# Untargeted analysis of TD-NMR signals using a multivariate curve resolution approach: application to the water-imbibition kinetics of Arabidopsis seeds

Silvia Mas Garcia, Jean-Michel Roger, Mireille Cambert, Corinne C. Rondeau-Mouro

# Untargeted analysis of TD-NMR signals using a multivariate curve resolution approach: application to the water- imbibition kinetics of Arabidopsis seeds

Silvia Mas Garcia[1,2*], Jean-Michel Roger[1,2], Mireille Cambert[3], Corinne Rondeau-Mouro[3]

[1] ITAP, INRAE, Institut Agro, University Montpellier, 34196 Montpellier, France

[2]ChemHouse Research Group, 34196 Montpellier, France

[3]INRAE, UR1466 OPAALE, 17 Avenue de Cucillé, CS 64427, F-35044 Rennes, France

[*]Author for correspondence:

E-mail: silvia.mas-garcia@inrae.fr

**Abstract**

The aim of this study is to investigate the ability of Time-Domain Nuclear Magnetic Resonance (TD-NMR) combined with Multivariate Curve Resolution Alternating Least Squares (MCR-ALS) analysis to detect changes in hydration properties of nineteen genotypes of Arabidopsis seeds during the imbibition process.

The Hybrid hard and Soft modelling version of MCR-ALS (HS-MCR) applied to raw TD-NMR data allowed the introduction of kinetic models to elucidate underlying biological mechanisms. The imbibition process of all investigated hydrated Arabidopsis seeds could be described with a kinetic model based on two consecutive first-order reactions related to an initial water absorption around the seed and a posteriori hydration of the internal seed tissues, respectively. Good data fit was achieved (LOF %= 0.98 and $r^2$ %=99.9), indicating that the hypothesis of the selected kinetic model was correct. An interpretation of the mucilage characteristics of the studied Arabidopsis seeds was also provided.

The presented methodology offers a novel and general strategy to describe in a comprehensive way the kinetic process of plant tissue hydration in a screening objective. This work also proves the potential of the MCR methods to analyse raw TD-NMR signals as alternative to the controversial and time-consuming pre-processing techniques of this kind of data, known to be an ill-conditioned and ill-posed problem.

**Keywords:** Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS), Hybrid soft- and hard-modeling-Multivariate Curve Resolution (HS-MCR), Imbibition process, Arabidopsis, mucilage, Time-Domain Nuclear Magnetic Resonance (TD-NMR).

## 1. Introduction

Arabidopsis (*Arabidopsis thaliana*) seeds belong to species known to release a gel from the seed coat on imbibition in water. This gel, made of polysaccharides, forms a sticky, gelatinous halo around the seed termed mucilage. Arabidopsis mucilage is composed of two distinct layers, a water-soluble non-adherent outer layer and an adherent inner layer, which differ in their polysaccharide composition and structure [1]. The formation of these layers and the relation and interaction between them are unclear. Some natural mutants of Arabidopsis release water-soluble mucilage while others display only a thin layer of adherent mucilage. Indeed, many proposals were claimed about the biological and ecological functions of this trait. As suggested by several studies, it may have a number of roles, including seed germination, seedling growth and seed dispersal through adhesion to soil or animal vectors [2]. Mucilage of Arabidopsis seeds seems to be influenced by environmental factors as temperature and water availability, as suggested by their natural native distribution and biodiversity from Europe to central Asia [3].

The scarcity of works about the variability in seed mucilage is still large and to the knowledge of the authors, no detailed kinetic imbibition works have been conducted on Arabipodsis seeds. Obtaining this information and providing the necessary methodology for this purpose and, in general, to phenotype water-imbibed seeds is one of the main goal of the present work.

The variability in mucilage can be studied by various techniques but few of them can simultaneously characterize water-soluble and adherent mucilage without hydrolysis or extraction in a non-destructive manner. In this context, time-domain Nuclear Magnetic Resonance (TD-NMR) provides new opportunities for non-destructive investigations of water distribution and transfer in plant tissues in large-scale studies and more particularly for fine phenotyping [4–7]. TD-NMR is based on longitudinal ($T_1$) and transverse ($T_2$) relaxation time

measurements. These parameters that measures the molecular dynamics through the magnetic properties of protons (from hydrogen), give access to molecular information (nature, size, physical state, interaction with other molecules), and to physico-chemical properties of the surrounding matrix (porosity size, mechanical properties, solvent accessibility …) [5] . It has been currently used to study the hydration properties of foodstuffs and became further exploited to characterize the water distribution and evolution of diverse plant tissues with the aim of having a better understanding of their functioning and adaptation to environmental changes [4,5] . Recently, TD-NMR relaxometry have been employed as an excellent concept-proof for non-invasive measurement of water uptake rates into imbibing seeds of a three Arabidopsis accessions [4]. In summary, for wild-type seeds, the water-soluble mucilage was found to retain water, but did not improve imbibition of internal seed tissues, indicating a role in maintaining seeds hydrated by trapping water around the seed. This should slow the rate of seed drying and prolong the imbibed state. More detailed investigations of the variability in mucilage traits have been carried out for characterizing both inner and outer mucilage traits for 19 natural variants identified previously as exhibiting atypical outer mucilage macromolecular properties [8,9]. This phenotyping study have pointed out questions about the complexity and the suitability of the classical signal processing of the TD-NMR data to extract useful relaxation information. This classical signal processing consists in Non-Negative Least Squares (NNLS) fitting procedures [10] or uses a numerical inversion of the Inverse Laplace transform (ILT) [11]. Such signal processing task can be rather fastidious, giving controversial results and concerning ILT, it is known to be an ill-conditioned and ill-posed problem, resulting in a large number of solutions that small noise in the data can easily affect.

In the present work, imbibition process of eighteen natural Arabidopsis variants and one reference wild-type accession has been investigated by TD-NMR in combination with

Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS) [12,13]. This method decomposes the raw mixed measurement into a small bilinear model of pure contributions that can help to interpret variations of mucilage properties in imbibition process of the seeds. Upon applying this technique, the general trend of change in concentration (kinetic profiles) of the individual components as a function of process time can be found. Hybrid soft- and hard-modelling-Multivariate Curve Resolution (HS-MCR) [14], a variant of MCR-ALS that allows for introducing kinetic models to describe the process behaviour, was used to interpret the meaning of the resolved kinetic process profiles and the rate constants related to the process can also be obtained.

Previous studies have shown the capability of the application of both MCR-ALS and HS-MCR to the analysis of spectroscopic data recorded in evolutionary systems [15,16]. However, as far as we know, MCR approaches have not been investigated for kinetic process studies of imbedded-water seeds monitored by TD-NMR. Given the state of the art, the aim of the present study is to provide an additional insight into the mechanisms involved in this kind of processes and illustrate the benefits of applying chemometric methods to analyse raw TD-NMR signals in a phenotyping context.

## 2. Material and Methods

### 2.1 Materials and sample preparation

Samples and NMR methods are described in [8] . Briefly, 19 natural accessions from the cultivation series d, comprising the wild reference Col-0 (identification number:186), were studied for their atypical outer mucilage properties [9]. The mature, ‘‘dry’’Arabidopsis seeds (approximately 8% water content) were first introduced into NMR tubes (10 mm dia.), before imbibing with 150 to 220 µl of deionised water, depending on the accession. Acquisitions of the NMR signal were carried out each 3 min 18 s from 5 min to 24 h for all genotypes. The

relaxation time evolution (time value and amplitude) remained constant after few hours after their water imbibition [8], and hence only the first 128 acquisition data (from 5 min to 7 h) were selected.

ILT on the relaxation decay curves obtained by MCR and HS-MCR analyses (see section 2.2) was performed using the homemade EMILIO-FID© software.

