and ^{13}C assignments were obtained with 2D ^{1}H-^{1}H COSY (Correlation SpectroscopY), ^{1}H-^{1}H TOCSY (Total Correlation SpectroscopY), and ^{1}H-^{13}C HSQC (Heteronuclear Single Quantum Coherences), all performed at the temperature of 353 K (75°C).

3.2.2 NMR Spectra acquisition

Water soluble samples (EM1-9)
The experiments were performed using a sequence that allowed the solvent presaturation.

^{1}H-NMR:
Time Domain (TD): 32K;
Sweep Width (SW): 10-12 ppm;
O1: about 2830 Hz (optimized for each spectrum);
Receiver Gain: 575-2048;
90 degrees pulse;
Delay (D1): 3.0 sec;
Number of Scans (NS): 256;

Dummy Scans (DS): 4;
Temperature: 353 K.

^{1}H-^{1}H-COSY:
Pulse program: cosyprqf;
Time Domain (TD): F_2=1024 F_1=512;
Sweep Width (SW): 10-12 ppm;
O1: about 2830 Hz (optimized for each spectrum);
Receiver Gain: 256;
90 degrees pulse;
Delay (D1): 2.5 sec;
Acquisition Time (AQ): F_2=0.085 sec F_1= 0.043 sec;
Number of Scans (NS): 48;
Dummy Scans (DS): 16;
Temperature: 353 K.

^{1}H-^{1}H-TOCSY:
Pulse program: mlevphpr;
Time Domain (TD): F_2=1024, F_1=512;
Sweep Width (SW): 10-12 ppm;
O1: about 2830 Hz (optimized for each spectrum);
Receiver Gain: 2048;
90 degrees pulse;
Mixing time: 90 ms;
Delay (D1): 2.5 sec;
Acquisition Time (AQ): F_1=0.071 sec; F_2=0.035 sec;
Number of Scans (NS): 32-48;
Dummy Scans (DS): 16;
Temperature: 353 K.

^{1}H-^{13}C HSQC:
Pulse program: hsqcphpr
Time Domain (TD): F_2=1024, F_1=512;
Sweep Width (SW): 10-12 ppm; 130-140 ppm
O1: about 2830 Hz (optimized for each spectrum); O2: 9054.17 Hz
Receiver Gain: 1024-2896;
90 degrees pulse;
$^{1}J_{C\text{-}H}$ coupling constant: 140 Hz

Delay (D1): 2.5 sec;
Acquisition Time (AQ): F_2=0.085 sec; F_1=0.013 sec;
Number of Scans (NS): 88-100;
Dummy Scans (DS): 16;
Temperature: 353 K.

Water insoluble samples (EM10-12)

^{1}H-NMR:
Time Domain (TD): 32K;
sweep width (SW): 11.0 ppm;
O1: 2700.6 Hz;
Receiver Gain: 812.7;
90 degrees pulse;
Delay (D1): 3.0 sec;
Number of Scans (NS): 256;
Dummy Scans (DS): 4;
Temperature: 353 K.

^{1}H-^{1}H-COSY:
Pulse program: cosyqf45;
Time Domain (TD): F_2=1024 F_1=512;
sweep width (SW): 11.0 ppm;
O1: 2700.6 Hz;
Receiver Gain: 812.7;
Delay (D1): 2.5 sec;
Acquisition Time (AQ): F_2: 0.077 sec; F_1: 0.039 sec;
Number of Scans (NS): 36;
Dummy Scans (DS): 16;
Temperature: 353 K.

^{1}H-^{1}H-TOCSY:
Pulse program: mlevph;
Time Domain (TD): F_2=1024, F_1=512;
sweep width (SW): 11.02 ppm;
O1: 2700.6 Hz;
Receiver Gain: 812.7;
90 degrees pulse;

Mixing time: 90 ms;
Delay (D1): 2.5 sec;
Acquisition Time (AQ): F_2=0.077 sec; F_1=0.039 sec;
Number of Scans (NS): 40;
Dummy Scans (DS): 16;
Temperature: 353 K.

^{1}H-^{13}C HSQC:
Pulse program: hsqcph;
Time Domain (TD): F_2=1024, F_1=512;
sweep width (SW): 11.02 ppm -140.00 ppm;
O1: 2700.6 Hz; O2: 9054.2 Hz
Receiver Gain: 574.7;
90 degrees pulse;
CNST2: 140 Hz;
Delay (D1): 2,5 sec;
Acquisition Time (AQ): F_2=0.077 sec, F_1=0.012 sec;
Number of Scans (NS): 84;
Dummy Scans (DS): 16;
Temperature: 353 K.

3.2.3 NMR Spectra processing

The ^{1}H spectra obtained after the Fourier transformation with a line broadening of 0.30 Hz, were processed performing a baseline correction and manual phase correction.
The chemical shifts are given in parts per million, using the internal DMSO residual signals as reference: ^{1}H = 2.500 ppm and ^{13}C = 40.40 ppm.

3.3 Results and discussion

NMR characterization of the polysaccharides

Hereafter we report the 1D and 2D NMR spectra performed on the different samples, followed by the structural information obtained from their interpretation.
All the samples were dissolved in D_2O twice and dried over P_2O_5. All the experiments were performed at high temperature (353 K) to increase the resolution of the spectra.
The assignments of ^{1}H and ^{13}C resonances were obtained from ^{1}H-^{1}H and ^{1}H-^{13}C two-dimensional correlation NMR experiments.
Due to high complexity of the ^{1}H spectra, for some samples (EM1, 3 and 4) it has not been possible to obtain an unambiguous information. Hereafter, the assignment of the spectra of the other samples are reported.

Sample EM8: 0,7 mL of 4mg/mL solution in P-buffer containing D_2O, with 2 mM TSP as internal standard, underwent the following experiments:

- ^{1}H-NMR T=353 K (fig. **3.1**)
- ^{1}H-^{1}H TOCSY T=353 K (fig. **3.2**)
- ^{1}H-^{13}C HSQC T=353 K (fig. **3.3**)

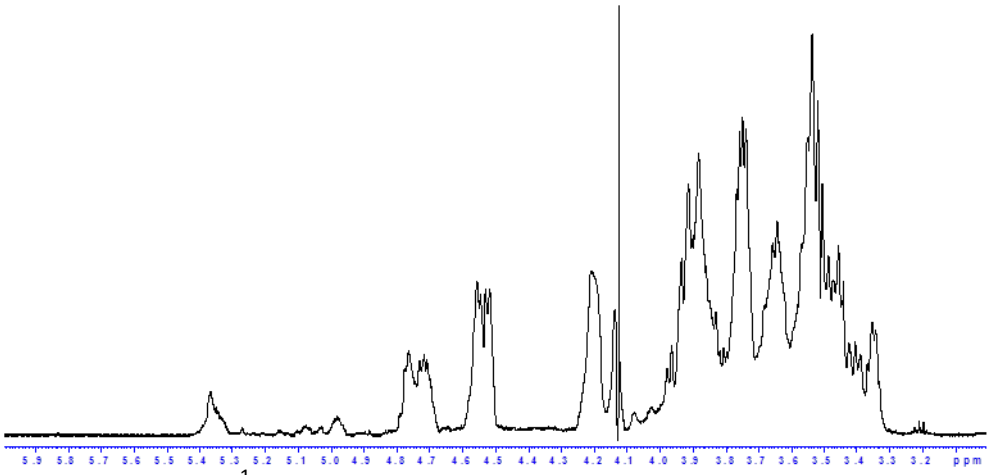


Figure 3.1: ^{1}H-NMR spectrum of sample EM8 4 mg/mL (353 K)

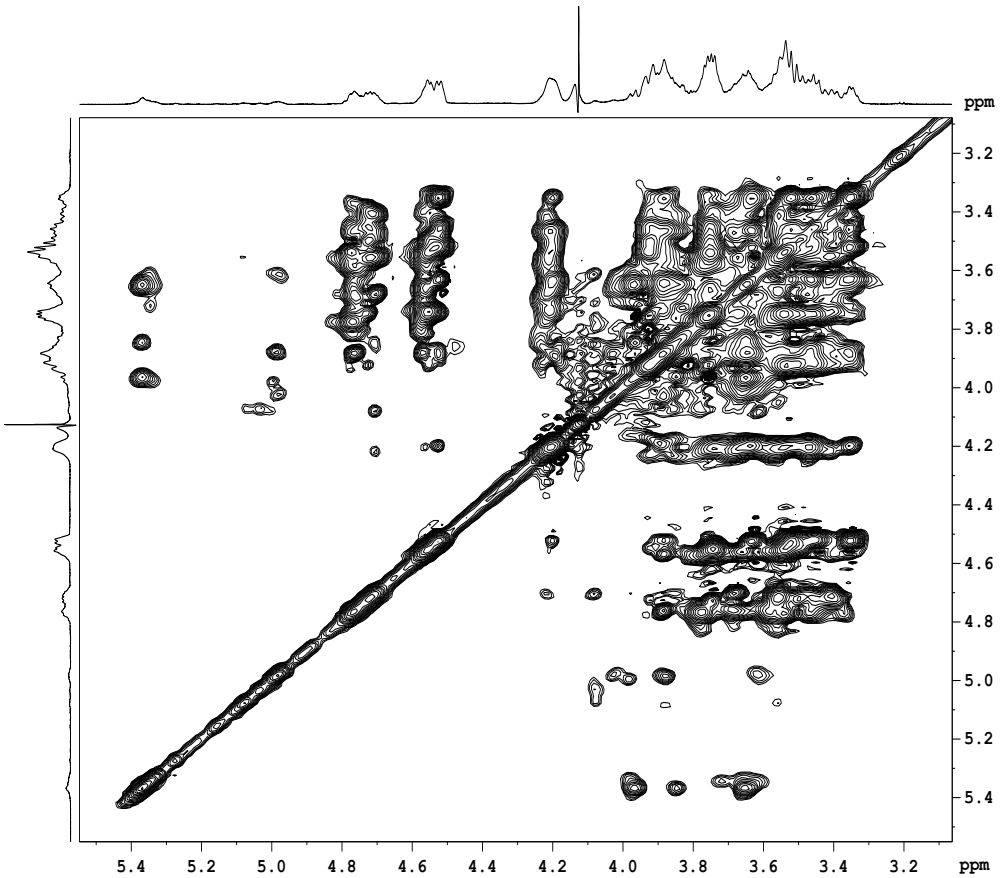


Figure 3.2: ^{1}H-^{1}H-TOCSY spectrum of sample EM8 4 mg/mL (T=353 K)

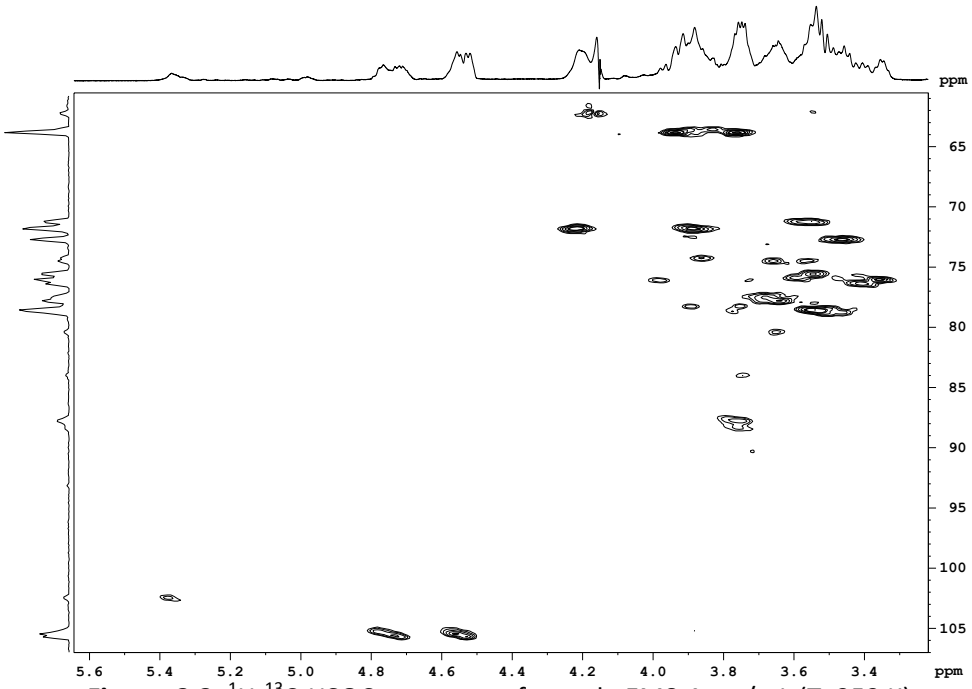


Figure 3.3: ^{1}H-^{13}C-HSQC spectrum of sample EM8 4 mg/mL (T=353 K)

This sample is constituted mostly (89%) by glucose, mainly in the β anomeric form, with a β(1→3) linkage (giving the signals of the anomeric protons in the region 4.68-4.80 ppm) and a β(1→6 or 4) linkage (giving the signals of the anomeric protons in the region 4.50-4.60 ppm).
In the sample a certain amount of α-glucose, even if in lower quantity, is present. It is involved in an α(1→4) linkage as showed by the signals of the anomeric protons in the region 5.31-5.40 ppm, and in an α(1→6) linkage, as showed by those in the region 4.95-5.01 ppm.
In table **3.2** we report the assignment of the monosaccharide units identified by the analysis of the correlations observed in the spectra.[76-78]

