4. Materials and methods

4.1 Methodology

In metabolome analysis of microbial cells the separation of intra- and extra-cellular metabolites is often required to gain insight into the control of metabolic pathways.

Since the metabolite levels reflect the ultimate response of a biological system to genetic or environmental changes, rapid stopping (quenching) of cell metabolism and extracellular enzymatic activity must be the first step for such methods. The quenching procedure aims to stop cellular metabolism and prevent the turnover of metabolites, maintaining the *in vivo* metabolite concentrations at a constant level [119]. The commonly used method is based on rapid sampling into a cold aqueous methanol solution, as proposed by de Koning and van Dam [120]. This method is preferred in microbial research because it is currently the only method that enables a reliable separation of intra- and extra-cellular metabolites while stopping cell metabolism, achieving the best reproducibility compared to the other methods.

However, membrane integrity study [119] showed that the yeast cell membrane is damaged by contact with both buffered and non-buffered cold methanol solution: the metabolite levels in the supernatant of quenched samples confirmed that the methanol solution caused damage to the yeast cell membrane.

Since the cells are usually separated from the quenching mixture by centrifugation (requiring sample handling at low temperatures to eliminate or at least minimize turnover of metabolites prior to cell extraction) increasing the speed and decreasing the time of centrifugation to separate the yeast cells after quenching resulted in a considerably reduced number of metabolites that leaked from the cells. This shows that the longer the time the cells stay in contact with the quenching solution, the higher is the leakage of intracellular metabolites, which is especially the case for amino acids [119].

So, following quenching, the second step in the sample preparation is the separation of the extracellular metabolites from the biomass and subsequent extraction of the intracellular metabolites, achieving minimal losses by degradation and avoiding further biochemical reactions.

The choose of the extraction procedure depends on the matrix tissue nature, on its origin (animal, vegetal, microbic) and on the informations level desired: only one compound class or all classes present? Ideally, the extraction of any type of tissue should be stress- and shock- free: by examining different extraction procedures [119] it has become evident that there is a strong influence of the extraction method on the metabolite profile of yeast cells. Moreover, excessive sample manipulation can introduce error in the chemical analysis, such as contamination and/or inadequate extraction of the analytes [121].

So we need an extraction protocol that is able to trap a higher diversity of intracellular metabolites of yeasts. This step, in particular, is the most critical in the analysis of total fat, neutral and polar lipids, and fatty acids composition.

First good extraction of polar lipids from animal tissues was achieved by Folch in 1957 [122] based on a chloroform:methanol 2:1 mixture. Again, de Koning and van Dam [120] adapted this methodology to extract polar metabolites from a yeast cell suspension. If the use of organic solvents is not critical, this extraction methods provides high yields, but it is important to note that it is also expensive because of the high amount of solvent required to carry out the extraction process, without considering chloroform and methanol toxicity.

Another popular extraction method for intracellular metabolites of yeasts is based on the use of boiling ethanol. However, for metabolome analysis, this method is not free of problems since not all metabolites are stable at the high temperature applied.

Classical acidic and alkalinic extraction methods are commonly used for the extraction of acid and alkaline stable compounds. These methods have been widely used in the extraction of metabolites from animal and plant tissues but for metabolome studies it is evident that not all metabolites are stable at the extreme pH and destruction of some nucleotides has been reported during acidic extractions [119].

The Bligh and Dyer method [123] is particularly suitable for the extraction of incubation medium, tissue homogenates or cell suspensions, instead [124,125]. It is certainly the most widely used method for total lipids determination and it is accepted as giving the most accurate results [121]. It all began back in the 1950s at the Halifax Laboratory of the Fisheries Research Board of Canada, where Dr. William J. Dyer and his group were investigating the deterioration of fish muscle proteins during frozen storage. Existing procedures were unsatisfactory and attention was focused on the use of mixtures of chloroform and methanol to isolate the lipids from moist biological materials. Bligh and Dyer introduced a method where extraction and partitioning are simultaneous, using a mixture of three solvents (chloroform, methanol and water), whereas the precipitated proteins (*pellet*) are isolated between two liquid phases.

This procedure, modified by our laboratories [126], allow us to obtain a chloroformic phase containing all the lipids and an hydro-alcoholic phase (characterized by a stronger extraction capacity respect to a simple acqueous one) containing amino acids, carbohydrates and all the others polar low molecular weight compounds.

Very often a third step in sample preparation is required: the concentration of the extract. Lyophilization, or freeze-drying, is commonly used to remove water from aqueous samples in order to concentrate and avoid thermal degradation. Alternatively, non-aqueous extracts can be concentrated by solvent evaporation. It should be noted, however, that both methods are unsuitable for volatile metabolites.

Our method allow us to obtain, all at once, two extract of different nature but directly confrontable each other, simplifying the original complex mixture. Moreover, analyzing two type of extract for the same sample it will be possible to identify, with a double test-stand, eventual anomalies in each sample (outliers, etc). Finally the obtained extracts are clear enough to be used as NMR spectroscopy good samples.

However, unwanted matrix components have to be removed, sample adulteration has to be excluded and the analytes of interest have to be well dissolved in the chosen NMR solvent. Furthermore, the physic-chemical characteristic of the samples (pH, ion strength...) are required to be constant over the whole sample set [63].

4.1.1 Samples obtainment

The *Kluyveromices lactis* strains used in this study are PM6-7A (*MATa ade-T600 uraA1-1*) and PM6-7A/VV30 (*MATa ade-T600 uraA1-1 rag8-1*). Cultures were grown under shaking condition at 28°C in YP (1% Difco yeast extract, 2% Difco Bacto-peptone) supplemented with 0.5 g/L of glucose. When the cells reached 0.7 OD₆₀₀, 1.6 g/L of glucose was added to the medium and the cells were grown till late exponential phase (OD₆₀₀ ~ 3). In this condition it was assumed that metabolic stationary state is reached: so culture cells were washed three times with cold distilled water (H₂O_{dd}). After the supernatant was eliminated, the wet mass is determined. Seven samples for each strain were obtained.