## 2.2 Multivariate curve resolution-alternating least squares (MCR-ALS)

Every *kth* imbibition experiment monitored by TD-RMN gives a data matrix, $\mathbf{D_k}$, which can be described assuming a bilinear model based on the following equation:

$$\mathbf{D_k} = \mathbf{C_k S^T} + \mathbf{E_k} \qquad \text{Equation 1}$$

where the rows of matrix $\mathbf{D_k}$ are the TD-NMR spectra collected at different imbibition times and the columns are the kinetic traces recorded at different relaxation times. $\mathbf{C_k}$ is the matrix of the concentration profiles of the compounds involved in the *kth* kinetic process and $\mathbf{S^T}$ is the matrix of their related pure TD-NMR spectra. Finally, $\mathbf{E_k}$ is the residual matrix with the unmodelled data variance.

MCR-ALS aims at resolving the underlying bilinear model (see equation 1) by using the sole information contained in the raw data set $\mathbf{D_k}$ (see Figure 1a). For this purpose, an alternating least squares procedure under constraints is used to iteratively optimize both $\mathbf{C_k}$ and $\mathbf{S^T}$ matrices until the error in the reproduction of the original data set ($\mathbf{D_k}$) is minimized as much as possible [12,13]. In order to initialize the iterative procedure, previous determination of number of contributions in the raw data set and the generation of $\mathbf{C_k}$ or $\mathbf{S^T}$ estimates are required. In this work, the determination of number of contributions were performed by Singular Value Decomposition (SVD) [17], knowing that singular values associated with chemical compounds are larger than singular values related to noise and to experimental error.

Initial estimates of $\mathbf{S^T}$ were obtained using SIMPLISMA which is a technique based on the detection of "purest" variables [18].

MCR solutions are not unique. Different combination of $\mathbf{C_k}$ and $\mathbf{S^T}$ can generate the mathematical solution to Equation 1.Therefore, in order to limit these number of possible solutions (rotational ambiguities) and provide meaningful shapes for the profiles in $\mathbf{C_k}$ and $\mathbf{S^T}$, some constraints during the ALS-optimization procedure must be imposed [13,19]. Constraints are chemical or mathematical properties that the profiles in $\mathbf{C_k}$ and/or $\mathbf{S^T}$ must fulfill. The constraints used in this study are non-negativity (both kinetic profiles and pure spectra of the resolved components must be positive), unimodality (kinetic profiles in imbibition processes present only one maximum per profile). Normalization of pure spectra is also used to avoid scaling fluctuations in the profiles during the optimization. Hard-modeling was also used as an additional constraint by imposing a kinetic model into the resolution, forcing the concentration profiles to obey the shapes described by this particular kinetic model. The introduction of this constraint gives rise to the hybrid soft- and hard-modeling variant of MCR (HS-MCR) that reduces the rotational ambiguity associated with the kinetic profiles obtained using exclusively soft-modelling constraints. Moreover, HS-MCR provides the kinetic rate constants as additional information to the resolution [14,20].

A great advantage of MCR-ALS is the possibility to analyse simultaneously kinetic experiments in a single multiset structure to provide more reliable results, less affected by ambiguity phenomenon [12,13,19]. In a phenotyping context, this means that resolved features would define much better general genotype traits analysed together than if they were analysed individually. In this study, the multiset structure was obtained by setting all data matrices $\mathbf{D_k}$ one on top of each other keeping the common relaxation time in the same column. The bilinear model in Equation 1 is now extended to the augmented data set as shown in Equation 2:

$\mathbf{D_{aug}=[D_1;D_2;...;D_n]=[C_1;C_2;...;C_n]S^T+[E_1;E_2;...;En]=C_{aug}S^T+E_{aug}}$ Equation 2

where $\mathbf{C_{aug}}$ is a column-wise augmented matrix formed by the $\mathbf{C_k}$ submatrices that contain the resolved kinetic profiles in the different imbibition experiments, and $\mathbf{S^T}$ is a single data matrix of pure TD-RMN spectra, assumed to be common and valid for all experiments (See Figure 1b).

The percentage of lack of fit (% LOF) and the explained variance (% $r^2$) are used to indicate the fit quality of the MCR-ALS results. These parameters are calculated according to the following expressions:

$$\% \text{ LOF} = 100 \times \sqrt{\frac{\Sigma e_{ij}^2}{\Sigma d_{ij}^2}} \quad \text{Equation 3}$$

$$\% \ r^2 = 100 \times \left(1 - \frac{\Sigma e_{ij}^2}{\Sigma d_{ij}^2}\right) \quad \text{Equation 4}$$

where $e_{ij}$ is equal to $d_{ij}$-$d_{ij}^*$, $d_{ij}^*$ are the values of the data set reproduced by the bilinear model and $d_{ij}$ the original values in the raw data set $\mathbf{D_k}$. In order to consider that the resolution results of an analysis are good, the variance explained must be sufficiently high and the concentration profiles and spectra obtained must be chemically meaningful and show shapes consistent with the variation in the raw data sets.

MCR-ALS results, i.e., kinetic profiles (**C** matrix) and pure spectra ($\mathbf{S^T}$ matrix) may be further used and processed to obtain additional information. In this work, inversion of the Laplace transform [21] was applied on the pure spectra in order to interpret the meaning of the resolved profiles and understand the water mobility during seed imbibition.

More details about the MCR-ALS method are given in Refs. [12,13,22] and a GUI to use the algorithm is freely available at http://mcrals.info.

**[Insert Figure 1]**

## 3. RESULTS AND DISCUSSION

This section shows the results obtained from individual and simultaneous analysis of imbibition experiments and from soft-modelling and hybrid hard and soft-modelling methodologies. A comprehensive description of the imbibition process of all Arabidopsis genotypes studied, taking into account literature interpretation, is also provided.

### 3.1 Individual MCR analysis

An exploratory analysis of the imbibition evolution by means of classical MCR-ALS has been carried out. As a way of example, Figure 2a shows the MCR-ALS resolved concentration profiles (**C**), the pure signals ($\mathbf{S^T}$) obtained for the wild genotype 186 over imbibition time. Resolution of three species was achieved with the lack of fit (LOF %) equal to 0.97% and the explained variance ($r^2$ %) equal to 99.9 %. The inclusion of a different number of species gave solutions worse mathematically or unreliable spectra or concentration profiles. The blue, green and red curves correspond respectively to short, transitional and long imbibition times. The shape of the resolved concentration profiles suggested a kinetic model based on two consecutive first-order reactions ( $A \xrightarrow{k1} B \xrightarrow{k2} C$ ). Therefore, HS-MCR was applied to the data and the postulated model was introduced as a hard-modeling constraint. Concentration profiles (**C**), the pure signals ($\mathbf{S^T}$) obtained from HS-MCR analysis resemble in shape to soft-modelling results (see Figure 2b) and similar good data fit was achieved (LOF %= 0.98 and $r^2$ %=99.9), indicating that the hypothesis of the selected kinetic model was correct. Moreover,

imbibition rate constants were obtained as additional information ($k_1$=5.01 x$10^{-4}$ $ms^{-1}$ and $k_2$=1.57 x$10^{-4}$ $ms^{-1}$).

**[Insert Figure 2]**

As mentioned before, classical signal processing in TD-NMR, based frequently on ILT, can be rather tedious and is known to be an ill-conditioned and ill-posed problem, resulting in a large number of solutions due to the fact that small noise in the data can easily affect. HS-MCR provides meaningful and noised-filtered spectra, and hence, the application of the classical ILT on these relaxation decay curves could be carried on in order to interpret the meaning of the resolved profiles.