Table 3.2: NMR ^{1}H and ^{13}C chemical shift assignments of EM8

→6) βGlcp(1→6 o 4)			
Position	Type	^{1}H δ (ppm)	^{13}C δ (ppm)
1	CH	4.52	105.5
2	CH	3.35	76.0
3	CH	3.52	78.6
4	CH	3.46	72.7

5	CH	3.64	77.7
6 / 6’	CH_2	4.20-3.88	71.8
→3)(6)βGlcp(1→6 o 4)			
Position	Type	^{1}H δ (ppm)	^{13}C δ (ppm)
1	CH	4.57	105.4
2	CH	3.53	75.5
3	CH	3.74	87.7
4	CH	3.56	71.2
5	CH	3.66	77.7
6 / 6’	CH_2	4.21-3.88	71.7
βGlcp(1→3)			
Position	Type	^{1}H δ (ppm)	^{13}C δ (ppm)
1	CH	4.72	105.6
2	CH	3.40	76.0
3	CH	3.55	78.5
4	CH	3.45	72.6
5	CH	3.49	78.4
6 / 6’	CH_2	3.92-3.74	63.8
→3)(6)βGlcp(1→3)			
Position	Type	^{1}H δ (ppm)	^{13}C δ (ppm)
1	CH	4.76	105.3
2	CH	3.54	75.3
3	CH	3.78	87.7
4	CH	3.59	71.2
5	CH	3.68	77.5
6 / 6’	CH_2	4.21-3.88	71.7
→4)αGlcp(1→4)			
Position	Type	^{1}H δ (ppm)	^{13}C δ (ppm)
1	CH	5.37	102.6
2	CH	3.65	74.5
3	CH	3.96	76.1
4	CH	3.64	80.5
5	CH	3.84	74.2
6 / 6’	CH_2	3.84/3.98	63.6
αGlcp(1→4)			
Position	Type	^{1}H δ (ppm)	^{13}C δ (ppm)
1	CH	5.34	102.6
2	CH	3.62	74.5
3	CH	3.72	76.0

4	CH	3.44	72.6
5	CH	3.73	75.7
6 / 6’	CH_2	3.85/3.96	63.6
→4)αGlcp(1→6)			
Position	Type	^{1}H δ (ppm)	^{13}C δ (ppm)
1	CH	4.99	-
2	CH	3.62	-
3	CH	4.02	-
4	CH	3.63	-
5	CH	3.88	-
6 / 6’	CH_2	3.85/3.88	-

Sample EM5: 0.7 mL of 4mg/mL solution in P-buffer containing D_2O, with 2 mM TSP as internal standard, underwent the following experiments:

- ^{1}H-NMR T=353 K (fig. **3.4**)
- ^{1}H-^{1}H TOCSY T=353 K (fig. **3.5**)
- ^{1}H-^{13}C HSQC T=353 K (fig. **3.6**)

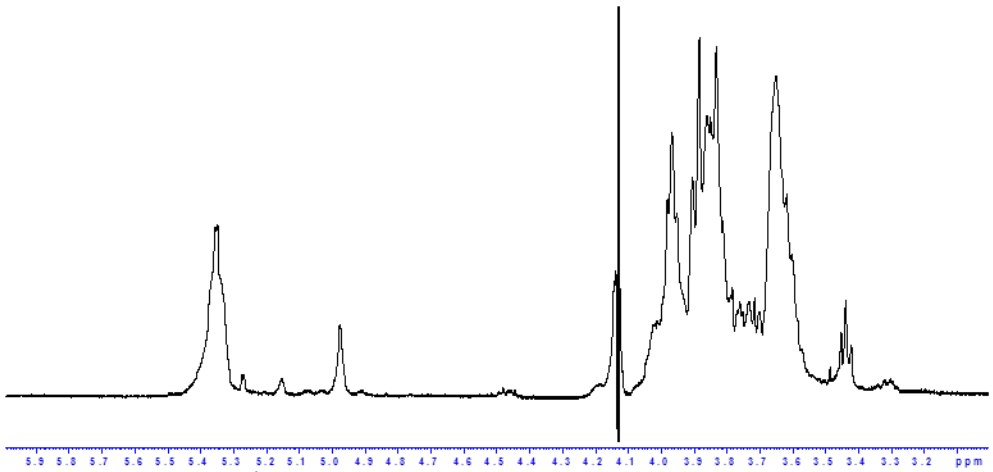


Figure 3.4: ^{1}H-NMR spectrum of sample EM5 4 mg/mL (353 K)

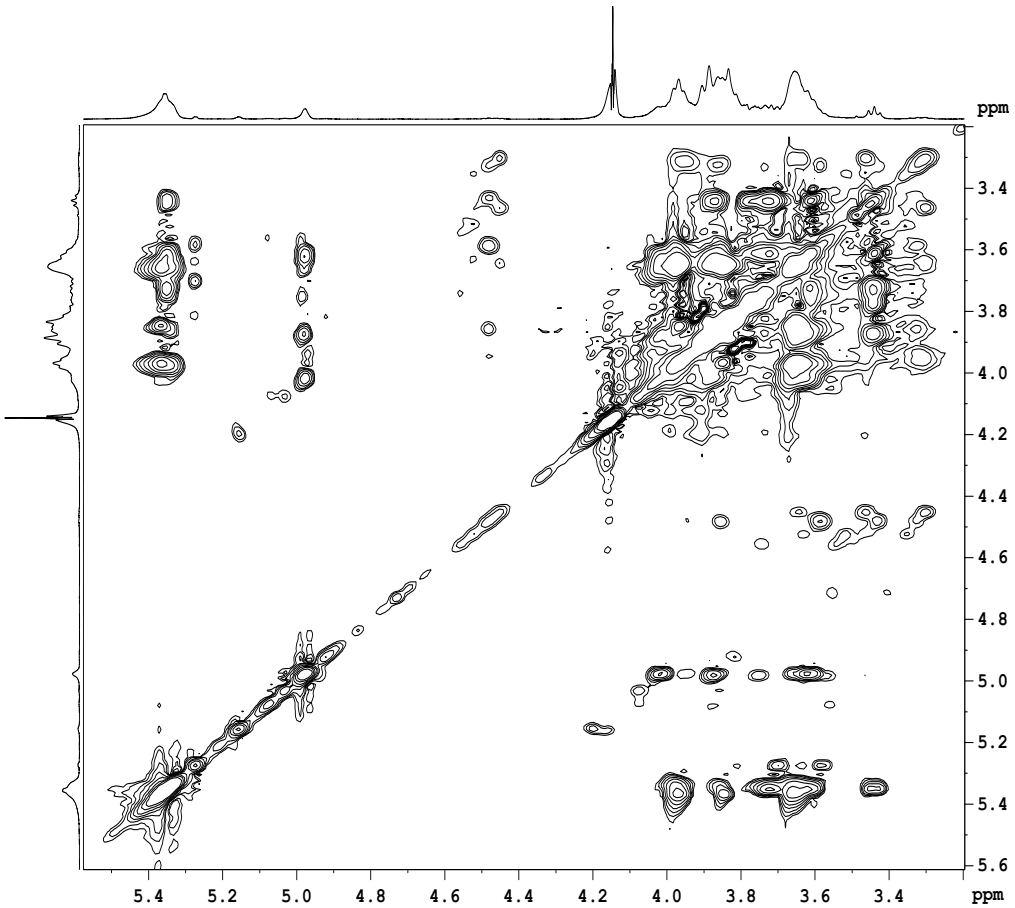


Figure 3.5: ^{1}H-^{1}H-TOCSY spectrum of sample EM5 4 mg/mL (T=353 K)

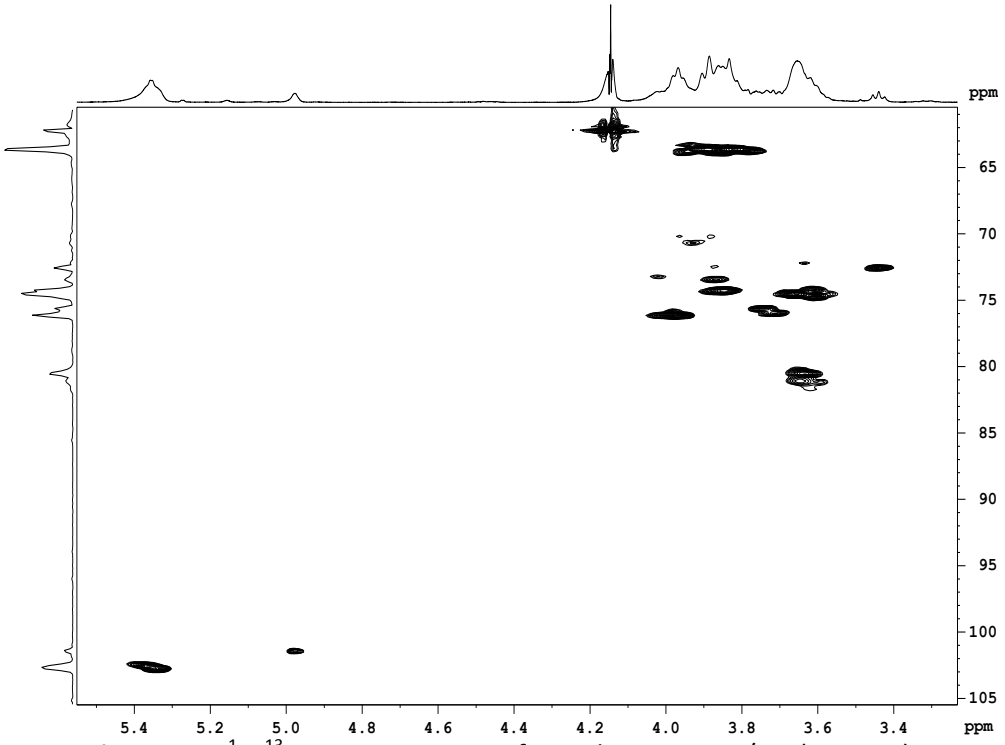


Figure 3.6: ^{1}H-^{13}C-HSQC spectrum of sample EM5 4 mg/mL (T=353 K)

This sample is mainly constituted by α-glucose, involved in an α(1→4) linkage, as shown by the signals of the anomeric protons in the region 5.30-5.48 ppm, and in an α(1→6) linkage, as shown by the signals in the region 4.95-5.01 ppm.

In table **3.3** we report the monosaccharide units, identified by the analysis of the correlations observed in the spectra.

Table 3.3: NMR ^{1}H and ^{13}C chemical shift assignments of EM5

→4)αGlcp(1→4)			
Position	Type	^{1}H δ (ppm)	^{13}C δ (ppm)
1	CH	5.37	102.5
2	CH	3.66	74.5
3	CH	3.96	76.1
4	CH	3.64	80.5
5	CH	3.84	74.3

6 / 6’	CH_2	3.84/3.98	63.6
αGlcp(1→4) terminal			
Position	Type	^{1}H δ (ppm)	^{13}C δ (ppm)
1	CH	5.34	102.7
2	CH	3.62	74.5
3	CH	3.72	76.0
4	CH	3.44	72.6
5	CH	3.75	75.7
6 / 6’	CH_2	3.87/3.82	63.6
→4)αGlcp(1→6)			
Position	Type	^{1}H δ (ppm)	^{13}C δ (ppm)
1	CH	4.97	101.4
2	CH	3.61	74.7
3	CH	4.02	76.1
4	CH	3.63	81.1
5	CH	3.88	73.5
6 / 6’	CH_2	3.84/3.88	62.7
αGlcp(1→6)			
Position	Type	^{1}H δ (ppm)	^{13}C δ (ppm)
1	CH	4.99	101.4
2	CH	3.62	-
3	CH	4.02	-
4	CH	3.63	-
5	CH	3.88	-
6 / 6’	CH_2	3.86/4.02	-

Sample EM6: 0.7 mL of 4mg/mL solution in P-buffer containing D_2O, with 2 mM TSP as internal standard, underwent the following experiments:

- ^{1}H-NMR T=353 K (fig. **3.7**)
- ^{1}H-^{1}H COSY T=353 K (fig. **3.8**)
- ^{1}H-^{1}H TOCSY T=353 K (fig. **3.9**)
- ^{1}H-^{13}C HSQC T=353 K (fig. **3.10**)

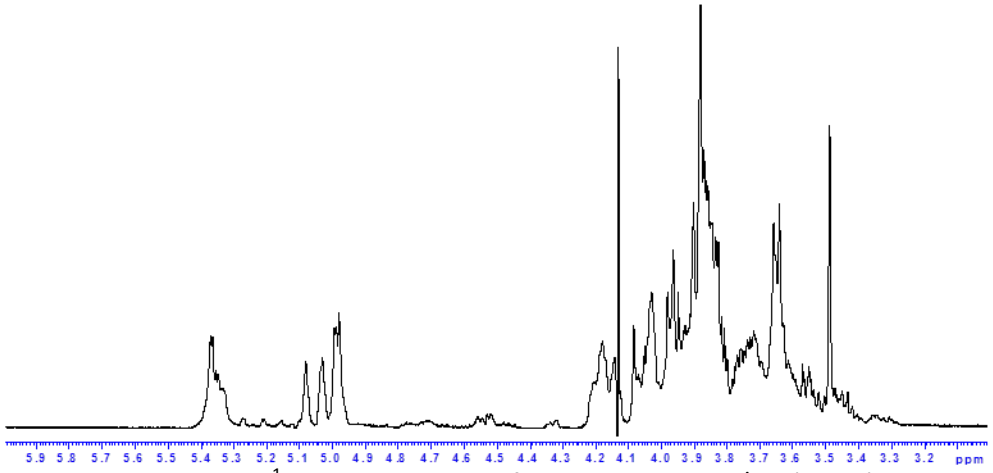


Figure 3.7: ^{1}H-NMR spectrum of sample EM6 4 mg/mL (353 K)