4.1.2 Labeling experiment

Four samples were prepared from the same cellular pool, for each strain: two exactly in the same way of the formers (named controls), whereas the others by adding $[1,2^{-13}C_2]$ -glucose as substrate (enriched samples). Samples obtainment procedure is the same and in this case it ensures the reachment of both metabolic and isotopic stationary state.

4.1.3 Metabolites extraction

The biomass was suspended in 900 μ L of cold methanol (-20°C) to quench intracellular metabolism. Then the sample was quantitatively transferred in centrifuge tubes with other methanol, reaching the final volume of 2 mL. Moreover 2 mL of chloroform and 1 mL of H₂O_{dd} were added. The mixture was kept overnight at 4°C. Polar and organic phases were separated by centrifugation at 10500 rpm at 4°C for 30 min. The polar phase (top layer) and organic phase (bottom layer) were separately collected, dried under N₂ flux and stored at -80°C until analyzed by NMR.

4.1.4 Sample preparation for ¹H-NMR analysis

The freeze-dried polar samples were re-dissolved in 600µL of D₂O phosphate buffer solution (pH = 7.4) containing 2 mM sodium 3-(tri-methylsilyl)propionate-2,2,3,3-d₄ (TSP, 98 atom % D, CAS n 24493-21-8, PW 172.28, Sigma-Aldrich, USA) as ¹H NMR reference, and transferred to 5 mm NMR glass tubes for analysis. The organic phases were re-dissolved in 600 µL of CDCl₃ containing 2 mM hexamethyl-di-siloxane (HMDSO, \geq 99.5%, NMR grade, CAS n 107-46-0, FW 162,38, Sigma-Aldrich, USA), which has a chemical shift of 0.055 ppm compared to tetramethylsilane (TMS), as an ¹H NMR reference.

4.1.5 Sample preparation for ¹³C-NMR analysis

Both polar and organic freeze-dried samples were re-dissolved in 0.5 mL of the same formers solutions containing the chemical shift reference (TSP and HMDSO, respectively). As quantitative reference a coaxial tube containing 70 μ L of pure acetonitrile (CH₃CN for HPLC, 99.9 % GLC, Carlo Erba, CAS n 75-05-8, PM 41.053; d(20°C) 0.783) was inserted in the samples.

4.2 NMR spectroscopy

The nuclei of all elements possess mass and charge. One or more isotopes of most nuclei also have spin, i. e., angular momentum.

Since spinning charge creates a magnetic field, there is a magnetic moment, μ , associated with the angular momentum. It is this property of matter that is exploited in nuclear magnetic resonance (NMR) spectroscopy. The magnetic moment, a vector quantity, can be aligned in the presence of an intense static magnetic field (H_0), generating a magnetization, then manipulated in space, i. e., caused to evolve in time under the influence of specific interactions and, finally, observed [127]. The NMR experiment consists of inducing transitions between the states of quantized magnetization. This is accomplished by irradiating the sample with radio frequency (rf) energy. The pulsed technique was developed largely in response to the need for much higher sensitivity in ${}^{13}C$ spectrometry. This higher sensitivity is achieved by exciting all of the nuclei of interest (¹³C or ¹H) simultaneously, as an instantaneous "scan". A short (microseconds, µs), powerful, rf pulse of center frequency v_1 applied along the x-axis generates the entire, desired frequency range; then all of the signals were collected simultaneously in a single scan. So acquiring NS scans the signal to noise ratio (S/N) can be improved [128]. The frequency of the rf (v_1) must exactly match the precession frequency (Larmor frequency) of the nuclei in order to cause the transition because the resonance condition were satisfied [127]. The result is the so-called free induction decay (FID), which may be described as a decaying interferogram. The signals collected represent the difference between the applied frequency v_1 and the Larmor frequency of each nucleus (carbon or proton). The complete framework of pulse sequences and time delays is collectively called the *pulse sequence*.

4.2.1 Sequences

a. 1D¹H-NMR experiment

The simplest 1D NMR experiments, shown in Figure 4.1, involve the application of a simple onepulse, followed by observation of the resulting FID in the time domain (during the acquisition time), with subsequent Fourier Transformation (FT) of the data to the frequency domain for presentation in a format that we, as chemists, can understand (Fig 4.2) [127]. The length of the rf excitation pulse (P1), generally at $\theta = 90^\circ$, should be optimised for each samples [129].



Figure 4.1: 1D ¹H pulse sequence [129]. Figure 4.2: FID Fourier Transformation (FT).

Morever one is usually interested in performing signal averaging, acquiring the FID for NS scans to improve S/N ratio, so, before each subsequent scan, it becomes necessary to wait many times T_1 to allows the system to relax and recover the full signal once again. A relaxation delay (D1) of 5 T_1 , leading to 99.3% recovery of longitudinal magnetisation, is generally accepted [130].



Figure 4.3: Presaturation sequence [130].

In all biofluids the large interfering signal arises from water: even if samples are re-dissolved in D_2O , there will be always present residual H_2O , indeed. This signal must be eliminated using an appropriate solvent-suppression method.

Presaturation sequence shown in fig. 4.3 involves low on-resonance irradiation at the rf of the solvent signal, during D1, for a 2 s period, followed by acquisition with a 90° pulse. The presaturation power is selected to attenuate the solvent resonance significantly. Naturally, higher powers will produce greater suppression, but these should not be so high as to reduce the neighbouring resonances. Typically "dummy" scans (DS) are collected to achieve a steady-state start condition [130].

b. 1D¹³C-NMR experiment

The calculation of each carbon enrichment is achieved by the acquisition of 1D ¹³C-NMR spectra. A fundamental problem in ¹³C NMR spectroscopy is its inherent lower sensitivity respect to ¹H, so that it is required a longer acquisition time to obtain an acceptable signal-to-noise ratio. However, there are several strategies for enhancing sensitivity, including increase the magnetic field strength: for this reason the acquisition of the ¹³C-NMR spectra was performed with a 600 MHz instrument at the CNR – Montelibretti laboratory (Rome, Italy) [87].