Figure 3 shows the results of applying ILT on the resolved HS-MCR spectra of the Figure 2b. In conformity with a previous work [4], five $T_2$ components were identified in imbibed seeds of the wild genotype 186. These components could be assigned to protons of water populations with different mobility and ratio. The first and second components, $T_2$(1) and $T_2$(2a,b), could be assigned to the solid phase of seeds (polysaccharides and proteins) and water protons in exchange with the hydroxyl groups found in these macromolecules, respectively. The $T_2$(3) and $T_2$(4) components could be assigned to oil with a superposed contribution from water as intracellular water and as in adherent mucilage, respectively. Finally, the fifth component, $T_2$(5), could correspond to water around the seeds in soluble mucilage.

Taking into account the potential meaning of these $T_2$ components, the first ILT spectrum (blue) characterized by a longer $T_2$(5), is characteristic of initial hydration of seeds resulting in the water uptake of the outer mucilage ($T_2$(5)) while the dry matter of seeds (characterized by T2(1)) remains quite intact. The second ILT spectrum (green) characterized by a high $T_2$ between the $T_2$(4) and $T_2$(5) of the blue spectrum and a second large $T_2$ peak between $T_2$(2b)

and $T_2(3)$, was understood as the state when water starts to move into internal tissues and in the adherent mucilage layer. Finally, the third spectrum (red) with well resolved $T_2(3)$ and $T_2(2a, b)$ was interpreted as the state when the internal tissues are well hydrated.

**[Insert Figure 3]**

HS-MCR imbibition kinetic profiles of the wild genotype 186 (figure 2) shows that the decay of the first imbibition profile (blue) is relatively fast (12007 s.), which means a rapid water uptake by the outer mucilage polysaccharides. An intermediate state (green), with fast emergence followed by a slowly decay evolution, is detected, which indicates a fast water uptake of the adherent mucilage layer over the first minutes and then a slow diminution of this water pool because of its transfer towards the internal seed tissues. Finally, the third profile (red) related to the water in internal seed tissues is increasing until the end of the process, at the expense of the outer and inner mucilage layers.

Regarding the imbibition rate constants, it could be seen that for the wild genotype 186, absorption of water in the inner mucilage layer is slightly faster than the hydration of the internal seed tissues ($k_1 > k_2$).

### 3.2. Simultaneous MCR analysis

An advantage of analyzing several experiments simultaneously is that the results are more reliable and less affected by ambiguity phenomenon. Moreover, the extracted features using a single multiset structure are common to all experiments. Thus, we can expect to extract more information about the Arabidopsis trait variability. To extend MCR-ALS to the simultaneous analysis of the imbibition experiments, it is required that the pure spectra of all species involved in the process do not change among experiments (linked to bilinear model shown in Figure 1b). No significant differences between resolved pure spectra of all 19 Arabidopsis

seeds were found (Figure not shown) with a Pearson correlation coefficient higher than 0.90 and, hence, performance of the simultaneous treatment of all 19 imbibition experiments is suggested.

Firstly, simultaneous analysis by classical soft-modelling MCR-ALS has been carried out (Figure not shown). A good description of the data set was obtained (LOF % ≤ 1,27 and $r^2$ %= 99.9). As in individual analysis, resolution of three species is achieved and suggested again the underlying kinetic model ( $A \overset{k1}{\rightarrow} B \overset{k2}{\rightarrow} C$ ). Therefore, simultaneous analysis of the different imbibition experiments by HS-MCR was performed, fitting each experiment to the mentioned kinetic model. Resolved spectral profiles from the multi-experiment MCR-ALS analysis were used as initial estimates. Figure 4 shows the resolved augmented concentration matrix ($\mathbf{C_{aug}}$) and the related pure signals ($\mathbf{S^T}$) of the simultaneous analysis. Application of ILT on the related pure signals ($\mathbf{S^T}$) of the simultaneous analysis was also carried out (see Figure 4c). The resolved concentration profiles, the pure signals and the ILT of the pure signals are consistent with the reference kinetic model and with the pure signals obtained in both individual and simultaneous analysis. A rather similar fit to simultaneous soft-modelling MCR-ALS was obtained (LOF % = 3.66 and r2 % = 99.6), strongly supporting the previously proposed kinetic model ( $A \overset{k1}{\rightarrow} B \overset{k2}{\rightarrow} C$ ). Moreover, imbibition rate constants, $k_1$ and $k_2$ (see table S1 in supplementary information), for each genotype were obtained as additional information.

**[Insert Figure 4]**

In order to better evaluate the imbibition trends of all genotypes, a log-transformation of obtained $k_1$ and $k_2$ was carried out. Figure 5 displays the 19 genotypes according to their values of log $k_1$ and log $k_2$. Genotypes 13, 167 and 517 (in red square) seem to follow

imbibition kinetics similar to the wild genotype 186, where absorption of imbibition water in the adherent mucilage layer (log $k_1$) is slightly faster than the hydration of the internal seed tissues (log $k_2$). HS-MCR imbibition kinetic profiles of these genotypes (see figure 4a) show also their kinetics similarity with the wild genotype 186. Genotypes 254, 257, 259, 335 and 472 (in black circle) absorb imbibition water rapidly (higher log $k_1$) but slow down the progression of water through the internal tissues (lower log $k_2$), especially the genotype 259 which presents the highest log $k_1$. Their kinetic profiles in Figure 4a present faster decay of the first profile (blue) and faster emergence of the intermediate state (green) than the genotype 186. This difference of imbibition kinetics between 254, 257 and 259 genotypes compared to the wild genotype 186 could be due to their lower content of soluble mucilage that facilitate the hydration of the adherent layer of mucilage [8,23]. On the other hand 335 and 472 genotypes show a higher soluble mucilage content but with lower intrinsic viscosity relative to 189. As mentioned before, the first ILT spectra (see blue spectra in Figure 4c), with a longer $T_2(5)$ component, should be interpreted as water uptake by the outer soluble mucilage polysaccharides. Previous results obtained on 189 (Col-0) and two natural mutants (Mum2 and Myb5) have shown that the outer soluble mucilage did not improve imbibition of internal seed tissues and, on the contrary, tent to retain water around the seeds [4]. Therefore, it is not surprising that genotypes that contain less soluble mucilage than the wild type present faster decay of this first profile (see blue profiles in Figure 4a) and, hence, higher log $k_1$. In case of 335 and 472 genotypes that release a higher outer mucilage content, the water uptake kinetic seems to be influenced by other physiological and/or physico-chemical characteristics. It may be due to the difference in the outer mucilage intrinsic viscosity instead of its quantity [23].

For genotypes 166, 301, 397 and 549 (in blue square) the absorption of imbibition water into the adherent layer of mucilage (log $k_1$) is slower than the hydration of the internal seed tissues (log $k_2$), especially in the genotype 301. The kinetics profiles shows in Figure 4a slower decay

of the first profile (blue) and almost negligible occurrence of intermediate species (green) in comparison with the wild genotype. These genotypes are also characterized by less soluble mucilage but with comparable or a higher adherent mucilage content compared to 189 [8]. The higher value of log $k_2$ (k2 is multiplied by 5 to 20 compared to 189) of these genotypes is supposed to be due to higher adherent mucilage content. It could be suggested that the adherent mucilage speeds up the transfer of water into internal tissues for 166, 301, 397 and 549 seeds.

In genotypes 77, 136, 178, 258, 261 and 456 (in green square) the water uptake by mucilage polysaccharides (log $k_1$) and water transfer to internal tissues of seeds (log $k_2$) present similar speed. Their kinetic profiles in Figure 4a shows slower decay of the first profile (blue), and slower emergence and decay evolution of the intermediate species (green) than the genotype 186 (smaller log $k_1$ and log $k_2$). Except for 178 and 258, this low kinetics might be correlated to a lower adherent mucilage content [8]. Compared to 189, the content of soluble mucilage varies but it is generally also less important (except for 258 and 456). .