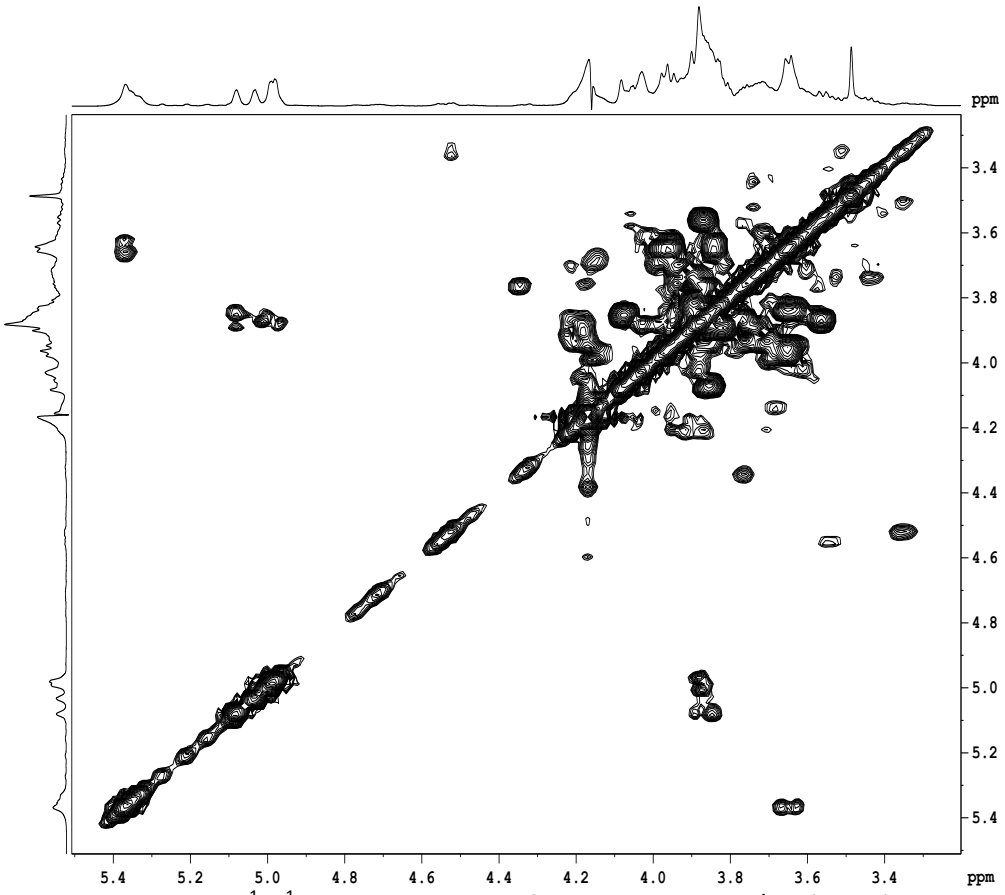


Figure 3.8: ^{1}H-^{1}H-COSY spectrum of sample EM6 4 mg/mL (353 K)

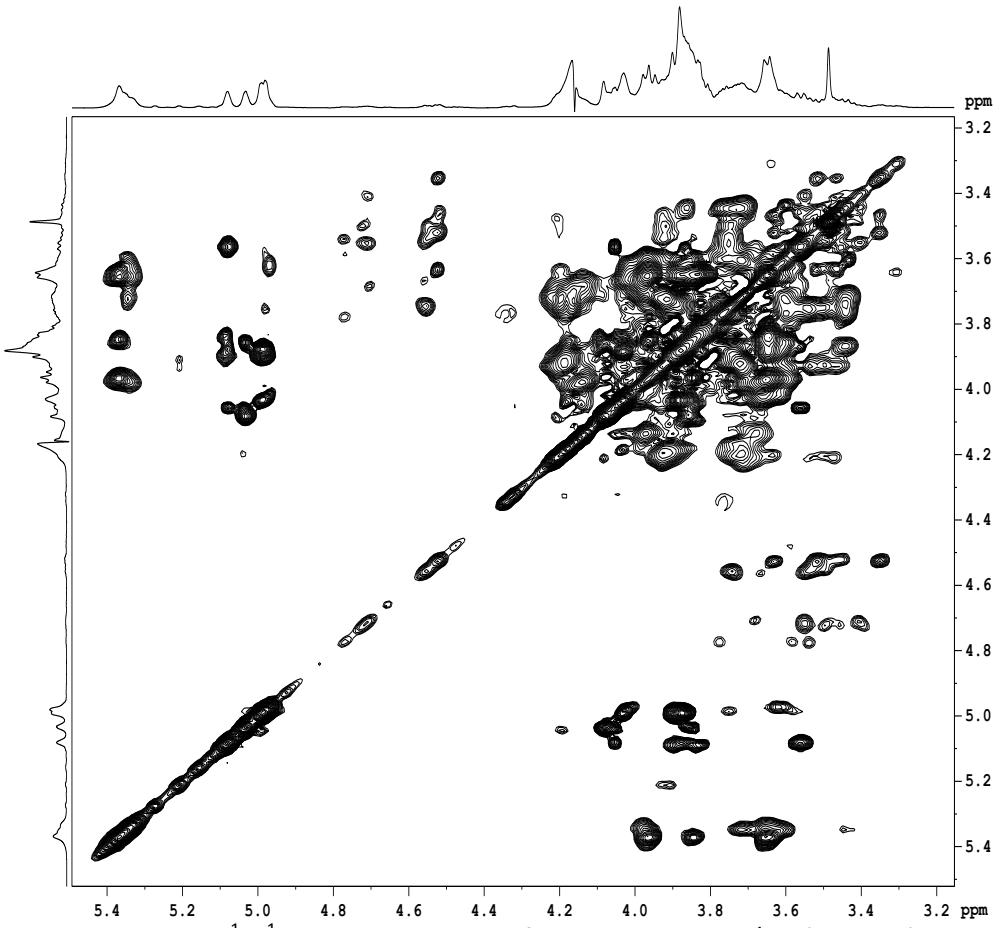


Figure 3.9: ^{1}H-^{1}H-TOCSY spectrum of sample EM6 4 mg/mL (T=353 K)

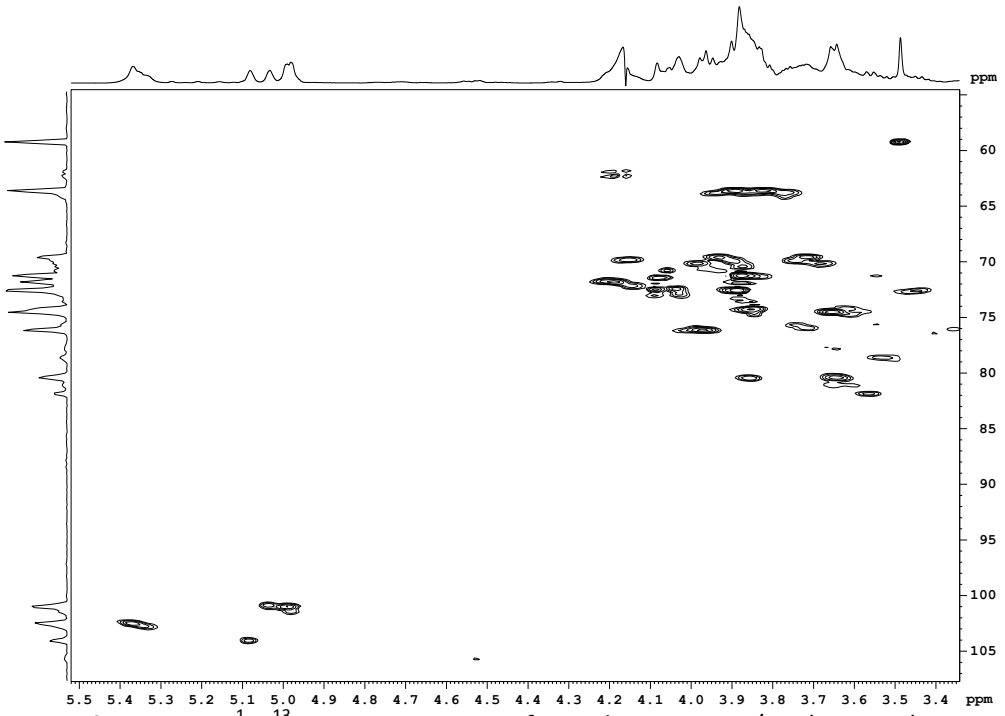


Figure 3.10: ^{1}H-^{13}C-HSQC spectrum of sample EM6 4 mg/mL (T=353 K)

Similarly to the sample EM5, it is mainly constituted by α-glucose, involved in an α(1→4) linkage, as shown by the signals of the anomeric protons in the region 5.30-5.48 ppm, and in an α(1→6) linkage, as shown by those in the region 4.950-5.01 ppm. The monosaccharide units that we observe in this region of the spectra correspond to those reported in table **3.3**.

The sample contains also a low amount of β-glucose with the anomeric proton signals in the spectrum regions 4.68-4.80 ppm and 4.50-4.60 ppm, that presents the same shapes already seen in the sample EM8 and reported in table **3.2.**

The identification of the two signals at 5.08 e 5.03 ppm is still in progress. Given the composition of the sample, it is reasonable to assume that they are due to anomeric proton of xylose. The chemical shift of these signals suggests the α[79,80]

Sample EM7: 0.7 mL of 4mg/mL solution in P-buffer containing D_2O, with 2 mM TSP as internal standard, underwent the following experiments:

- ^{1}H-NMR T=353 K (fig. **3.11**)

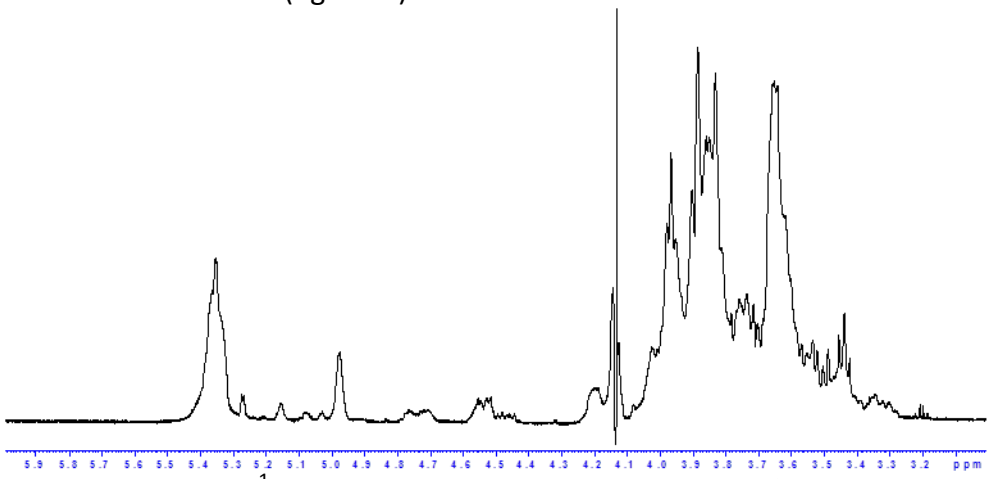


Figure 3.11: ^{1}H-NMR spectrum of sample EM7 4 mg/mL (353 K)

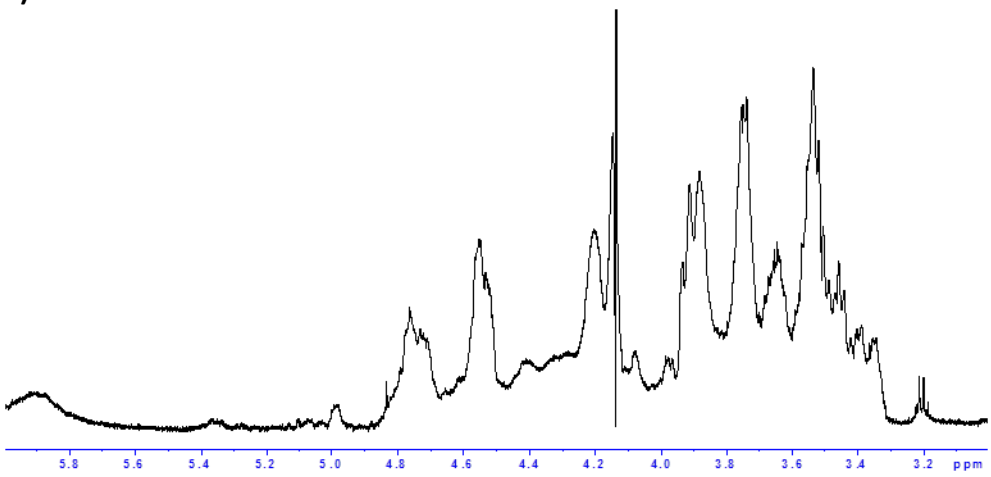
To obtain information on the structure of the sample EM7, it has been sufficient to perform a simple ^{1}H-NMR spectrum. In fact, see figure **3.11**, the spectrum is the result of the perfect superposition of the ^{1}H-NMR spectra of the samples EM5 and EM8.
In this case the principal component of the sample is α–glucose (the anomeric proton signals fall in the spectrum regions 5.30-5.45ppm and 4.95-5.01 ppm). But we can also see the signals corresponding to the β-glucose already seen in the EM8 sample, included in the regions 4.68-4.80 and 4.50-4.60 ppm.
Starting from the integration of the anomeric peaks, we obtained the following relative abundances of the different forms of glucose:
64.% α –Glcp with 1→4 linkage
13% α-Glcp with 1→6 linkage
6.% β-Glcp with 1→3 linkage
9% β-Glcp with 1→6 o 4 linkage