Observation of ¹³C-NMR signals fine structure due to the ¹J_{CC} coupling, essential for the interpretation of such spectra, crucially depends on the efficient removal of the large ¹J_{CH} scalar coupling interactions [32], achievable using a variety of proton decoupling techniques. All decoupling procedures are based on the same principle: the broadband irradiation of all ¹H resonances with a high-power amplitude-modulated and monochromatic ¹H frequency. The irradiation saturates the Boltzman energy levels of the ¹H spectrum, removing the macroscopic magnetization of all protons and making the scalar ¹H couplings disappear from the ¹³C spectrum [2]. There are two main proton decoupling methods:

- 1. broad band decoupling (BB) uses a continuous irradiation of proton frequencies with a train of rectangular pulses of identical duration and opposite phase;
- composite pulse decoupling (WALTZ) uses blocks of ¹H pulses of different pulse lengths and phases. This sequence are more efficient than BB, allowing decoupling ¹³C spectra using less decoupler power and diminishing also the potential dielectric sample heating.

In addition to the removal of scalar ¹H couplings, ¹H irradiation can cause an increase in the ¹³C signal intensity because of the Nuclear Overhauser Enhancement (NOE) effect. NOE arises as a result of the redistribution of spin populations of a nucleus when the spin transitions of another coupled nuclei are perturbed from their equilibrium populations.

The magnetization transfer process from ¹H to ¹³C, called heteronuclear NOE, increase the intensity of the ¹³C resonance to three-fold (200%) because $\gamma_{\rm H} / \gamma_{\rm C} = 4$ (Eq. 1).

$$\eta_{I}{S} \ll = \frac{\gamma_{S}}{2\gamma_{I}}$$
 100 Eq. 1
S is always refer to the perturbed (or Source) spin and I to the enhanced (or Interesting) spin.

Different decoupling schemes are possible that produce proton-coupled or proton-decoupled ¹³C NMR spectra containing (or not containing) NOE enhancement, like figure 4.4 shows [2].



- a) If the ¹H decoupler is not used in conjunction with ¹³C excitation (fig 4.4 a), a proton-coupled ¹³C spectrum is obtained. The multiplet structure of the proton-coupled ¹³C resonance depends on the number of protons coupled to the observed carbon and follows the rule: 2nI + 1 (for n = 1, a doublet is obtained).
- b) If the decoupler is gated only during the acquisition of the ¹³C FID (inverse gated decoupling), a proton-decoupled spectrum devoid of NOE is obtained (fig 4.4 b). On the contrary, if the decoupler is gated only during the relaxation delay and not during the acquisition (direct gated decoupling), a proton-coupled NOE-enhanced ¹³C spectrum is obtained, generating ¹H-coupled ¹³C spectra with complete NOE enhancement (not shown).
- c) Decoupling during the complete cycle time generates a proton-decoupled NOE-enhanced ¹³C spectrum (fig. 4.4 c).

Observation of the fine structure relative to the coupling pattern between adjacent ¹³C, fundamental for isotopomers identification, crucially depends on the efficient removal of the large ${}^{1}J_{CH}$ scalar coupling interactions [32]. This was accomplished using an inverse-gated decoupling; furthermore, this avoids the effects of NOE that could compromise ${}^{13}C$ quantitative analysis.

c. Two-dimensional experiments

A general 2D experiment can be characterized by four time-periods [131]. During each time-period, one uses RF pulses, time-delay or a combination of pulses with time delay(s) depending on the need of the experiment.



- 1. The first time-period in a pulse sequence is called *preparation period*. Typically, a $\pi/2$ RF pulse may be applied during this period to create transverse magnetization. When a pulse sequence is applied repeatedly for achieving time averaging, a relaxation delay (D1) precedes the $\pi/2$ RF pulse to allow the system to regain thermal equilibrium from the perturbations caused to the spin-system by the previous pulse sequence. A decoupler pulse may be applied to saturate the solvent signal during this period.
- 2. During the *evolution period* (t₁), the transverse magnetization is allowed to precess under the free Hamiltonian (\mathcal{H} in the absence of RF; comprising of the terms \mathcal{H}_{δ} and \mathcal{H}_{J} in solution state). During this period, each spin is characterized by its specific precessional frequency and acquires a specific phase. The evolution time t₁ is incremented in steps of Δ t₁: such incremental delays constitute the second time-dimension, often called "indirect dimension".
- 3. During the *mixing period*, transfer of magnetization (coherence transfer) takes place between the spins which are coupled to each other (through J coupling, dipole-dipole interaction or chemical exchange). In other words, spins are made to interact or mix during this time.
- 4. For each t_1 incremental period, a separate FID is acquired during the detection period, as a function of time t_2 . Each FID thus collected, reflects the history of the spin system as a function of t_1 .

Thus, one acquires a 2D dataset S (t_1 , t_2) consisting of detected FID signal as a function of two time variables (t_1 and t_2). A double-Fourier transformation of the resulting data matrix provides the 2D NMR spectrum with two frequency domains (f_1 and f_2) [127]:

$$S(t_1, t_2) \rightarrow s(f_1, f_2)$$

Viewing events with the vector model, the first pulse of a general sequence (for example a 90_x), places the equilibrium magnetisation in the x-y plane along the +y axis, after which it will precess (or evolve) according to this chemical shift offset. After a time period t₁, the vector moves through an angle of $360vt_1$ degrees and is then subject to the second pulse (in general, another 90_x). This pulse places the y-component along the -z axis whilst the x-component, unaffected, producing the detected FID [130].