Despite the exceptions mentioned above, it can be hypothesized that the adherent mucilage has a different role compared to the soluble mucilage in the imbibition processes. Soluble mucilage seems to trap water around the seed to slow down its hydration [8], whereas the adherent mucilage appears to accelerate the hydration of the internal tissues. However, to support this hypothesis and provide better insights in these kinetic imbibition processes, the results from MCR should be interpreted with more precise composition and physicochemical properties of the two mucilage layers. Indeed, the interpretation of the seed imbibition behaviour cannot only be interpreted relative to the polysaccharide content into the mucilage layers. The results on 335 and 472 genotypes indicates that the intrinsic viscosity of the outer mucilage could also affect the water ingress into seeds. The polysaccharide assemblies and

the global mucilage porosity should also influence the water accessibility to the internal tissues of seeds.

In future works, the incorporation of these physicochemical properties and the MCR results of more replicates of the studied imbibition experiments is planned to allow for a more comprehensive and reliable study of the variability in seed mucilage.

## 4. Conclusions

Results obtained in this work have shown the power of combining TD-NMR whith multivariate curve resolution methods to study the kinetic processes of imbibed-water seeds. In particular, in this work, the application of MCR-ALS and of HS-MCR-ALS methods on TD-NMR monitoring experiments has given a good description of the kinetic pathway involved in the imbibition process of the nineteen genotypes of Arabidopsis seeds. By using the resolution power of these data analysis methods when they are applied to the simultaneous analysis of all investigated genotypes, it was possible to estimate common features to all experiments and to determine their kinetic profiles and rate constants. It is concluded that the imbibition process of all investigated Arabidopsis seeds could be described with a kinetic model based on two consecutive first-order reactions related to an initial water absorption around the seed and into the adherent mucilage layer and a posteriori hydration of the internal seed tissues, respectively. Calculated rate constants seemed to depend on both soluble and adherent mucilage content. However, physico-chemical properties and organisation of the mucilage polysaccharides should also affect the water uptake kinetics.

The methodology proposed here is worth considering as an alternative strategy for modelling kinetic profiles obtained by TD-NMR data. This methodology will permit widening the studies of the variability seed mucilage in phenotyping context dealing with genetic diversity and the biological function of mucilage as an adaptive trait. It offers the opportunity for much

wider applications for initial screening of plants based on their water uptake property in connection with different physiological and/or metabolic criteria.

## 5. Acknowledgements

This work was supported by the Agence Nationale de Recherche program (grant number: ANR-14-CE19-0001). The IJPB benefits from the support of Saclay Plant Sciences-SPS (ANR-17-EUR-0007). The authors kindly acknowledge H. M North (IJPB-INRAE-AgroParisTech Versailles) and M.-C. Ralet (INRAE Nantes) for having accepted the use of Arabidopsis data before the publication of the Cambert et al. (2021) data-paper. The NMR experiments were performed using the NMR facilities of the AgroScans Platform (Rennes, France).

## 6. References

[1] A. Macquet, M.C. Ralet, J. Kronenberger, A. Marion-Poll, H.M. North, In situ, chemical and macromolecular study of the composition of Arabidopsis thaliana seed coat mucilage, Plant Cell Physiol. (2007). doi:10.1093/pcp/pcm068.

[2] X. Yang, J.M. Baskin, C.C. Baskin, Z. Huang, More than just a coating: Ecological importance, taxonomic occurrence and phylogenetic relationships of seed coat mucilage, Perspect. Plant Ecol. Evol. Syst. 14 (2012) 434–442. doi:10.1016/j.ppees.2012.09.002.

[3] M.A. Koch, The plant model system Arabidopsis set in an evolutionary, systematic, and spatio-temporal context, J. Exp. Bot. (2019). doi:10.1093/jxb/ery340.

[4] S. Saez-Aguayo, C. Rondeau-Mouro, A. Macquet, I. Kronholm, M.C. Ralet, A. Berger, C. Sallé, D. Poulain, F. Granier, L. Botran, O. Loudet, J. de Meaux, A. Marion-Poll, H.M. North, Local Evolution of Seed Flotation in Arabidopsis, PLoS Genet. 10 (2014) 13–15. doi:10.1371/journal.pgen.1004221.

[5] L.A. Colnago, Z. Wiesman, G. Pages, M. Musse, T. Monaretto, C.W. Windt, C. Rondeau-Mouro, Low field, time domain NMR in the agriculture and agrifood sectors: An overview of applications in plants, foods and biofuels, J. Magn. Reson. (2021). doi:10.1016/j.jmr.2020.106899.

[6] L.A. Colnago, M. Engelsberg, A.A. Souza, L.L. Barbosa, High-throughput, non-destructive determination of oil content in intact seeds by continuous wave-free precession NMR, Anal. Chem. 79 (2007) 1271–1274. doi:10.1021/ac062091+.

[7] R.A. Prestes, L.A. Colnago, L.A. Forato, L. Vizzotto, E.H. Novotny, E. Carrilho, A rapid and automated low resolution NMR method to analyze oil quality in intact oilseeds, Anal. Chim. Acta. 596 (2007) 325–329. doi:https://doi.org/10.1016/j.aca.2007.06.022.

[8] M. Cambert, A. Berger, C. Sallé, S. Esling, D. Charif, T. Cadoret, M. Ralet, H.M. North, C. Rondeau-Mouro, Datasets of seed mucilage traits for Arabidopsis thaliana

natural accessions with atypical outer mucilage, Sci. Data. 8 (2021) 79. doi:10.1038/s41597-021-00857-3.
[9] I. Fabrissin, G. Cueff, A. Berger, F. Granier, C. Sallé, D. Poulain, M.C. Ralet, H.M. North, Natural variation reveals a key role for rhamnogalacturonan I in seed outer mucilage and underlying Genes, Plant Physiol. (2019). doi:10.1104/pp.19.00763.
[10] R.F. Ling, C.L. Lawson, R.J. Hanson, Solving Least Squares Problems., J. Am. Stat. Assoc. (1977). doi:10.2307/2286501.
[11] R.L. Parker, Y.Q. Song, Assigning uncertainties in the inversion of NMR relaxation data, J. Magn. Reson. (2005). doi:10.1016/j.jmr.2005.03.002.
[12] R. Tauler, Multivariate curve resolution applied to second order data, Chemom. Intell. Lab. Syst. 30 (1995) 133–146. doi:10.1016/0169-7439(95)00047-X.
[13] R. de Juan, Anna; Rutan, S. and Tauler, Two-Way Data Analysis: Multivariate Curve Resolution – Iterative Resolution Methods, in: B. Brown, S. D. ; Tauler, R. and Walczak (Ed.), Compr. Chemom., Elsevier, 2010: pp. 325–344.
[14] A. De Juan, M. Maeder, M. Martínez, R. Tauler, Combining hard- and soft-modelling to solve kinetic problems, Chemom. Intell. Lab. Syst. (2000). doi:10.1016/S0169-7439(00)00112-X.
[15] S. Mas, A. Carbó, S. Lacorte, A. De Juan, R. Tauler, Comprehensive description of the photodegradation of bromophenols using chromatographic monitoring and chemometric tools, Talanta. 83 (2011) 1134–1146. doi:10.1016/j.talanta.2010.06.042.
[16] A. Jayaraman, S. Mas, R. Tauler, A. de Juan, Study of the photodegradation of 2-bromophenol under UV and sunlight by spectroscopic, chromatographic and chemometric techniques, J. Chromatogr. B Anal. Technol. Biomed. Life Sci. (2012). doi:10.1016/j.jchromb.2012.03.038.
[17] N. Lord, G.H. Golub, C.F. Van Loan, Matrix Computations, Math. Gaz. (1999). doi:10.2307/3621013.
[18] W. Windig, C.E. Heckler, F.A. Agblevor, R.J. Evans, Self-modeling mixture analysis of categorized pyrolysis mass spectral data with the SIMPLISMA approach, Chemom. Intell. Lab. Syst. (1992). doi:10.1016/0169-7439(92)80104-C.
[19] R. Tauler, A. Smilde, B. Kowalski, Selectivity, local rank, three-way data analysis and ambiguity in multivariate curve resolution, J. Chemom. 9 (1995) 31–58. doi:10.1002/cem.1180090105.
[20] M. Maeder, A.D. Zuberbühler, Nonlinear Least-Squares Fitting of Multivariate Absorption Data, Anal. Chem. (1990). doi:10.1021/ac00219a013.
[21] Maximum Entropy and Bayesian Methods, 1989. doi:10.1007/978-94-015-7860-8.
[22] J. Jaumot, A. de Juan, R. Tauler, MCR-ALS GUI 2.0: New features and applications, Chemom. Intell. Lab. Syst. 140 (2015) 1–12. doi:10.1016/j.chemolab.2014.10.003.
[23] D. Poulain, L. Botran, H.M. North, M.C. Ralet, Composition and physicochemical properties of outer mucilage from seeds of Arabidopsis natural accessions, AoB Plants. (2019). doi:10.1093/aobpla/plz031.