Sample EM9: 0,7 mL of 4mg/mL solution in P-buffer containing D_2O, with 2 mM TSP as internal standard, underwent the following experiments:

- ^{1}H-NMR T=353 K (fig. **3.12**)
- ^{1}H-^{1}H COSY T=353 K (fig. **3.13**)
- ^{1}H-^{1}H TOCSY T=353 K (fig. **3.14**)

- ^{1}H-^{13}C HSQC T=353 K (fig. **3.15**)

a)



b)

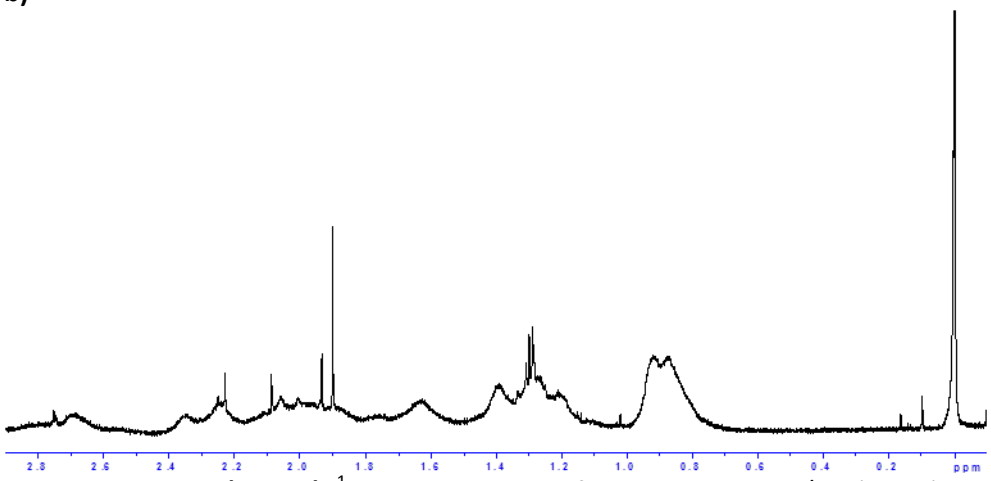


Figure 3.12 a) and b): ^{1}H-NMR spectrum of sample EM9 4 mg/mL (353 K)

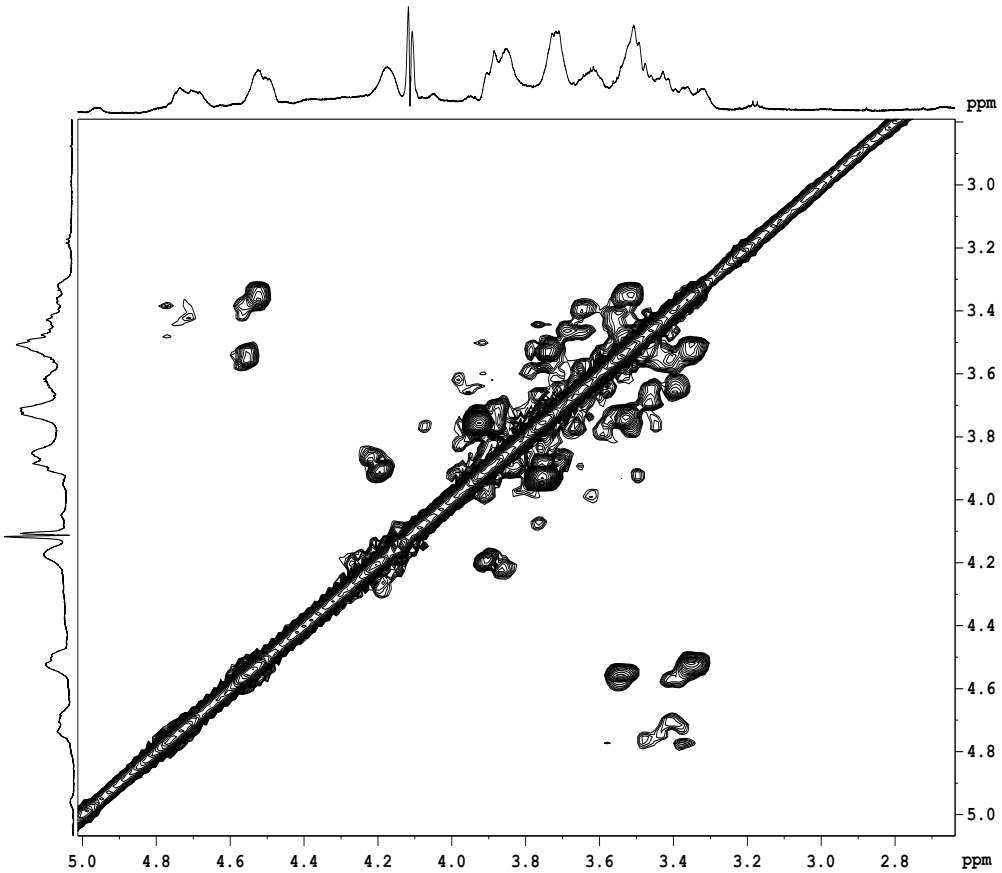


Figure 3.13: ^{1}H-^{1}H-COSY spectrum of sample EM9 4 mg/mL (353 K)

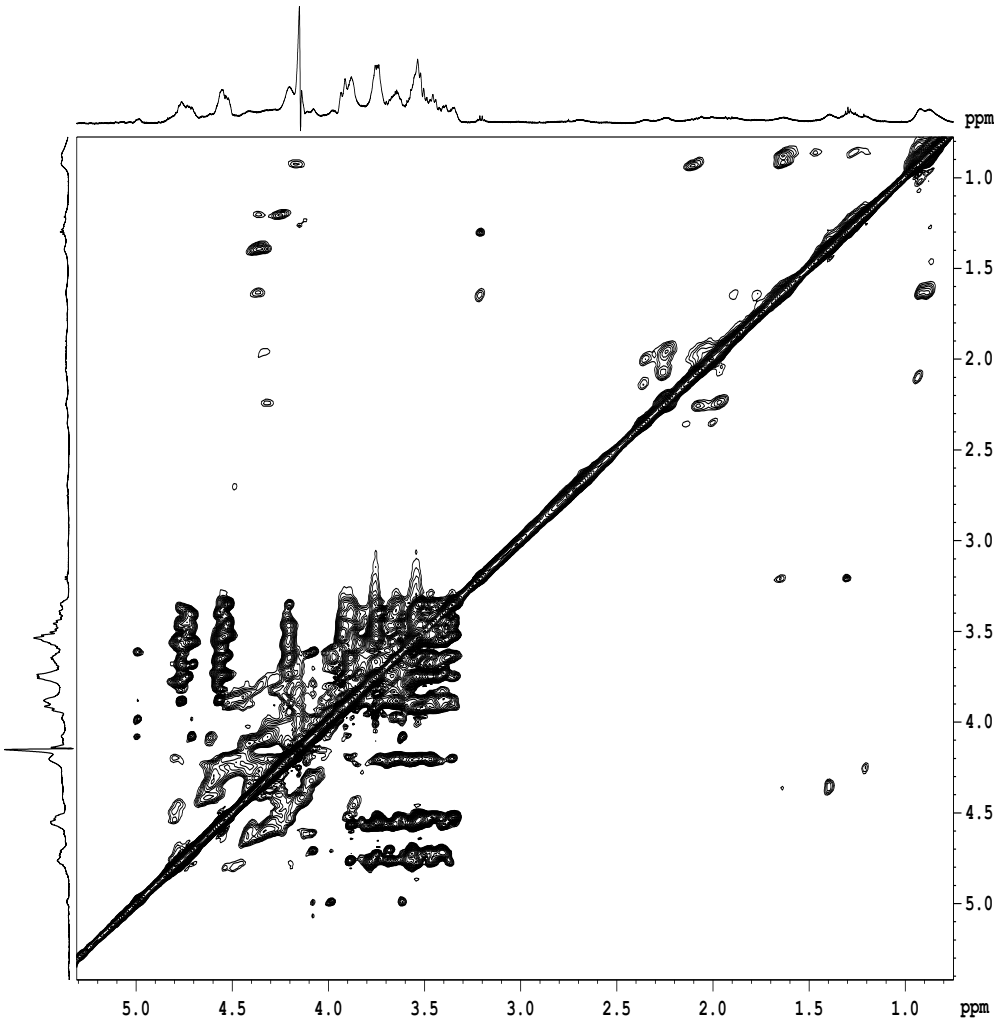


Figure 3.14: ^{1}H-^{1}H-TOCSY spectrum of sample EM9 4 mg/mL (353 K)

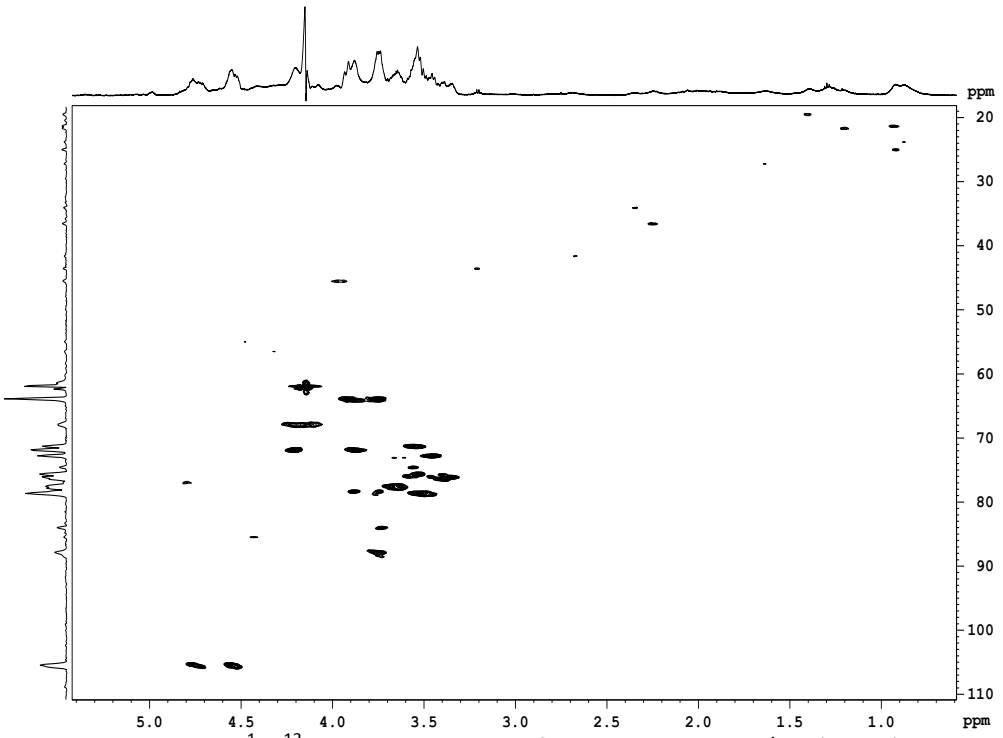


Figure 3.15: ^{1}H-^{13}C-HSQC spectrum of sample EM9 4 mg/mL (353 K)

The glucose contained in this sample (66%) has the same forms seen in the EM8 sample: the anomeric proton involved in a linkage β(1→3) gives a signal in the range 4.68-4.80 ppm, while the one involved in the linkage β(1→6 or 4) gives the signal in the range 4.50-4.60 ppm (table **3.2**)
The signal at 4.99 ppm was attributed to an anomeric proton of an α-rhamnose molecule.[81,82] However it was not possible to identify with certainty the signal of its methyl, because of the presence of many correlation peaks in the corresponding region of the spectrum.