Now imagine repeating the experiment a number of times with the t_1 interval increased by a uniform amount each time and the resulting FIDs stored separately. Subjecting each of the acquired t_2 FIDs to Fourier transformation, a series of spectra are produced and each resulting signal intensity (or amplitude) results modulated as a function of time according to sin 360vt₁ (fig. 4.5 a), as t_1 increases from zero and x-magnetisation developed during the evolution period.

With longer values of t_1 , relaxation effects diminish the intensity of the transverse magnetisation, according to T_2^* , so the signal intensities show a steady (exponential) decay in addition to the amplitude modulation. The intensity of the resonance as a function of time therefore represents another FID for the t_1 time domain (referred to as an interferogram) that has been generated artificially or indirectly (fig. 4.5 b).



Figure 4.5 a: Amplitude modulation of a singlet resonance as a function of the evolution period t_1 . At longer values of t_1 , signal intensity is diminished by spin relaxation [130].



Figure 4.5 b The variation in peak intensity of the amplitude-modulated resonance of Fig. 4.5a produces a FID (interferogram) for the t_1 domain [130].

A second FT with respect to t_1 then gives the spectrum in two frequency dimensions S (f_1 , f_2). Thus, the signals in a 2D spectrum are characterized by a pair of frequencies (f_1 and f_2). Depending on the pulse sequence used, the two frequencies may correspond to the same spins (homonuclear 2D) or different spins (heteronuclear 2D). In a homonuclear 2D spectrum, the diagonal peaks ($f_1 = f_2$) correspond to the 1D NMR experiment. The off-diagonal peaks (commonly called *cross peaks*) carry important information about interactions between different spins [131].

The advantages of 2D spectroscopy can be readily conceived. The advent of 2D spectroscopy has resolved several problems associated with resolution, solvent suppression and assignments, which one encounters while dealing with large molecules or complex mixture, because in presence of more than one chemical species, each one can be characterized individually. In fact, since the information is spread in two dimensions, the chemical shifts get resolved by the cross peaks, even when one of the two chemical shifts is in a crowded region suffering from spectral overlaps.

d. 2D homonuclear TOCSY

TOCSY (TOtal Correlation SpectroscopY) is an important and widely used 2D technique. It provides information about coupled spin systems complete networks and helps in overcoming the problems of peak overlap during resonance assignments [131]. Pulse sequence is shown in fig. 4.6a.



Figure 4.6 a: TOCSY pulse sequence.

Figure 4.6 b: Spin lock pulse scheme.

The experiment begins by exciting the spins with a 90_x pulse such that they all lie along the +y axis in the rotating frame. At this point the spin-lock (SL) rf field (H₁) is applied, this time along the +y axis. In its simplest form, the spin-lock is a continuous low-power pulse of constant phase applied for a period of typically tens of milliseconds (D9). The multiple-pulse SL sequence most commonly used is MLEV-17 [129]. The composite pulse cycle is essentially MLEV-16. This is followed by a $(\pi)_x$ pulse used to remove the overall effect of pulse imperfections that may occur during the composite pulse cycle. The two trim pulses (usually set to 2-3 msec each) at the beginning and end of the mixing period serve to defocus any magnetization that is not parallel to the x-axis and ensures proper phasing of 2D spectrum into absorption mode [131].

However, it is convenient to imagine spin lock to be composed of a continuous sequence of closely spaced 180_y pulses bracketed by infinitely small periods δ (fig. 4.6 b). Each δ –180– δ period constitutes nothing more than a homonuclear spin-echo sequence in which a 180° pulse refocuses the evolution of chemical shifts, so that no shift evolution has occured and all spin vectors remain along the +y axis. Extending this argument for the whole of the mixing period, it is apparent that no net chemical shift evolution occurs during t_m and hence the nuclear vectors are said to have been *spin-locked* in the rotating frame along, in this case, the y axis. In contrast, homonuclear spin–spin couplings continue to evolve [130].

In summary, during the SL mixing, all chemical shift differences in the rotating frame are removed whereas spin–spin couplings remain active. Nuclei that share a coupling and have very similar (or coincident) chemical shifts relative to their coupling constant are said to be strongly coupled. Under such conditions, forced upon all spin systems by the application of the SL field during the mixing period, protons loose their unique identity and this provides the mechanism by which coherence may be shared over all spins within the same spin system. Since magnetisation may travel in either direction along a spin chain, 2D TOCSY spectra are symmetrical about the diagonal.

The requirement for nuclei to experience identical local fields during the mixing time for transfer to occur between them is also referred to as the Hartmann–Hahn match. More formally and more generally, this may be expressed as follows:

$$\gamma_A H_{1A} = \gamma_X H_{1X}$$

where H_{1A} and H_{1X} are the H_1 fields experienced by spins A and X and γ_A and γ_X are their magnetogyric ratios, respectively. The inclusion of the γ s in this means the nuclides involved in the exchange need not be the same [130]. Homonuclear vicinal coupling coherence is established during the evolution period t_1 . The extent to which magnetization is propagated from the vicinal neighbor to its vicinal neighbor, and so on, is a function of the mixing time and the size of the homonuclear coupling constants between the vicinally coupled protons in question. The longer the mixing time, the further magnetization is propagated through the contiguous homonuclear vicinal coupling network [127].

e. Inverse heteronuclear shift correlation

Heteronuclear shift correlations are 2D techniques that correlate different nuclei establishing correlations through the chemical bonds. In addition, the experiment may be used as a means of spreading the resonances of a complex proton spectrum according to the chemical shift of the directly attached nucleus, utilising the typical greater dispersion of the X spin chemical shifts to assist with proton interpretation.

The high sensitivity of modern correlation techniques often provides a fast method for indirectly determining chemical shifts of the X-nucleus and avoids the need for its direct observation altogether, offering considerable time savings [130].