**Figure Captions**

**Figure 1.** (a) Multivariate curve resolution-alternating least squares (MCR-ALS) resolution applied to a single data matrix **Dk** (b) MCR-ALS resolution applied to the column-wise augmented data matrix **Daug**.

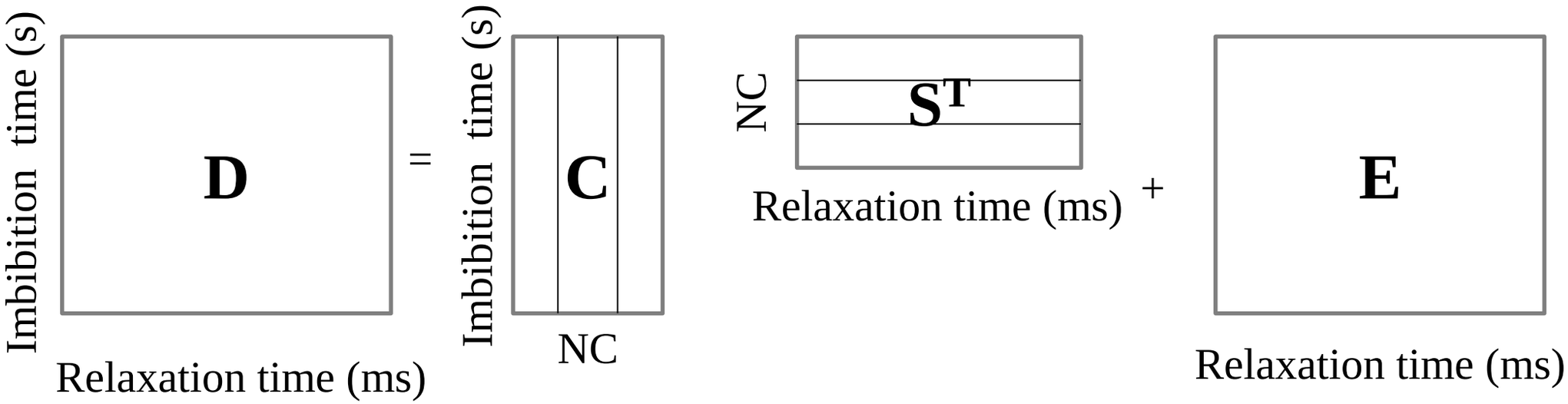
**Figure 2**. MCR results from imbibition measurements of the wild genotype 186: (a) classical MCR-ALS resolved kinetic profiles and related pure spectra. (b) HS-MCR resolved kinetic profiles and related pure spectra.

**Figure 3.** $T_2$ distributions obtained by ILT of the pure spectra of the classical HS-MCR-ALS resolved kinetic profiles for the wild genotype 186.

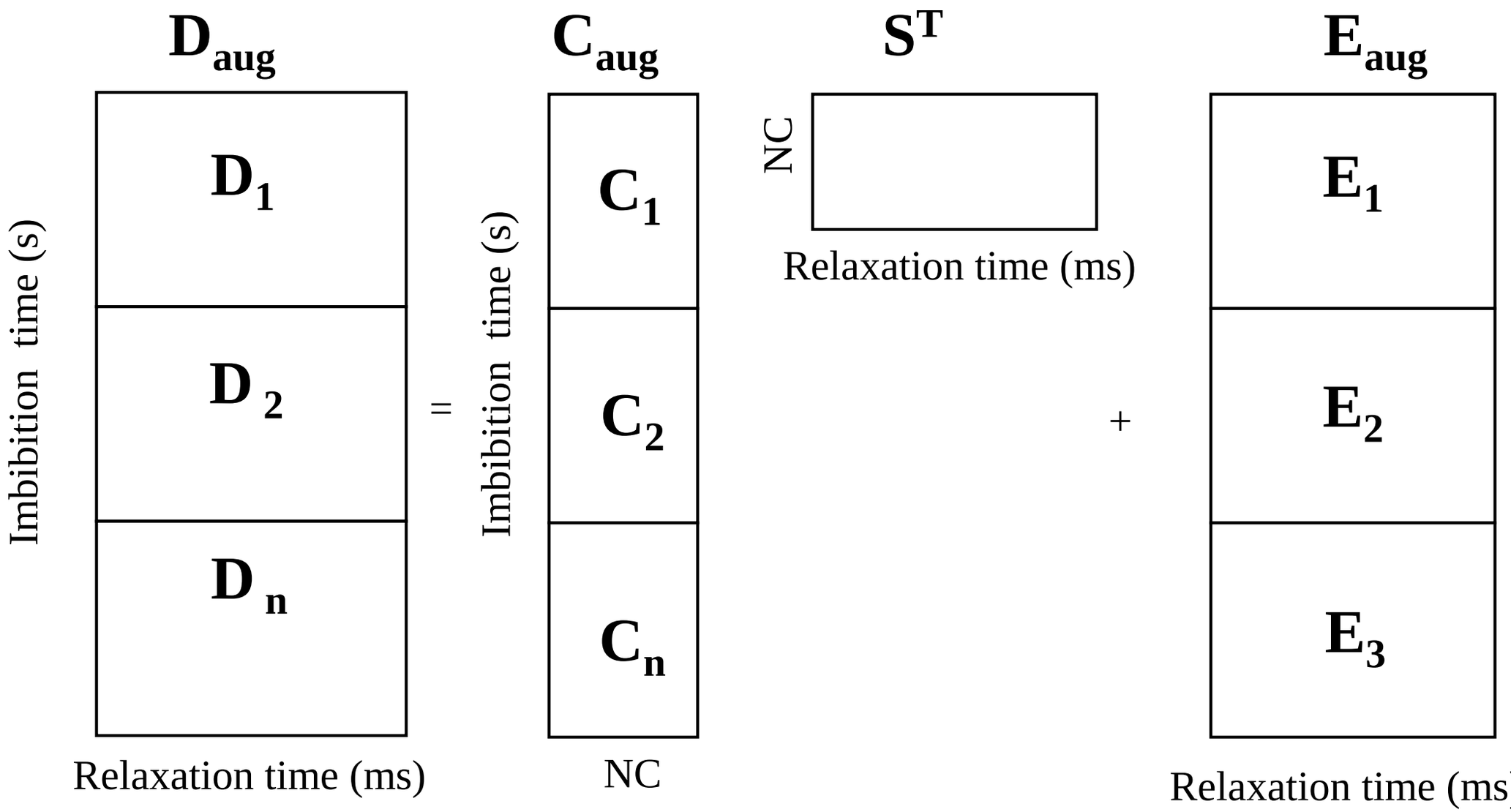
**Figure 4**. HS-MCR results for the multiset analysis of 19 genotypes: a) kinetic profiles of each genotype from the augmented concentration matrix ($\mathbf{C_{aug}}$), b) related pure signals, c) the ILT processing of pure signals.