Table 3.4: NMR ^{1}H and ^{13}C chemical shift assignments of EM9

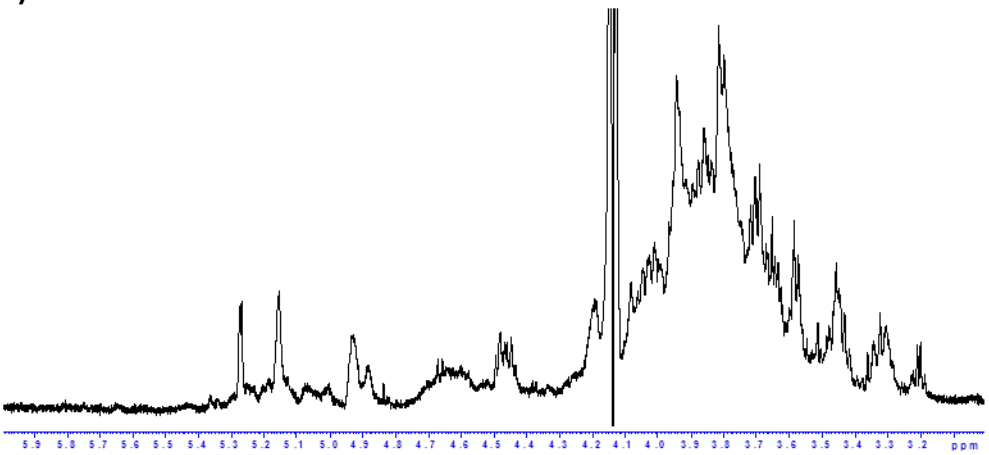
→3)αRha(1→			
Position	Type	^{1}H δ (ppm)	^{13}C δ (ppm)
1	CH	5.08	104.6
2	CH	4.00	73.0
3	CH	4.08	n. a.
4	CH	3.61	73.0

5	CH	3.88	71.8
6	CH_3	n. a.	n. a.

Sample EM2: 0,7 mL of 4mg/mL solution in P-buffer containing D_2O, with 2 mM TSP as internal standard, underwent the following experiments:

- ^{1}H-NMR T=353 K (fig. **3.16**)
- ^{1}H-^{1}H COSY T=353 K (fig. **3.17**)
- ^{1}H-^{1}H TOCSY T=353 K (fig. **3.18**)
- ^{1}H-^{13}C HSQC T=353 K (fig. **3.19**)

a)



b)

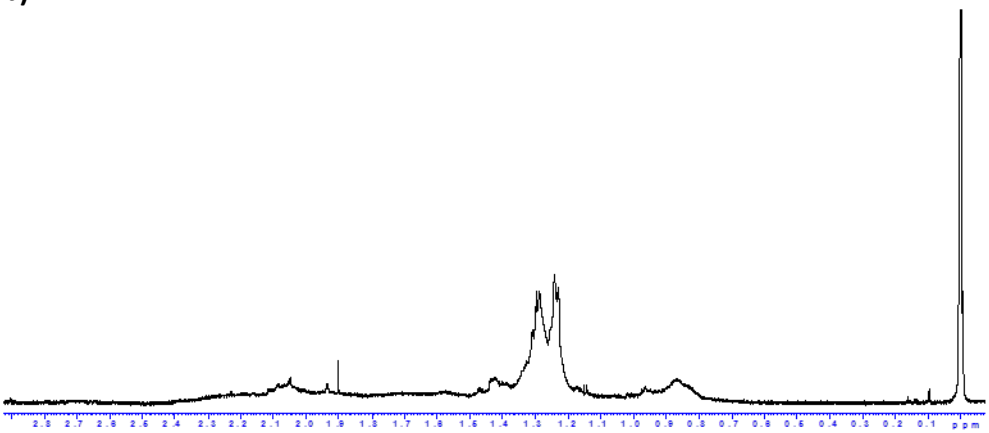


Figure 3.16 a) and b): ^{1}H-NMR spectrum of sample EM9 4 mg/mL (353 K)

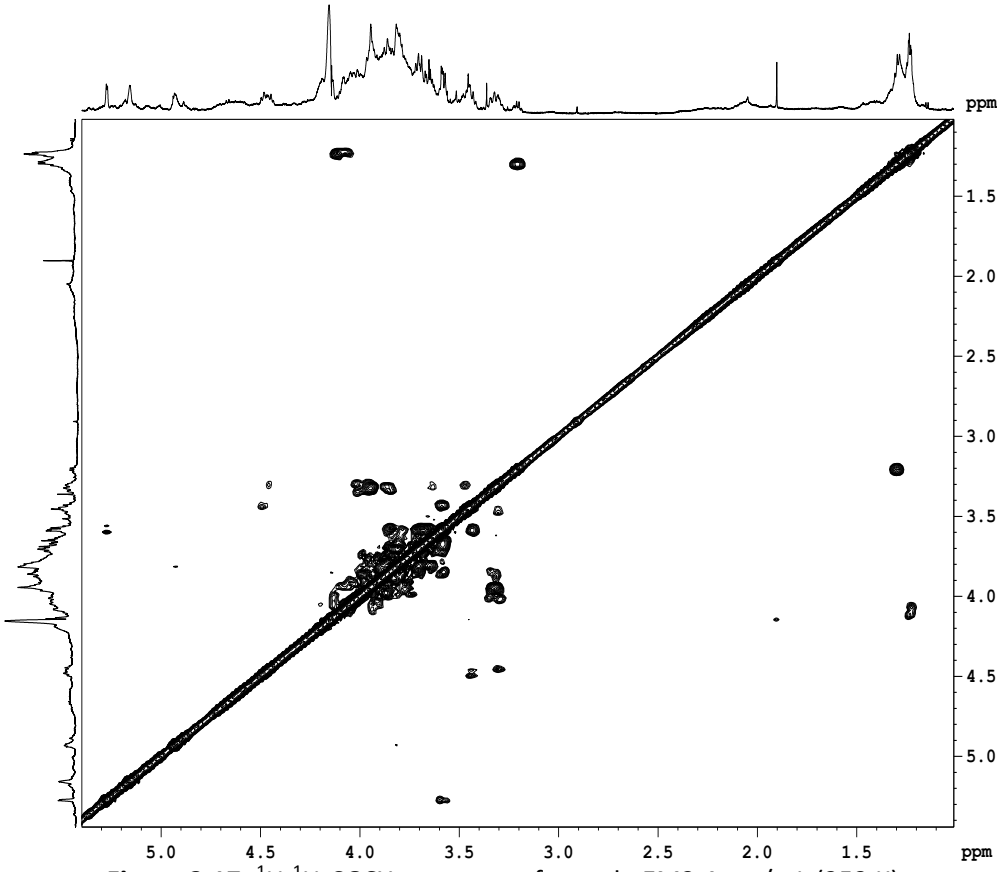


Figure 3.17: ^{1}H-^{1}H-COSY spectrum of sample EM2 4 mg/mL (353 K)

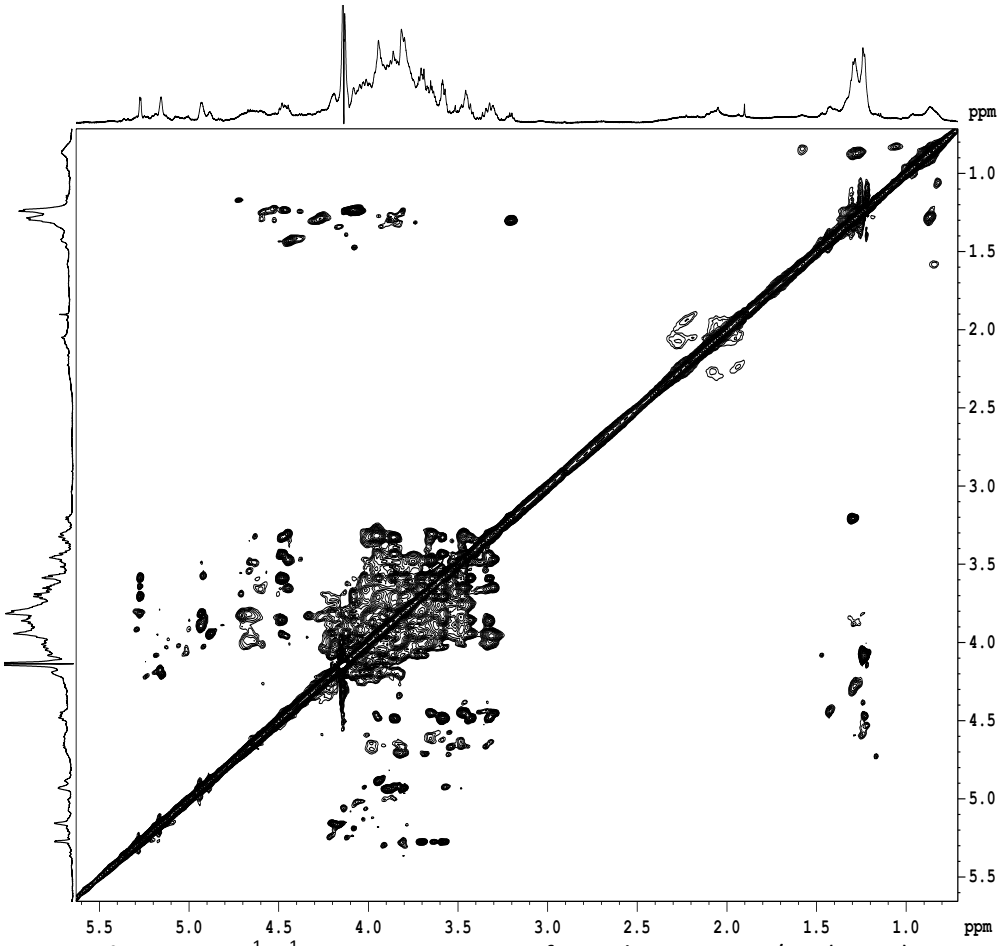


Figure 3.18: ^{1}H-^{1}H-TOCSY spectrum of sample EM2 4 mg/mL (353 K)

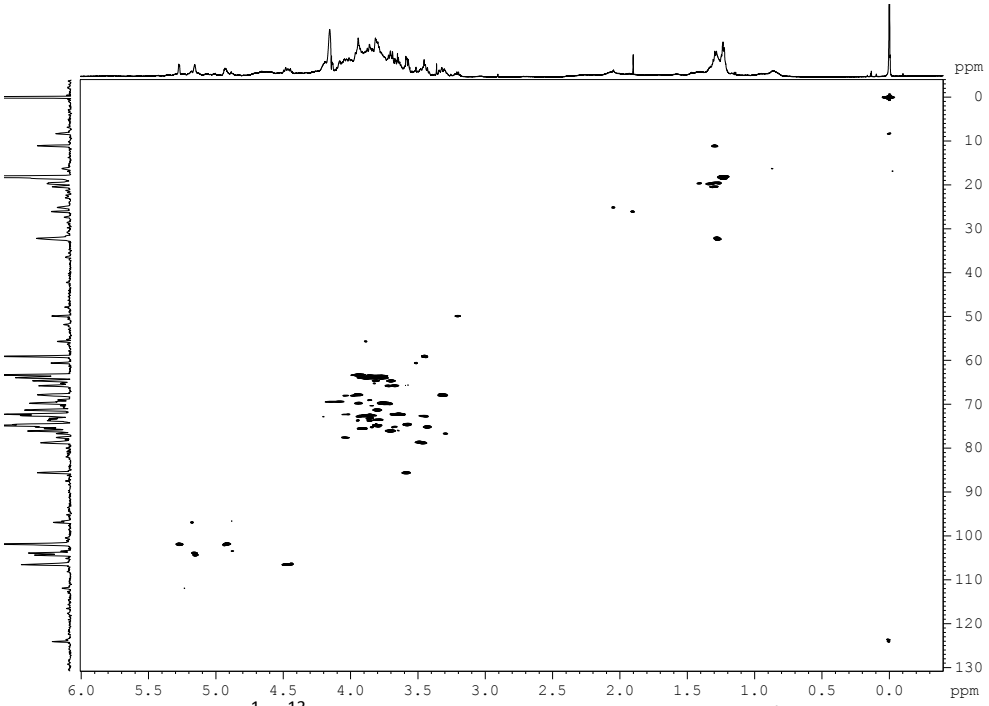


Figure 3.19: ^{1}H-^{13}C-HSQC spectrum of sample EM2 4 mg/mL (353 K)

Unfortunately, because of the complexity of the spectra, it has been possible to identify only two monosaccharide units: the α-mannose linked in position 1 and 3 at 5.16 ppm, and the α-mannose linked in position 1, 3 and 6.[83,84] at 5,15 ppm. (Table **3.5**)

Table 3.5: ^{1}H and ^{13}C chemical shift assignments of EM2

→3)α-Man (1→			
Position	Type	^{1}H δ (ppm)	^{13}C δ (ppm)
1	CH	5.161	101.8
2	CH	4.186	69.4
3	CH	4.040	77.7
4	CH	3.943	67.9
5	CH	4.085	72.3
6 / 6’	CH_2	3.819/3.920	63.9
3-6)-α-Man (1→ (tentative assignment)			
Position	Type	^{1}H δ (ppm)	^{13}C δ (ppm)
1	CH	5.153	104.3

2	CH	4.179	72.8
3	CH	3.84	75.2
4	CH	4.034	67.9
5	CH	4.030	72.3
6 / 6’	CH_2	4.012/ n. A.	70.8

Sample EM10: 0.7 mL of 5mg/mL solution in DMSO-d_6 underwent the following experiments:

- ^{1}H-NMR T=353 K (fig. **3.20**)
- ^{1}H-^{1}H COSY T=353 K (fig. **3.21**)
- ^{1}H-^{1}H TOCSY T=353 K (fig. **3.22**)
- ^{1}H-^{13}C HSQC T=353 K (fig. **3.23**)

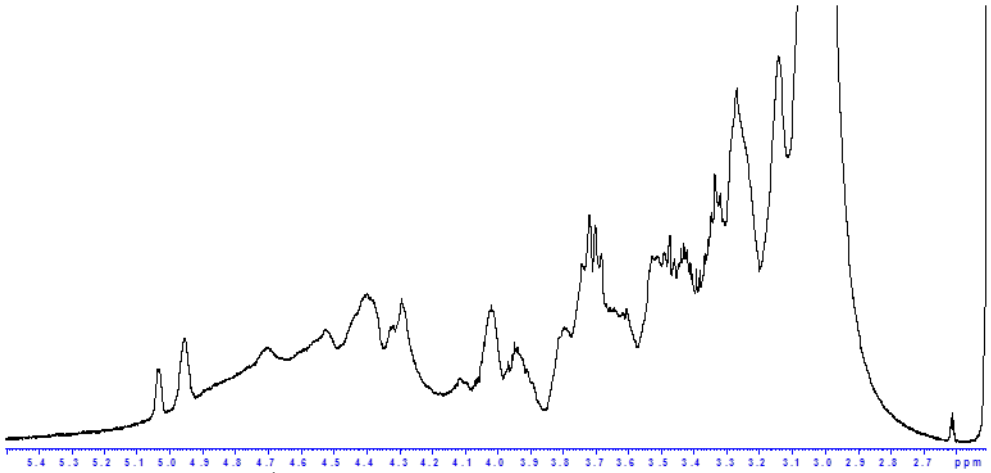


Figure 3.20: ^{1}H-NMR spectrum of sample EM10 5 mg/mL (353 K)