The advantage of inverse experiments is that the nucleus with the highest γ (usually ¹H) is detected, with greatly increased sensitivity (theoretically, if $X = {}^{13}C$, eight times more sensitive than the other schemes using ¹H excitation and X detection) [32,129].

These inverse sequences aim at detecting only those protons that are bound to a spin- $\frac{1}{2}$ X, or in other words, only the satellites of the conventional proton spectrum. But, in the case of ¹³C, only 1 in every 100 proton spins will contribute to the 2D spectrum (the other 99 being attached to NMR-inactive ¹²C). When the FID is recorded, all protons will induce a signal in the receiver on each scan and the unwanted resonances, which clearly represent the vast majority, must be removed if the correlation peaks must be revealed.

Suppression of the dominant parent resonance can be performed by inverting the first ¹³C pulse on alternate scans, so that the phase of the ¹³C satellites is inverted whereas the ¹²C-bound protons remain unaffected. Simultaneous inversion of the receiver will thus lead to cancellation of the unwanted resonances with corresponding addition of the desired satellites. This two-step procedure is called phase cycle.

The problem with this scheme is that clean suppression of the unwanted resonances is unlikely to be achieved by phase cycling alone, with residual signals contributing to undesirable bands of t_1 noise in the resulting spectra, which may mask the genuine correlations [130].

To make the inverse shift correlation sequences generally applicable, the more recent and certainly most effective approach is through the incorporation of pulsed field gradients (PFGs) into the sequences. The gradient-selected sequence operates by employing a suitable combination of gradients that refocus only those responses that have followed the desired transfer pathway. But this retention of only one of the two possible transfer pathways is the reason for the reduced S/N ratio.

2D heteronuclear correlation spectroscopy is widely used in biomolecular NMR. However these methods still need considerable spectrometer time, especially when employed with samples of very low concentration [90].

The two most common inverse chemical shift correlation experiments are HSQC and HMBC, which are used to determine which ¹H of a molecule are bonded to which ¹³C nuclei (or other X nuclei).

f. HSQC (Heteronuclear Single Quantum Correlation)

The experiment utilizes an INEPT step to transfer single quantum magnetization from proton to the directly bonded heteronuclide, immediately prior to the beginning of the evolution time t_1 (fig. 4.7). The X-nucleus single quantum magnetisation evolves during t_1 with the proton 180° pulse at its midpoint refocusing ¹H–X coupling evolution. So, decoupling such heteronuclear coupling interaction, only heteronuclear chemical shifts remain in f_1 (indirect dimension). Following evolution, magnetization is transferred back to the protons and refocused, using a "retro-INEPT" pulse sequence, to produce, once again, in-phase ¹H magnetization, which is detected with broadband decoupling during the acquisition time t_2 [127,131].

HSQC crosspeaks reveal the directly-bound ¹³C for each ¹H resonance considered.



Figure 4.7: HSQC pulse sequence [130].

g. HMBC (Heteronuclear Multiple Bond Correlation)

HSQC experiment presented above depends upon the presence of a proton bound to the heteroatom, for example 13 C, so it is therefore unable to provide assignments for non-protonated centres and does not produce unambiguous carbon assignment.

An alternative approach in such cases is to establish correlations between carbons and neighbouring protons over more than one bond, so-called long-range or multiple-bond correlations, most commonly through the proton-detected HMBC experiment (fig. 4.8).



Figure 4.8: HMBC pulse sequence [130].

The sequence starts with proton excitation followed by evolution of ¹H-magnetisation during the Δ_{LR} preparation period. This delay is set to a sufficiently long time to allow the small, long-range ¹H-¹³C couplings to evolve, producing the anti-phase displacement of vectors required for the subsequent generation of heteronuclear multiple-quantum coherence. This anti-phase magnetisation may be transferred to the coupled partner by the action of the subsequent rf pulse: the role of this first ¹³C pulse is to generate ¹H-¹³C multiple-quantum coherence that cannot be directly observed.

Since this multiple-quantum coherence contains terms for both transverse ¹H and ¹³C magnetisation, during the subsequent t_1 period it will evolve under the influence of both ¹H and ¹³C chemical shifts. To remove the effect of proton shifts during t_1 , a spin-echo is incorporated by placing a proton 180° pulse at the midpoint of t_1 so that, by the end of the evolution period, these shifts have refocused and thus have no influence in f_1 . Evolution of the carbon shifts is unaffected by the proton pulse instead, so these remain to produce the desired frequency labelling according to ¹³C shifts only, characteristic of the indirectly detected f_1 dimension.

The final carbon pulse then reconverts the multiple-quantum coherence back to observable singlequantum ¹H-magnetisation.

For sensitivity reasons, no refocusing period is performed in HMBC so that long-range heteronuclear couplings are anti-phase at the start of t_2 , precluding the application of 13 C-decoupling.

The result is a 2D spectrum with ¹H shifts represented in f_2 , ¹³C shifts in f_1 and crosspeaks indicating long range connectivities. Crosspeak intensities depends upon, among other things, both the magnitude of the long-range coupling and the value selected for Δ_{LR} .

4.2.2 Acquisition

Defects in lineshape can be minimized by using exactly the same sample volume (0.6 mL) and by optimizing the magnetic field homogeneity (manual shimming) before data acquisition [67].

¹H NMR spectra were acquired at 298 K using a Bruker Avance III 400 spectrometer (Bruker BioSpin GmbH, Germany) equipped with a magnet operating at 9.4 Tesla, where the ¹H nucleus resonates at 400.13 MHz. The probe-head was a 5 mm diameter multinuclear PABBO BB-1H/D (Z108618/0044) equipped with z-gradient.

The pulse sequence adopted for ¹H NMR spectra acquisition was a presaturation - single 90° detection pulse - acquire – delay sequence where the D1 relaxation delay was optimised to 2.5 s to allow the acquisition of 64k data point in about 5.5 s, satisfying full relaxation conditions. The length of the detection pulse was calibrated previously to the acquisition of each spectrum, the spectral width was set to 6009.62 Hz (15 ppm) and 64 scans were collected for each spectrum.