**Figure 5.** Scatter plot of the 19 genotypes with regard to the kinetic constants log $k_1$ and log $k_2$. The coloured squares and circle correspond to the groups of genotypes with similar imbibition kinetics.

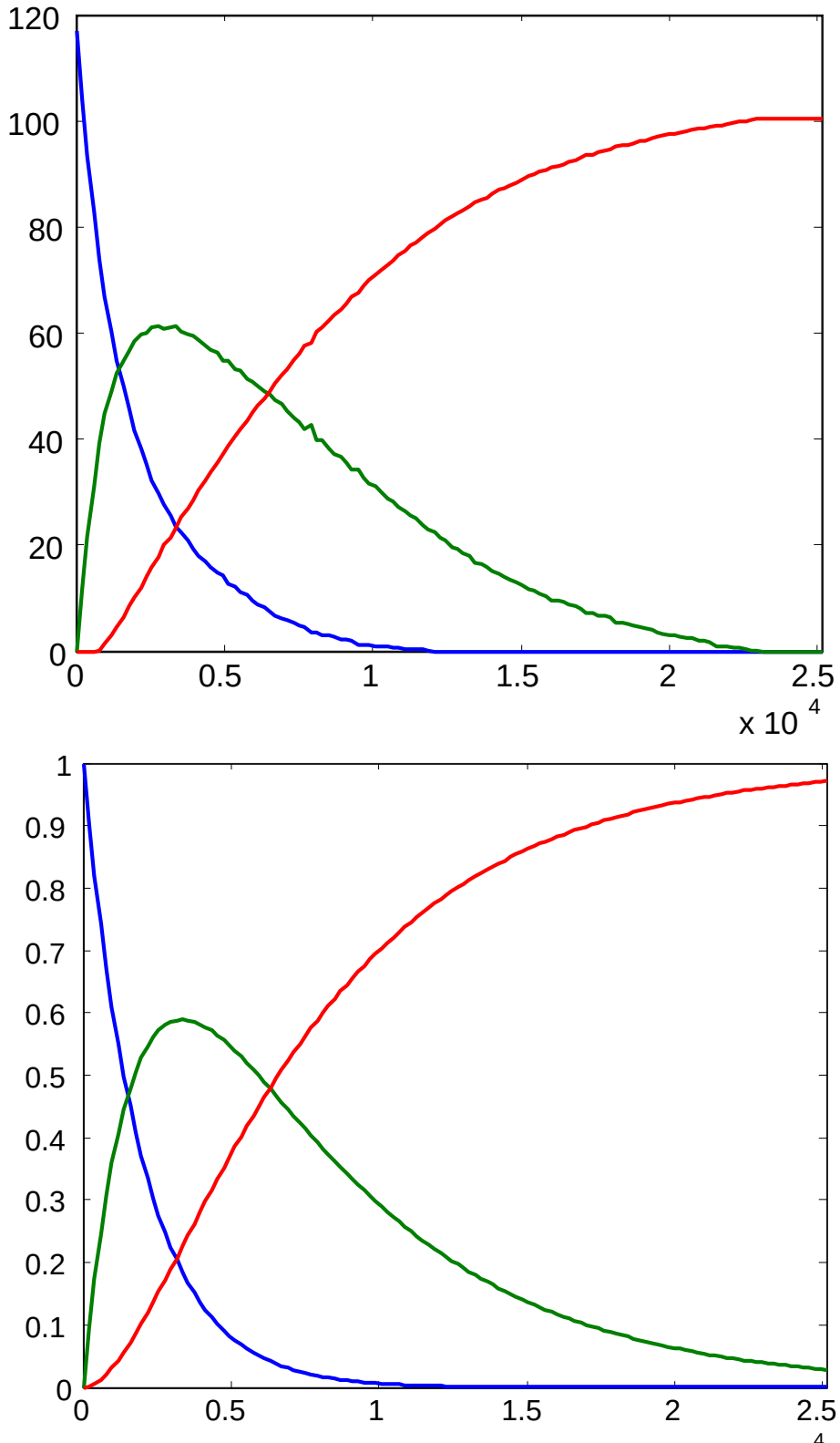
**(a)**

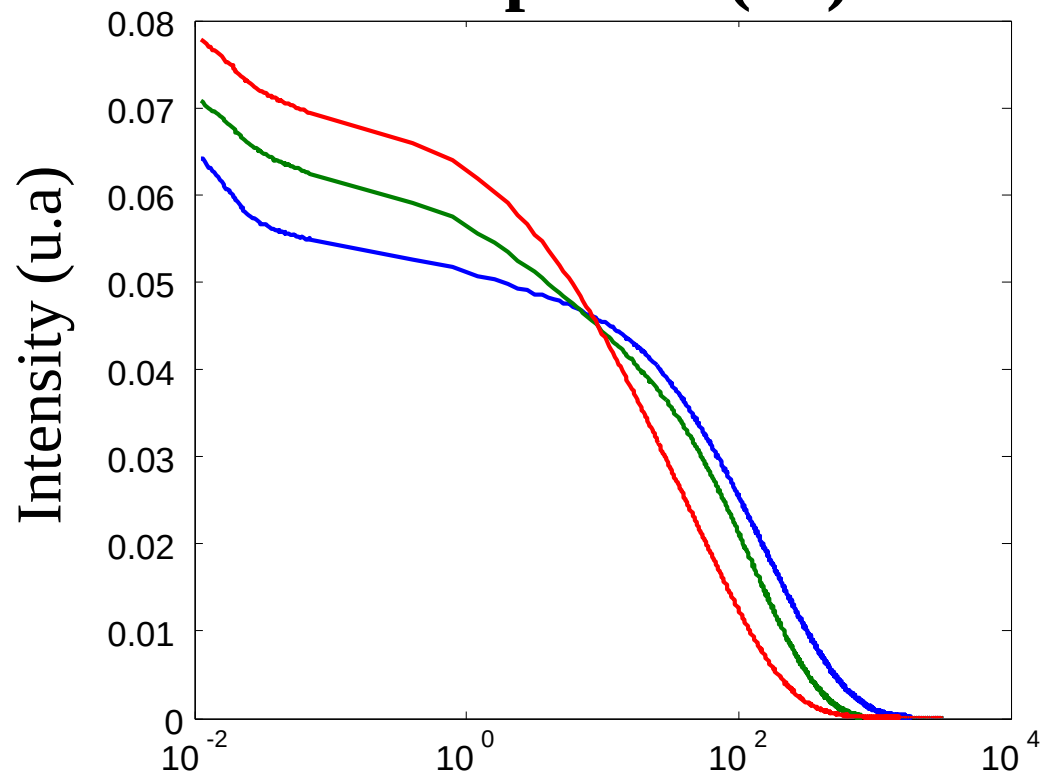

Imbibition time (s)
D
Relaxation time (ms)
=
Imbibition time (s)
C
NC
NC
S^T
Relaxation time (ms)
+
E
Relaxation time (ms)

**(b)**


D_aug
D_1
D_2
D_n
Imbibition time (s)
Relaxation time (ms)
=
C_aug
C_1
C_2
C_n
Imbibition time (s)
NC
S^T
NC
Relaxation time (ms)
+
E_aug
E_1
E_2
E_3
Relaxation time (ms)