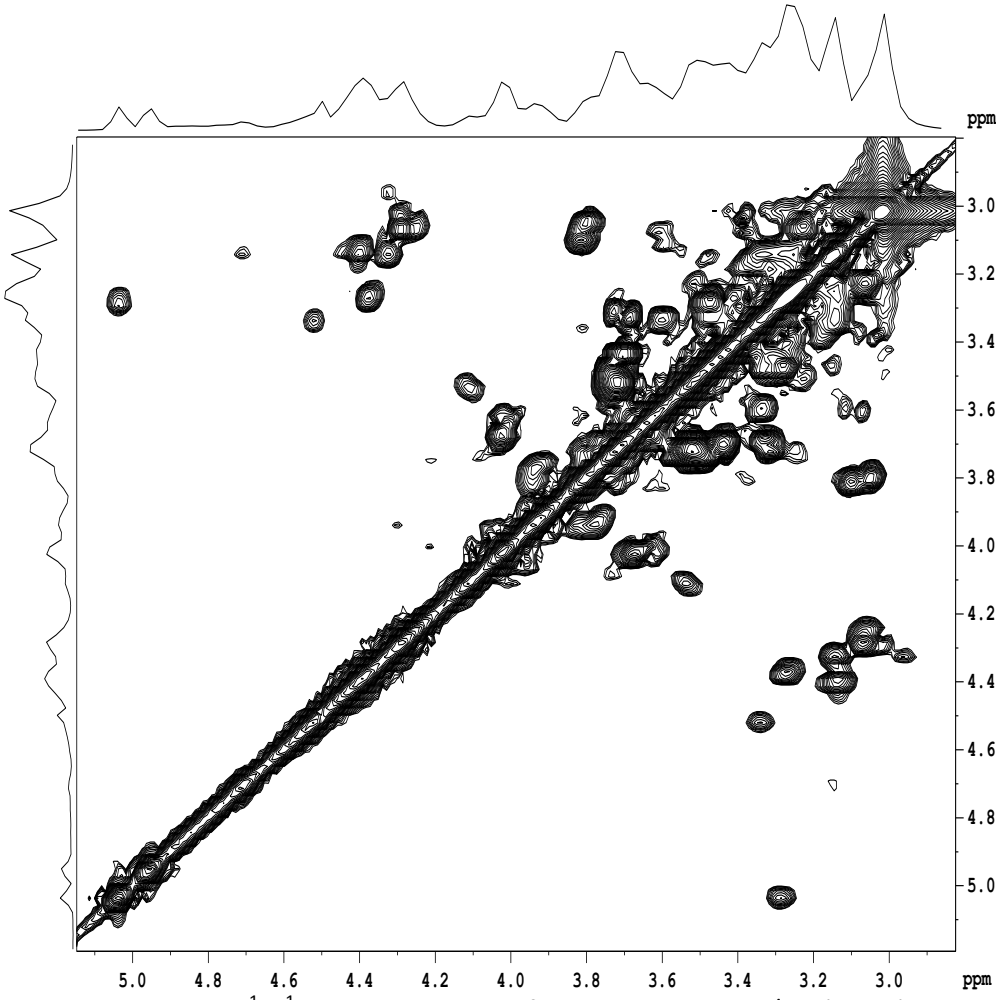


Figure 3.21: ^{1}H-^{1}H-COSY spectrum of sample EM10 5 mg/mL (353 K)

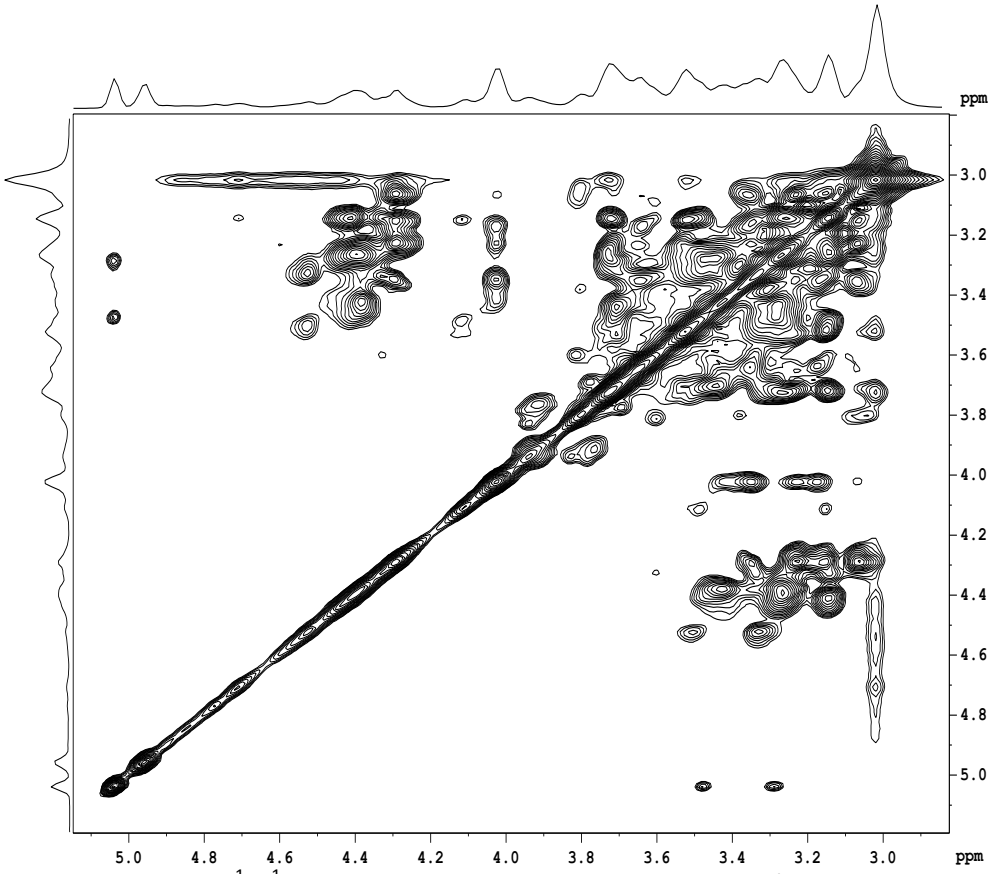


Figure 3.22: ^{1}H-^{1}H-TOCSY spectrum of sample EM10 5 mg/mL (353 K)

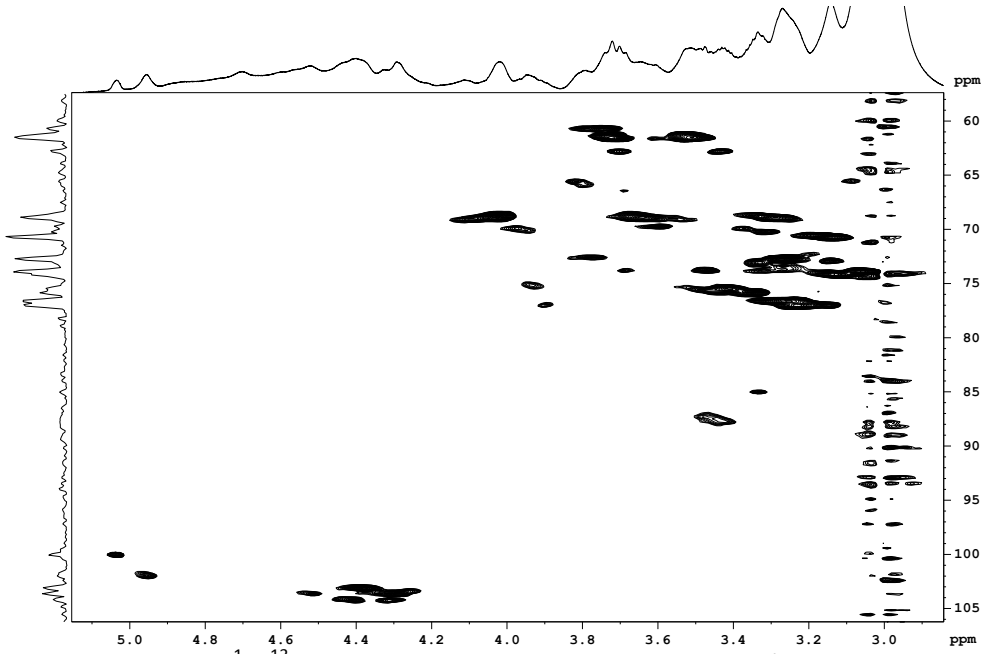


Figure 3.23: ^{1}H-^{13}C-HSQC spectrum of sample EM10 5 mg/mL (353 K)

This sample is composed mainly by glucose, but it presents also a certain amount of mannose, xylose and glucuronic acid. The analysis of the spectra allowed the identification of signals corresponding to two different forms of glucose, both β anomers. The first one is linked in position 1 and 6, the second one in position 1,3 and 6.[59] (Table **3.6**)

Table 3.6: ^{1}H and ^{13}C chemical shift assignments of EM10

→6)βGlcp(1→6)			
Position	Type	^{1}H δ (ppm)	^{13}C δ (ppm)
1	CH	4.28	103.7
2	CH	3.06	73.9
3	CH	3.22	76.9
4	CH	3.15	70.7
5	CH	3.34	75.8
6 / 6’	CH_2	3.69-4.04	68.8

→3)(6)βGlcp(1→6)			
Position	Type	^{1}H δ (ppm)	^{13}C δ (ppm)
1	CH	4.52	103.6
2	CH	3. 33	73.1
3	CH	3.50	87.6
4	CH	3.31	70.2
5	CH	3.49	75.5
6 / 6’	CH_2	4.02-3.68	68.9

Sample EM12: 0,7 mL of 5mg/mL solution in DMSO-d_6 underwent the following experiments:

- ^{1}H-NMR T=353 K (fig. **3.24**)
- ^{1}H-^{1}H COSY T=353 K (fig. **3.25**)
- ^{1}H-^{1}H TOCSY T=353 K (fig. **3.26**)
- ^{1}H-^{13}C HSQC T=353 K (fig. **3.27**)

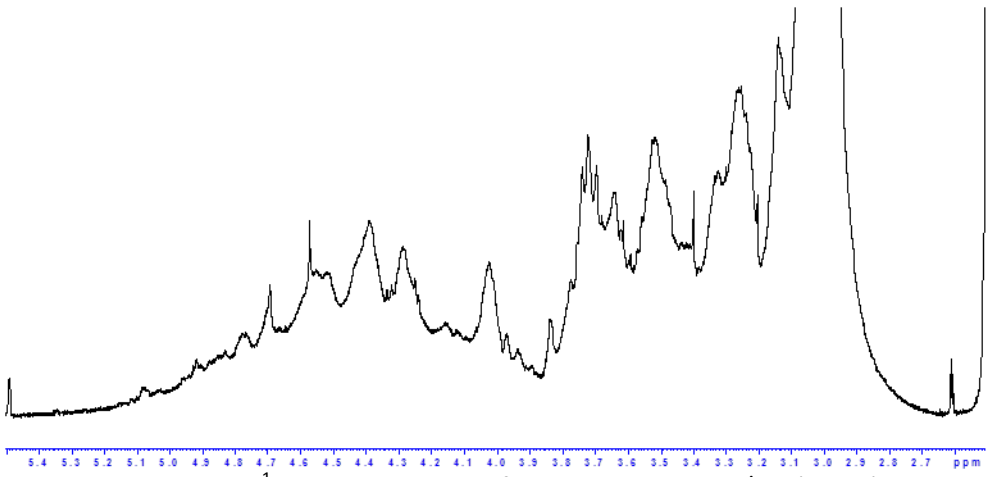


Figure 3.24: ^{1}H-NMR spectrum of sample EM12 5 mg/mL (353 K)

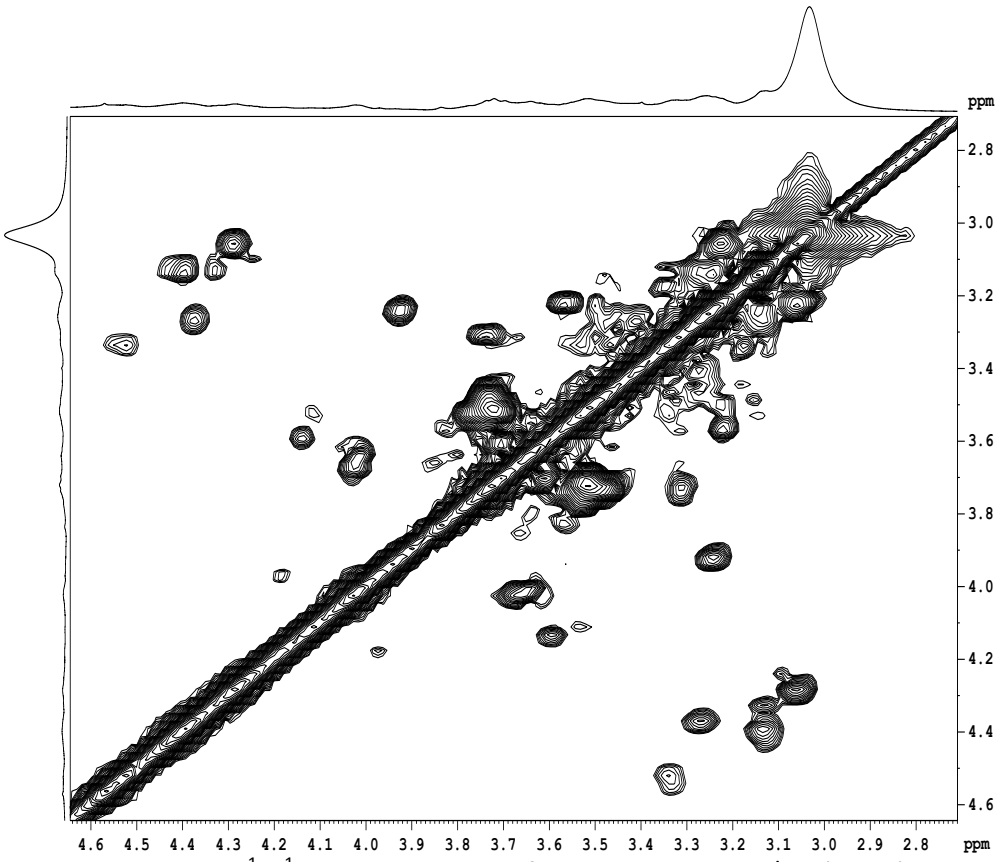


Figure 3.25: ^{1}H-^{1}H-COSY spectrum of sample EM12 5 mg/mL (353 K)