¹³C NMR spectra were acquired at 298 K using a Bruker Avance AQS600 spectrometer operating at the proton frequency of 600.13 MHz.

The pulse sequence adopted was an inverse gated decoupling: the decoupler was gated only during the acquisition of the ¹³C FID, after a single 90° detection pulse, calibrated before the acquisition of each spectrum. The D1 relaxation delay was set to 7 s for all the spectra, acquiring 64k data point in about 0.83 s. The spectral width was set to ~ 39370 Hz (~ 260 ppm) and 8192 scans were collected for each spectrum to achieve an acceptable signal-to-noise ratio.

4.2.3 Processing

¹H and ¹³C 1D-NMR spectra were processed by using the 1D-NMR Manager ver. 12.0 software (Advanced Chemistry Development, Inc., Toronto, Ontario, Canada). 2D-NMR experiments were processed by using Bruker Topsin ver. 1.3.b software (Bruker Biospin Corporation, Billerica, Massachusetts, U.S.A.).

4.2.4 Assignment

The assignment of the peaks to specific metabolites was achieved using standard two-dimensional ¹H-¹H TOCSY, ¹H-¹³C HSQC and HMBC experiments and confirmed by comparison with literature data [68,132-144].

4.2.5 Quantitative Analysis

The addition of a reference compound at known concentration (TSP, HMDSO 2 mM) had quantitative purpose: since the area of each signal is proportional to the number of proton that generates the signal itself, each integral value was first normalized for this number of protons and then the concentration of each integrated metabolite were obtained through a simple proportion. Finally, was applied an ulterior normalization considering the wet weight of each sample, to make biological samples with different biomass comparable.

Relative to organic extracts, reposing to the assigned resonances, fatty acid (FA) content for each sample was calculated [135]. Unsaturated fatty acids (UFA) were derived from the integral of the allylic protons, resonating at 2.02 ppm; polyunsaturated fatty acids (PUFA) were represented by the sum of linoleic and linolenic acid contents, calculated from the integral of the diallylic CH₂ (~ 2.8 ppm) and terminal CH₃ (~ 0.88 ppm) signals, respectively; Monounsaturated fatty acids (MUFA) were calculated as UFA – PUFA; finally saturated fatty acids (SFA) were calculated subtracting UFA content from the total amount of fatty acids revealed in the sample, represented by the integral of the β CH₂ signal at 1.59 ppm.

4.3 Statistical analysis

When an experiment is performed it is essential also to know how the results can be analyzed [145]. Statistics is the study of data collection, organization, analysis, interpretation and presentation: it ensures that data collection is undertaken so that valid conclusions are produced, reporting results and summarizing data in ways (tables and graphs) comprehensible to those who must use them [146].

The goal is to investigate causality and, in particular, to draw a conclusion on the effect of changes in the values of predictors (or independent variables) on dependent variables (or response). Practically, statistical analysis gives us a way to quantify the confidence we can have in our data interpretation [147].

4.3.1 Univariate analysis - Student's t-test

In probability and statistics, Student's *t*-distribution (or simply the *t*-distribution) is a family of continuous probability distributions that arises when estimating the mean of a normally distributed population in situations where the sample size is small and population standard deviation is unknown. It plays a role in a number of widely used statistical analyses, including the Student's *t*-test for assessing the statistical significance of the difference between two sample means, assuming that the means of these two populations are equal (null hypothesis).

The statistical significance is expressed by the *p-value* (generally 0.05) representing the probability (5%) that the difference observed between the two means is accidental.

The independent samples *t*-test (or unpaired *t*-test) is used when two separate sets of independent and identically distributed samples are obtained, one from each of the two populations being compared, to determine if the two sets of data are significantly different from each other.

The assumptions underlying such unpaired *t*-tests are that:

- each of the two populations being compared should follow a normal distribution under the null hypothesis;
- the two populations being compared should have the same variance;
- the data used to carry out the test should be sampled independently from the two populations being compared.

Statistical comparison between the experimental groups (i. e. the means of the metabolites concentration between wild type and mutant samples) were performed by unpaired Student's *t*-test. Differences with a p value of less than 0.05 were considered significant [68].

4.3.2 Multivariate analysis

Detecting and identifying correlation pattern changes is the first and often crucial step in understanding underlying biochemical changes in the metabolism of an organism [63].

Similar to other omic technologies, NMR spectroscopy is a suitable technique for the simultaneous quantification of a large number of metabolites simultaneously, without any a priori hypothesis regarding the involved biochemical pathways and leading to the identification of a specific metabolic fingerprint [68].

Such technique generates multivariate data, consisting of many different variables (metabolites concentration) recorded for each observation (sample) [148]. This torrent of data necessitates the development of algorithms and approaches to analyze them.

The complexity of NMR profiles can be reduced and the interpretation of biological influences made easier by using multivariate computer-based pattern recognition methods [63].

Pattern recognition and related multivariate statistical approaches can be used to discern significant patterns in complex data sets and are particularly appropriate in situations in which there are more variables than samples in the data set [62].

The general aim of these methods is to describe objects (in this case, the samples) by reducing the dimensionality of complex data set, facilitating the visualization of inherent patterns that could not be found by analyzing each variable separately [149].

The simplest multivariate statistical method is Principal Component Analysis (PCA): it generates an unsupervised overview of natural clustering among biochemically similar samples, grouping them according to their similarity without using knowledge of sample class [59,63].

PCA analysis was carried out using Unscrambler 9.8 Software (CAMO, Oslo, Norway). Prior to PCA, data are often pre-treated, in order to transform them into a form suitable for the analysis: re-shaping the data improves the interpretability of the model.