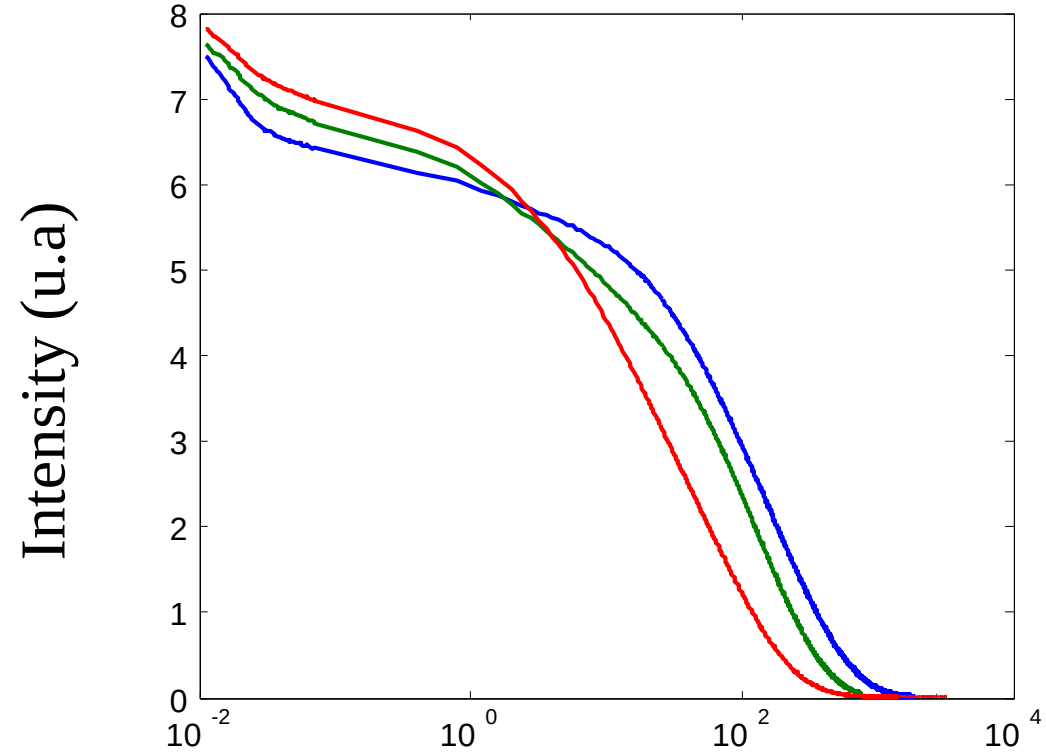
(a)