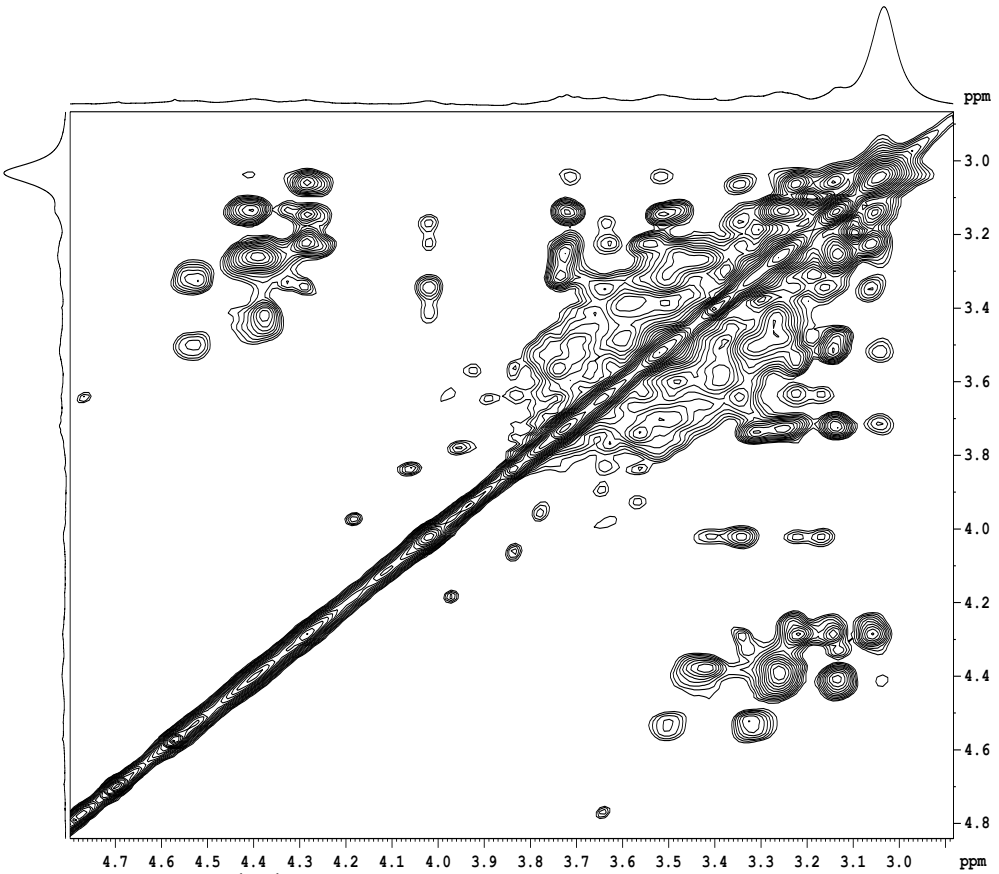


Figure 3.26: ^{1}H-^{1}H-TOCSY spectrum of sample EM12 5 mg/mL (353 K)

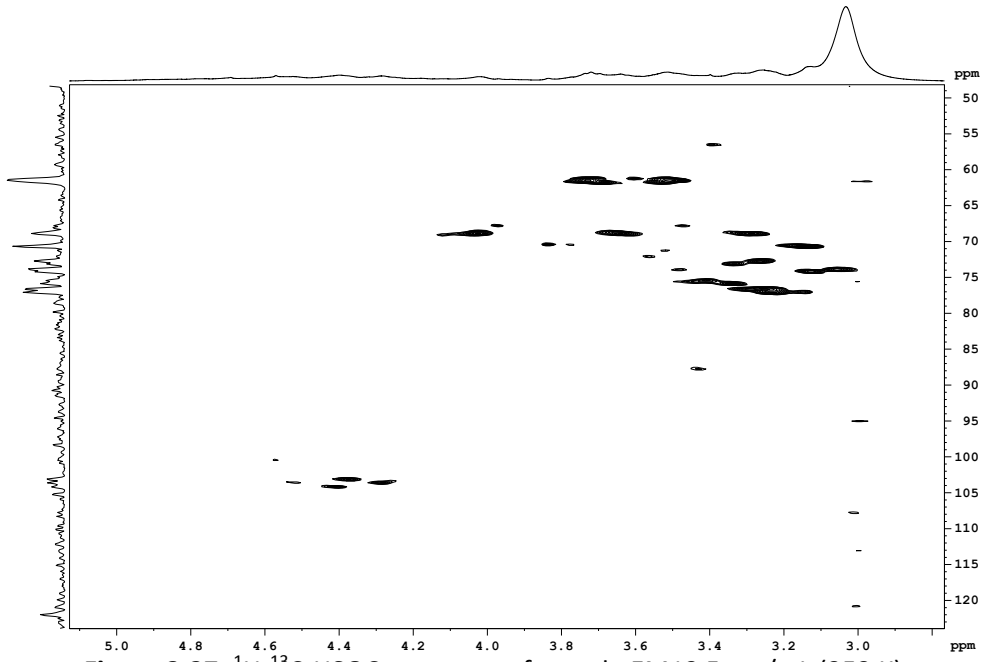


Figure 3.27: ^{1}H-^{13}C-HSQC spectrum of sample EM12 5 mg/mL (353 K)

All the spectra of sample EM12 are superimposable with those of EM10. Therefore we can assume that the monosaccharide units identified in EM10 are present also in EM12 (table **3.6**). The only difference arises for the peaks in the part of the spectra near 5.0 ppm: while in the EM10 ^{1}H spectra this region shows some signals, in the same region of the spectra of EM12 no signal is appreciated.

Unfortunately it has not yet been possible to identify the corresponding monosaccharides and we can't say whether this difference is due to the selenium, that was added in the culture medium, from which EM12 has been obtained.

3.4 Conclusions

Different polysaccharides, obtained from the mycelium and from the culture of *M. Erinaceum*, showed an immunostimolatory activity.
The NMR characterization of these polysaccharides has been carried out, allowing the identification, for many of them, of the constituent monosaccharides and of their connectivity in the polysaccharides.
Although further studies are needed to fully understand the composition of these samples, some aspects of their composition have been uncovered. The information obtained by the NMR study of these molecules may help to understand the mechanism by which the immunostimulatory activity, shown by some of these samples, is carried out.

4. References

1. Duus, J.; Gotfredsen, C.H.; Bock, K. *Chem. Rev.* **2000**, 100*,* 4589-4614

2. Dwek, R. A. *Chem. Rev.* **1996**, 96, 683-720.

3. Rudd, P. M.; Dwek, R. A. *Crit. Rev. Biochem. Mol. Biol.* **1997**, 32, 1-100.

4. Rudd, P. M.; Endo, T.; Colominas, C.; Groth, D.; Wheeler, S. F.; Harvey, D. J.; Wormald, M. R.; Serban, H.; Prusiner, S. B.; Kobata, A.; Dwek, R. A. *Proc. Nat. Acad. Sci. U.S.A.* **1999**, *96*, 13044-13049.

5. Duus, J. Ø.; St. Hilaire, P. M.; Meldal, M.; Bock, K. *Pure Appl. Chem.* **1999**, 71, 755-765.

6. Ru¨ diger, H.; Siebert, H.-C.; Solis, D.; Jimenez-Barbero, J.; Romero, A.; von der Lieth, C.-W.; Diaz-Maurino, T.; Gabius, H.- J. *Curr. Med. Chem.* **2000**, 7, 389-416.

7. Laine, R. A. *Pure Appl. Chem.* **1997**, 69, 1867-1873.

8. Gabius, H.-J. *Naturwissenschaften* **2000**, 87, 108-121.

9. Anumula, K. R.; Dhume, S. T. *Glycobiology* **1998**, 8, 685-694.

10. Rassi, Z. E.; Mechref, Y. S. *Capillary Electrophoresis, Theory and Practice*; CRC Press: London, 1997; Chapter 7, pp 273- 363.

11. van der Hoeven, R. A. M.; Hofte, A. J. P.; Tjaden, U. R.; van der Greef, J.; Torto, N.; Gorton, L.; Marko-Varga, G.; Bruggink, C. *Rapid Commun. Mass Spectrom.* **1998**, 12, 69-74.

12. Tseng, K.; Hedrick, J. L.; Lebrilla, C. B. *Anal. Chem.* **1999**, 71, 3747-3754.

13. Makino, Y.; Omichi, K.; Hase, S. *Anal. Biochem.* **1998**, 264, 172-179.

14. Carlwood, J.; Birrell, H.; Organ, A.; Camilleri, P. *Rapid Commun. Mass Spectrom.* **1999**, 13, 716-723.

15. Steinberg, P.; Fox, A. *Anal. Chem.* **1999**, 71, 1914-1917.

16. Ahn, Y. H.; Yoo, J. S. *Rapid Commun. Mass Spectrom.* **1999**, 13, 1985-1990.

17. Venkataraman, G.; Shriver, Z.; Raman, R.; Sasisekharan, R. *Science* **1999**, 286, 537-542.

18. Geyer, H.; Geyer, R. *Acta Anat.* **1998**, 161, 18-35.

19. Rudd, P. M.; Guile, G. R.; Ku¨ ster, B.; Harvey, D. J.; Opdenakker, G.; Dwek, R. A. *Nature* **1997**, 388, 205-207.

20. Rudd, P. M.; Mattu, T. S.; Zitzmann, N.; Mehta, A.; Colominas, C.; Hart, E.; Opdenakker, G.; Dwek, R. A. *Biotechnol. Genet. Eng. Rev.* **1999**, 16, 1-21.

21. Thoru, Y. *Trends Glycosci. Glycotechnol.* **1999**, 11, 227-232.

22. Sato, Y.; Suzuki, M.; Nirasawa, T.; Suzuki, A.; Endo, T. *Anal. Chem.* **2000**, 72, 1207-1216.

23. Lindberg, B.; Lo¨nngren, J. *Methods Enzymol.* **1978**, 50, 3-33.

24. Bodenhausen, G.; Ruben, D. J. *Chem. Phys. Lett.* **1980**, 69, 185-189.

25. Mu¨ ller, L. *J. Am. Chem. Soc.* **1979**, 101, 4481-4484.

26. Bax, A.; Griffey, R. H.; Hawkins, B. L. *J. Magn. Reson.* **1983**, 55, 301-315.

27. Bax, A.; Summers, M. F. *J. Am. Chem. Soc.* **1986**, 108, 2093-2094.

28. Jansson, P.-E.; Kenne, L.; Widmalm, G. *Carbohydr. Res.* **1987**, 168, 67-77.

29. Lemieux, R. U.; Kullnig, R. K.; Bernstein, H. J.; Schneider, W. G. *J. Am. Chem. Soc.* **1958**, 80, 6098-6105.

30. Bock, K.; Pedersen, C. *J. Chem. Soc., Perkin Trans. 2* **1974**, 293-297.

31. Bubb W. A. *School of Molecular and Microbial Biosciences,* University of Sydney, New South Wales **2006**, Australia

32. Hakomori S. "A rapid permethylation of glycolipid and polysaccharide, catalyzed by methylsulfinyl carbanion in dimethyl sulfoxide." *J Biochem* (Tokyo) **1964**, 55, 205–208.

33. Lemieux, R. U.; Bock, K.; Delbaere, L. T. J.; Koto, S.; Rao, V. S. *Can. J. Chem.* **1980**, 58, 631-653.

34. Vinogradov, E. V.; Petersen, B. O.; Thomas-Oates, J. E.; Duus, J. Ø.; Brade, H.; Holst, O. *J. Biol. Chem.* **1998**, *273*, 28122-28131.

35. van Halbeek, H. *Encyclopedia of Nuclear Magnetic Resonance*; John Wiley & Sons Ltd: Chichester, **1996**; 1107-1137.

36. Bock K., Tho¨gersen H. "Nuclear magnetic resonance spectroscopy in the study of mono- and oligosaccharides." *Annual Reports on NMR Spectroscopy* **1982**, 13, 1–57.

37. Bradbury JH, Jenkins GA. **1984.** Determination of the structures of trisaccharides by 13C-N.M.R. spectroscopy. Carbohydr Res 126:125–156.

38. Coxon B. **1980.** Carbon-13 nuclear magnetic resonance spectroscopy of food-related disaccharides and trisaccharides. In: Lee CK, ed. Developments in food carbohydrate, Vol. 2. London: Applied Science; 351–390.

39. Jones, C.; Mulloy, B. *Methods in Molecular Biology; Spectroscopic Methods and Analyses*; Humana Press Inc.: Totowa, NJ, **1993**; Chapter 6, 149-167.