Variables often have substantially different numerical ranges, indeed. A variable with a large range has a large variance, whereas a variable with a small range has a small variance. Since PCA is a maximum variance projection method, it follows that a variable with a large variance is more likely to be expressed in the modeling than a low-variance variable. In order to give all variables equal weight in the data analysis, we standardized them (scaling or weighting). There are many ways to scale the data, but the most common technique is the **autoscaling**, that apply an unit variance scaling followed by a mean-centering (fig. 4.9).

In the first part of this standard procedure for pre-processing, one calculates, for each variable (column), the standard deviation (s_k) and obtains the scaling weight as the inverse standard deviation $(1/s_k)$. Subsequently, each column of the data matrix is multiplied by $1/s_k$. In this way, each scaled variable has equal (unit) variance. Then, with mean-centering the average value of each variable is calculated and then subtracted from the data.



Figure 4.9: Autoscaling procedure scheme.

Briefly, the idea beneath the multivariate Principal Component Analysis (PCA) is that each subject is represented by a single data point in a n-dimensional space where each measured variable $(X_1 ... X_n)$ is one of the coordinate axis (fig. 4.10). The goal of PCA is the reorientation of the cloud of data points along new axes, that are algebraically expressed by a linear combination of the original variables, to maximize the amount of the overall variance embedded in the data.

Statistically, PCA finds lines, planes and hyperplanes in the n-dimensional space that approximate the data as well as possible in the least squares sense. It is easy to see that a line or a plane that is least squares approximation of a set of data points makes the variance of the co-ordinates on the line or plane as large as possible. So one can say that PCA derives a model that fits the data as well as possible in the least squares sense or, alternatively, that PCA may be understood as maximizing the variance of the projection co-ordinates.



Figure 4.10: N-spatial representation of the samples (yellow balls) respect to the measured variables (X_n).

So, PCA performs a linear transformation of the original variables associated with samples into a set of new variables, the principal components (PCs), which are orthogonal to each other and explain progressively less variance in the data set [148].

In detail, the first principal component (PC1) is the line in the n-dimensional space that best approximates the data in the least squares sense. This line goes through the average point. Each observation may now be projected onto this line in order to get a co-ordinate value along the PC-line (*score*).

Extending the model with the second PC, this is also represented by a line in the n-dimensional variable space, which is orthogonal to the first PC. This line also passes through the average point and improves the approximation of the X-data as much as possible.

So, the first two PC together define a plane into the n-dimensional variable space. By projecting all the observations onto this low-dimensional sub-space and plotting the results, it is possible to visualize the structure of the investigated data set.

Extract a smaller number of PC, that account for most of the variation in the original multivariate data, has the effect of summarize the data without including vast amount of irrilevant noise and with little loss of information [148].

The co-ordinate values of the observations on this plane are called *scores* (fig. 4.11) and hence the plotting of such a projected configuration is known as a *scores plot*.

For example, in a two dimensional scores plot, each observation (sample) will be characterized by two values, one along the PC1 and another along PC2, indicating their co-ordinates relating the respective PC (score value). This scores plot allows the analysis of the distribution and grouping of the samples in the new variable space: observation close to each other have similar properties, whereas those far from each other are dissimilar with respect to the profile given by the linear combination of the measured variables and that describe some aspect of the investigated system, often unknown at the beginning.

Moreover, scores plot analysis allows to identify *outliers*: conceptually they are observations that are extreme or that do not fit the PCA model. In particular outliers have high leverage on the model, i. e., strong power to pull the PCA model toward themselves so that they may "consume" one PC just because of their existence. For this reason if they are found, they must be removed from the data.

A diagnostic showing outliers is given by Hotelling's T^2 . This statistic is a multivariate generalization of Student's t-test defining, when used in conjuction with a score plot, a tolerance region corresponding to, for istance, 95% or 99% confidence [150].





Figure 4.11: Scores values are obtained by the projection of each observation (sample) to each PC-axes.

Figure 4.12: The direction of each PC in relation to the original variable, indicates how the original variables "load" into the corresponding PC.

Moreover, since each component is a linear combination of the original input variables, each variable has a weight (*loading*) which indicates the strenght of influence that variable has on the overall profile for a set of samples. Geometrically, the principal component loadings express the orientation of the model plane in the n-dimensional variable space. The direction of PC1 in relation to the original variables is given by the cosine of the angles α_1 , α_2 , α_3 ... α_n (fig. 4.12).

These values indicate how the original variables $X_1 \dots X_n$, "load" into (contribute to) PC1: hence, they are called loadings. These weights are normalized from 0 to 1 and can also be visually inspected in the so-called *loadings plot* allowing the identification of influencing variables because it displays the relationships between all the variables measured, at the same time [63].

Variables contributing similar information are grouped together, that is, they are correlated. Positive correlation means that when the numerical value of one variable increase or decrease, the numerical value of the other variable has a tendency to change in the same way. When variables are negatively (inversely) correlated they are positioned on opposite sides of the plot origin, in diagonally opposed quadrants. Furthermore, the distance to the origin also conveys informations: the further away from the plot origin a variable lies, the stronger impact that variable has on the model.

4.4 Isotopomer analysis

Considering only the ¹²C and ¹³C isotopes in the carbon backbone of a molecule M with n carbon atoms, remembering that an **isotopomer** of M is one of the 2ⁿ possible labeling states in which this molecule can be encountered, the corresponding **isotopomer fraction** denotes the percentage of molecules in this specific labeling state.

The **positional enrichment** at the ith carbon atom within a metabolite M is then the sum of all isotopomer fractions of M where the ith carbon atom is labeled, i. e. the percentage of isotopomers labeled at the ith specific position [32]. An important difference between these two concepts is that the isotopomer fractions of M always add up to 100%, whereas positional labeling fractions have no such constraint (fig. 4.13) [151].





For the precise determination of isotopomer abundances from ${}^{13}C$ fine structures, three phenomena have to be considered [32]:

nuclear spin relaxation: it may distort the resonance intensities originating from differently labeled molecules. The incorporation of ¹³C introduces additional relaxation pathways for ¹³C and ¹H via ¹³C-¹³C and ¹³C-¹H dipolar interactions. Since ¹³C-¹³C dipolar interactions are very small compared to the one-bond ¹³C-¹H dipolar interaction, significant effects will usually be registered only for non-protonated carbons.

- 2. Strong spin-spin scalar coupling effects: whenever the scalar coupling constant ${}^{1}J_{CC}$ involving two carbons is comparable to or larger than the corresponding chemical shift difference, significant distortions of the line intensities of a given multiplet occur and the integration of the signal for the quantitative analysis is compromised because the spectrum become of superior order. However, this effects are largely absent when using modern high-field NMR spectrometer.
- 3. ¹³C isotope effects on carbon chemical shifts: the chemical shift of a given carbon depends slightly on the number of attached ¹³C nuclei [152]. Due to these ¹³C isotope effects multiplet components arising from isotopomers with ¹³C-¹³C spin pairs are shifted by ≈ 1 4 Hz to higher field relative to the singlet line. Consequently, the superposition of the multiplets arising from different isotopomers is not symmetric with respect to the central singlet line.

Accurate isotopomer abundances can usually be obtained from the spin–spin scalar coupling fine structure detected either in ¹H-decoupled 1D ¹³C-NMR or 2D [¹³C-¹H] HSQC spectra. Since these interactions allow identification of pairs of ¹³C-nuclei within the carbon backbone, the determination of a complete set of relative isotopomers becomes, at least in principle, feasible.

However, only the one-bond scalar couplings $({}^{1}J_{CC})$ are usually large enough (30 – 60 Hz) to be well resolved in 13 C-NMR spectroscopy, while long-range couplings are often too small (< 5 Hz) and moreover the dispersion of the ${}^{1}J_{CC}$ values is often insufficient to distinguish the signals from two different spin pairs ${}^{13}C^{a}$ – ${}^{13}C$ and ${}^{13}C^{b}$ – ${}^{13}C$ [32].

These are the reasons why complete isotopomers analyses are often not possible for molecules with more than three carbon atoms and ¹³C-NMR spectral analysis is limited to revealing the abundance of groups of isotopomers.

4.4.1 Fractional enrichment

The increased intensity of the resonances observed in the spectrum obtained after infusion of ¹³C-labeled substrate, indicates that the ¹³C label has been incorporated in the metabolites.

The interpretation of ¹³C NMR spectra in terms of flux through metabolic pathways requires the quantification of the ¹³C incorporated in specific carbons. This is normally done by expressing ¹³C incorporation as a fractional ¹³C enrichment in carbon C_i (Y_{Ci}). The fractional ¹³C enrichment Y_{Ci} is defined as the amount of ¹³C relative to the total carbon (¹³C+¹²C) present in C_i [2].

$$Y_{Ci} = {}^{13}C_i / ({}^{13}C_i + {}^{12}C_i)$$

Special precautions must be taken in the determination of the concentration of ${}^{13}C$ in C_i by ${}^{13}C$ NMR because ${}^{13}C$ resonances from carbons with the same ${}^{13}C$ concentration may have different intensities or areas depending on the particular pulsing condition used, the relaxation behavior of the observed carbon, its NOE effect, acquisition and repetition times.

The intensity, area, or magnetization M(t) of a ¹³C carbon resonance is function of M(0) (the area of the resonance under equilibrium magnetization conditions), Φ (the flip angle used in the ¹³C pulse), η (the NOE enhancement factor), t (the total cycle time), T_1 (the longitudinal relaxation time), and a (the acquisition time) [2].

M(0) is proportional to the concentration of ¹³C. Thus, if the ¹³C spectrum is acquired under fully relaxed conditions, a comparison of the area M(0) in the sample with the area of the same resonance in a standard solution (not labeled) of known concentration would give a value for the ¹³C concentration in the sample. In practice, this procedure is not recommended because T_1 values of ¹³C in extracts vary from approximately 3 s in methyl and methylene resonances to 30 s or more in quaternary carbons like those of carboxylic acids.

Thus to recover completely the equilibrium magnetization M(0) in the sample, the slowest carbons would require a total cycle time of approximately 150 s. This is incompatible with reasonable acquisition times [2].

In a ¹³C NMR spectrum, the signal of the labeled carbon atom appears to be enhanced compared to that of natural abundance carbon ($\approx 1.1\%$ ¹³C in each position) [90]. So ¹³C-NMR quantification strategies that may be used to determine fractional ¹³C enrichments comprise the direct comparison of the areas of the ¹³C resonances in the sample 1D ¹³C spectrum with those of the same resonances in the spectrum of a standard solution of known concentration, acquired under identical pulsing conditions. Moreover, note that the ¹H NMR signal from a hydrogen atom attached to ¹³C is split into a pair of signals: so, the relative area of the ¹³C satellites to the total area of the ¹H resonance, or comparison of the intensities of the split and unsplit peaks, gives a direct measure of the enrichment of the carbon atom [83].

For this work, two enriched samples for each strains (WT and rag8 mutant) were obtained performing labelling experiments with [1,2-¹³C₂]glucose as substrate.

The difference between the area of the natural abundance signal (1.1% Y_{Ci}) and the total area of the corresponding multiplet resonance in the spectrum obtained with ¹³C-labeled substrate (enriched samples), acquired under identical pulsing conditions, reveals the net amount of ¹³C incorporated in the corresponding carbon [2]; then, normalising for the metabolite concentration, the fractional enrichment is obtained.

So, NMR is ideal for discerning the fractional enrichment at specific carbon atoms in particular metabolites. These fractional enrichments give primary data from which metabolic fluxes can be deduced, making possible to construct an appropriate flux model that matches the underlying metabolic network [57].