40. Van Halbeek, H. *Methods in Molecular Biology; Spectroscopic Methods and Analyses*; Humana Press Inc.: Totowa, NJ, **1993**; Chapter 5, 115-148.

41. Casadei, M. A.; Matricardi, P.; Fabrizi, G.; Feeney, M.; Paolicelli, P. *European Journal of Pharmaceutics and Biopharmaceutics* **2007**, 67, 682–689

42. Coviello, T.; Palleschi, A.; Grassi, M.; Matricardi, P.; Bocchinfuso, G.; Alhaique, F. "Scleroglucan: a versatile polysaccharide for modified drug delivery", *Molecules* **2005,** 10, 6–33.

43. Kitamura, S.; Hirano, T.; Takeo, K.; Fukada, H.; Takahashi, K.; Falk, B.H.; Stokke, B.T. "Conformational transitions of schizophyllan in aqueous alkaline solution" *Biopolymers* **1995**, 39, 407–416.

44. Guo, B.; Elgsaeter, A.; Stokke, B.T. "Scleroglucan gel volume changes in dimethylsulphoxide/water and alkaline solutions are partly caused by polymer chain conformational transitions" *Carbohydr. Polym*. **1999,** 39, 249–255.

45. Touitou, E.; Alhaique, F.; Riccieri, F.M.; Riccioni, G.; Santucci, E. "Scleroglucan as sustained release oral preparations. Part I. In vitro experiments", *Drug Des. Deliv.* **1989**, 141–148.

46. Alhaique, F.; Carafa, M.; Riccieri, F.M., Santucci, E.; Touitou, E. "Studies on the release behaviour of a polysaccharide matrix", *Pharmazie* **1993**, 48, 432–435.

47. Rizk, S.; Duru, C.; Gaudy, D.; Jacob, M.; Ferrari, F.; Bertoni, M.; Caramella, C. "Physicochemical characterization and tabletting properties of scleroglucan", *Int. J. Pharm.* **1994**, 112, 125–131.

48. Crescenzi, V.; Dentini, M.; Silvi, F.; Paci, M.; Paradossi, G.; Bellini, L.D.; Righetto, Z. "Studies of physical and chemical gels based on microbial polysaccharides", *Bioactive Compatible Polym*. **1995**, 10, 235–248.

49. Coviello, T.; Dentini, M.; Rambone, G.; Desideri, P.; Carafa, M.; Murtas, E.; Riccieri, F.M.; Alhaique, F. "A novel co-crosslinked polysaccharide: studies for a controlled delivery matrix", *J. Control. Release*, **1998**, 55, 57–66.

50. Francois, N.J.; Rojas, A.M.; Daraio, M.E. “Rheological and drugrelease behaviour of a scleroglucan gel matrix at different drug loading” *Polym. Int.* **2005**, 54, 1613–1619.

51. Casadei, M.A.; Pitarresi, G.; Benvenuti, F.; Giannuzzo, M. “Chemical gels of scleroglucan obtained by cross-linking with 1,x-dicarboxylic acids: synthesis and characterization”, *J. Drug Deliv. Sci. Technol.* **2005**, 15, 145–150.

52. Christensen, B.E.; Aasprong, E.; Stokke, B.T. “Gelation of periodate oxidised scleroglucan (scheraldehyde)”, *Carbohydr. Polym.* **2001**, 46 241–248.

53. Coviello, T.; Grassi, M.; Rambone, G.; Alhaique, F. “A crosslinked system from scleroglucan derivative: preparation and characterization”, *Biomaterials* **2001,** 22, 1899–1909.

54. Coviello, T.; Grassi M.; Rambone, G.; Santucci, E.; Carafa, M.; Murtas, E.; Riccieri, F.M.; Alhaique, F. “Novel hydrogel system from scleroglucan: synthesis and characterization”, *J. Control. Release* **1999,** 60, 367–378.

55. Coviello, T.; Alhaique, F.; Parisi, C.; Matricardi, P.; Bocchinfuso, G.; Grassi, M. “A new polysaccharidic gel matrix for drug delivery: preparation and mechanical properties”, *J. Control. Release* **2005**, 102, 643– 656.

56. Coviello, T.; Grassi, M.; Lapasin, R.; Marino, A.; Alhaique, F. “Scleroglucan/borax: characterization of a novel hydrogel system suitable for drug delivery”, *Biomaterials* **2003,** 24, 2789–2798.

57. Coviello, T.; Grassi, M.; Palleschi, A.; Bocchinfuso, G.; Coluzzi, G.; Banisheib, F.; Alhaique, F. “A new scleroglucan/borax hydrogel: swelling and drug release studies”, *Int. J. Pharm*. **2005**, 289, 97–107.

58. Ensley, H.E.; Tobias, B.; Pretus, H.A.; McNamee, R.B.; Jones, E.L.; Browder, I.W.; Williams, D.L. “NMR spectral analysis of a waterinsoluble (1→3)-beta-D-glucan isolated from Saccharomyces cerevisiae.” *Carbohydrate Research*, **1994**, 258, 307-311.

59. Tada N., Harada T., Nagi-Miura N., Adachi Y., Nakajima M., Yodomae T., Ohno N., "NMR characterization of the structure of a β-(1→3)-D-glucan isolate from cultured fruit bodies of Sparassis crispa" *Carbohydrate research* **2007,** 342, 2611-2618**.**

60. Malinowska, E.; Krzyczkowski, W.; Łapienis, G., Herold, F. *Food Research International,* **2010**, 43, 988-995

61. Mizuno T. "Bioactive substances in *Hericium erinaceus* (Bull.: Fr.) Pers. (Yamabushitake), and its medicinal utilization." *Int J Med Mush,* **1999**, 1, 105-19.

62. Zhang A, Sun P, Zhang J, Tang C, Fan J, Shi X, et al. "Structural investigation of a novel fucoglucogalactan isolated from the fruiting bodies of the fungus *Hericium erinaceus*." *Food Chem.* **2007**, 104, 451-6.

63. Dong, Q.; Jia, L.; Fang, J. "A β-D-glucan isolated from the fruiting bodies of Hericium erinaceus and its aqueous conformation." Carbohydrate Research **2006**, 341, 791–795.

64. Jia, L.; Liu, L.; Dong, Q.; Fang, J. "Structural investigation of a novel rhamnoglucogalactan isolated from the fruiting bodies of the fungus Hericium erinaceus." *Carbohydrate Research* **2004**, 339, 2667-2671.

65. Mizuno, T. Yamabushitake, "Hericium erinaceus: Bioactive substances and medicinal utilization." *Food Reviews International* **1995**, 11, 173-178.

66. Mizuno, T.; Wasa, T.; Ito, H.; Suzuki, C.; Ukai, N. " Antitumor-active polysaccharides isolated from the fruiting body of Hericium erinaceum, an edible and medicinal mushroom called yamabushitake or houtou." *Bioscience, Biotechnology, and Biochemistry* **1992**, 56, 347-348.

67. Ooi, V. E. C.; Liu, F. "A review of pharmacological activities of mushroom polysaccharides." *International Journal of Medicinal Mushrooms* **1999**, 1, 195-206.

68. Wang, Z.; Luo, D.; Liang, Z. "Structure of polysaccharides from the fruiting

body of Hericium erinaceus Pers." *Carbohydrate Polymers* **2004**, 57, 241–247.

69. Zhang, M.; Cui, S. W.; Cheung, P. C. K.; Wang, Q. "Antitumor polysaccharides from mushrooms: a review on their isolation process, structural characteristics and antitumor activity." *Trends in Food Science and Technology* **2007**, 18, 4–19.

70. Zhang, A.; Zhang, J.; Tang, Q.; Jia, W. J.; Yang, Y.; Liu, Y., et al. "Structural elucidation of a novel fucogalactan that contains 3-O-methyl rhamnose isolated from the fruiting bodies of the fungus, Hericium erinaceus." *Carbohydrate Research* **2006**, 341, 645–649.

71. Park, Y. S.; Lee, H. S.; Won, M. H.; Lee, J. H.; Lee, S. Y.; Lee, H. Y. "Effect of an exo-polysaccharide from the culture broth of Hericium erinaceus on enhancement of growth and differentiation of rat adrenal nerve cells." *Cytotechnology* **2002,** 39, 155–162.

72. Yang, B. K.; Park, J. B.; Song, C. H. "Hypolipidemic effect of an exobiopolymer produced from a submerged mycelial culture of Hericium erinaceus." *Bioscience, Biotechnology, and Biochemistry* **2003**, 67, 1292-1298.

73. Jones, T.M.; Albersheim, P. *Plant Physiol.* **1972**, 49, 926-936

74. Lv, Y.; Yang, X.; Zhao, Y.; Ruan, Y.; Yang, Y.; Wang, Z. *Food Chemistry* **2009**, 112, 742-746

75. Malinowska, E.; Krzyczkowski, W.; Herold, F.; Łapienis, G.; S´ lusarczyk , J.; Suchocki, P.; Kuras´, M.; Turło, J. *Enzyme and Microbial Technology* **2009**, 44334-343

76. Petersen, B.O.; Krah, M.; Duus, J.O.; Thomsen, K.K. *Eur. J. Biochem.* **2000**, 267, 361-369)

77. Dinadayala, P.; Lemassu, A.; Granovski, P.; Cérantola, S.; Winter, N.; Daffé, M. *J. Biol. Chem.* **2004**, 279, 12369-12378

78. Aimanianda, V.; Clavaud, C.; Simenel, C.; Fontaine, T.; Delepierre, M.; Latge, J.-P. *J. Biol. Chem.* **2009**, 284, 13401-13412

79. Watt, D. K.; Brasch, D. J.; Larsen, D. S.; Melton, L. D. *Carbohydrate Polymers* **1999**, 39 165–180

80. Schols, H.A.; Bakx, E. J.; Schipper D.; Voragen, A. G. J. *Carbohydrate Research*, **1995**, 279, 265-279.

81. Vliegenthart, J.F.G.; Huisman, M.M.H.; Fransen. C.T.M.; Kamerling, J.P.; Schols, H.A.; Voragen, A.G.J. *Biopolymers*, **2001**, 58, 279-294.

82. Bosco, M.; Miertus, S.; Dentini, M.; Segre, A.L. *Biopolymers,* **2000**, 54, 115-126.

83. Cérantola, S.; Lemassu-Jacquier, A.; Montrozier, H. *Eur. J. Biochem*., **1999**, 260373-383.

84. Cescutti, P.; Bosco, M.; Picotti, F.; Impallomeni, G.; Leitão, J. H.; Richau, J. A.; Sá-Correia I. *Bioch. And Biophys. Research Communications*, **2000**, 273, 1088-1094.

Regards

Giunta al termine di questo percorso è un piacere poter porgere i miei ringraziamenti a coloro che hanno reso possibile tutto questo.
Primo fra tutti ringrazio il Professor Bruno Botta: grazie per avermi permesso di lavorare nel suo laboratorio sin dai tempi della mia tesi di laurea e, a seguire poi, nel corso del dottorato. Grazie per la sua disponibilità e per aver costituito un meraviglioso esempio di quanto si possa amare il proprio lavoro. La ringrazio per avermi guidato, ispirato, corretto…
Ringrazio tutti coloro che hanno partecipato alla realizzazione del mio lavoro, in particolare il Prof. Boffi, il Prof. Tafi, il Prof. Delle Monache, la Dott. D'Aquarica, e la Dott. Fabiana Subrizi, il cui contributo scientifico è stato di fondamentale importanza.
Un grazie infinite va anche alla Prof. Luisa Mannina, alla Prof. Donatella Capitani e a tutto il laboratorio di Risonanza Magnetica Nucleare "A.L. Segre" dell'Istituto di Metodologie Chimiche del CNR di Montelibretti: li ringrazio per avermi permesso di vivere una meravigliosa esperienza, accogliendomi fra loro, e per la disponibilità dimostratami. Un pensiero speciale va al Dott. Anatoli Sobolev, al Dott. Marco Gobbino e alla Dott. Raffaella Petroccione, con cui ho trascorso un'indimenticabile periodo.
Grazie anche alla Prof. Maria Antonietta Casadei e alla Dott. Federica Corrente, con cui ho avuto l'immenso piacere di collaborare.
A great thanks to Dr. Eliza Malinowska and the laboratory of Prof. Herold of the Medical University of Warsaw for their willingness and for givng me the opportunity to collaborate with them.
Grazie di cuore a tutti i ragazzi del laboratorio di chimica organica del Prof. Botta: li ringrazio per tutti i momenti meravigliosi che abbiamo condiviso, per il loro sostegno e per aver reso così piacevoli i giorni trascorsi insieme. Con loro vorrei ringraziare anche il Prof. Giovani Zappia, che ha sempre messo a disposizione di tutti la sua conoscenza, l'esperienza ma soprattutto una dose inestinguibile di buonumore.
Grazie a Carmela, Irene e Fabiola: il mio rapporto con voi è iniziato come "colleghe di laboratorio", ma si è immediatamente trasformato in una amicizia che si è rafforzata col tempo, dentro, ma soprattutto fuori dal laboratorio. Grazie per tutto quello che siete state e che siete per me.

Un grazie speciale va a Dorina: nove anni trascorsi fianco a fianco, giorno dopo giorno, nel segno di quella che è diventata una delle più profonde amicizie. Spero di aver dato a lei tanto quanto lei ha dato a me.
Ringrazio in fine le persone che hanno costituito la colonna portante di ogni mia esperienza: la mia famiglia. La loro presenza, il sostegno, l'affetto e il bene dimostratomi in ogni occasione sono stati e saranno sempre alla base di tutto.