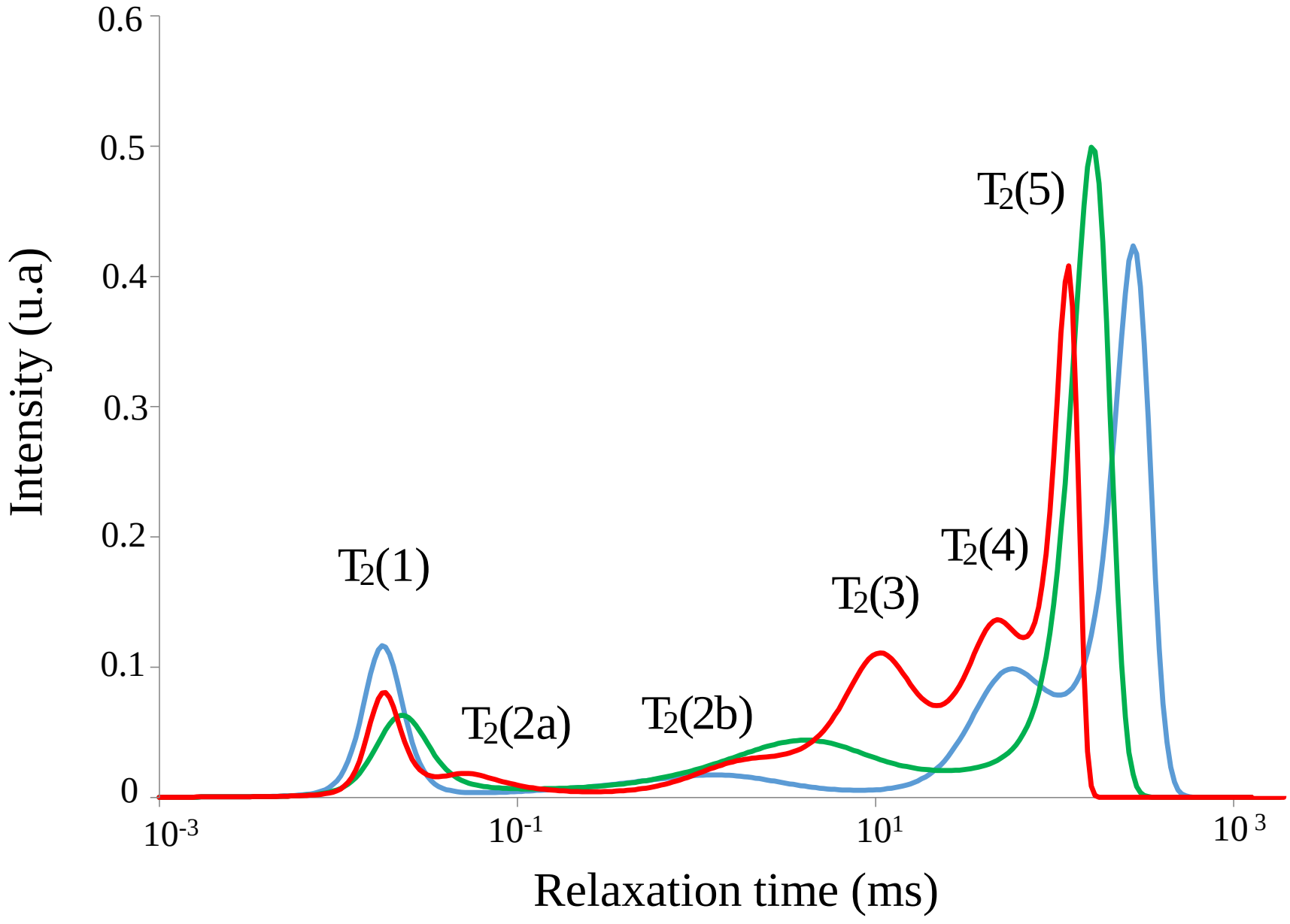

Kinetic profiles (C)
120
100
80
60
40
20
0
0
0.5
1
1.5
2
2.5
x 10 4

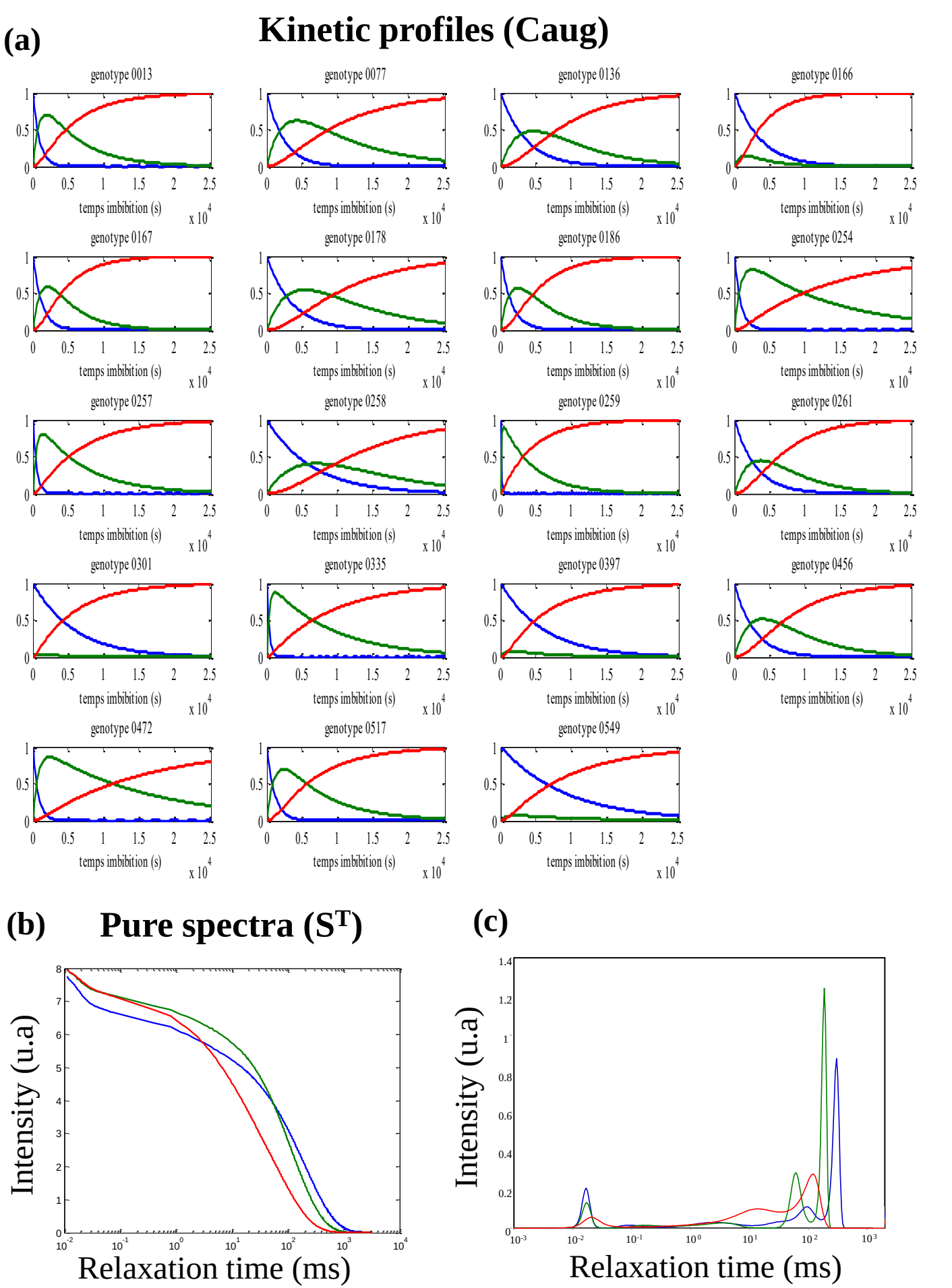

Pure spectra (S^T)
Intensity (u.a)
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
10 -2
10 0
10 2
10 4

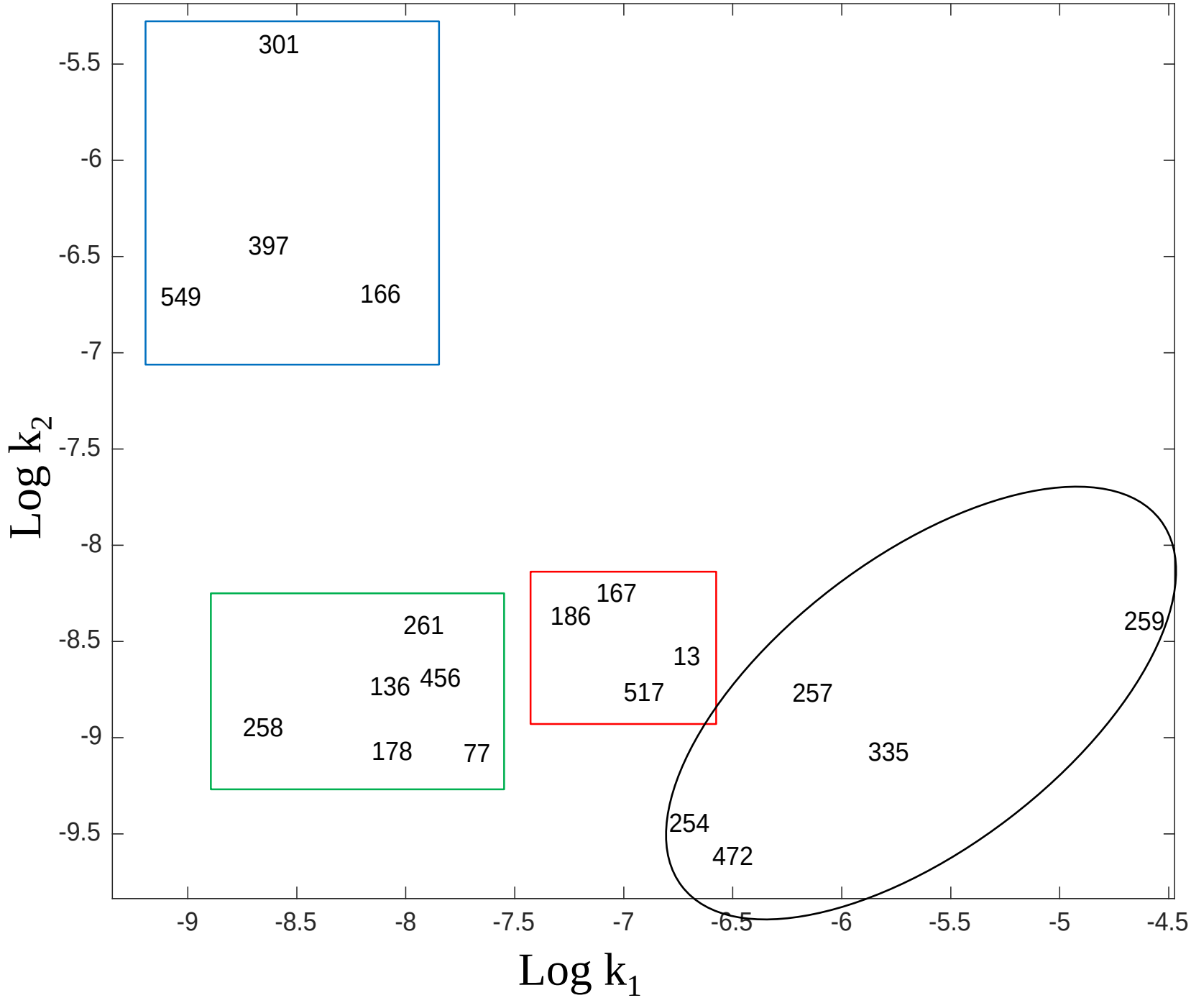
(b)


1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.5
1
1.5
2
2.5
x 10 4
Imbibition time (s)

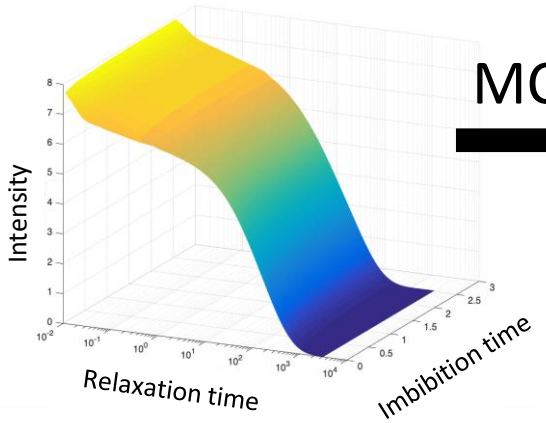

Intensity (u.a)
8
7
6
5
4
3
2
1
0
10 -2
10 0
10 2
10 4
Relaxation time (ms)


0.6
0.5
0.4
0.3
0.2
0.1
0
Intensity (u.a)
T2(1)
T2(2a)
T2(2b)
T2(3)
T2(4)
T2(5)
10-3
10-1
101
10 3
Relaxation time (ms)


(a)
Kinetic profiles (Caug)
genotype 0013
genotype 0077
genotype 0136
genotype 0166
genotype 0167
genotype 0178
genotype 0186
genotype 0254
genotype 0257
genotype 0258
genotype 0259
genotype 0261
genotype 0301
genotype 0335
genotype 0397
genotype 0456
genotype 0472
genotype 0517
genotype 0549
temps imbibition (s)
x 10^4
(b)
Pure spectra (S^T)
Intensity (u.a)
Relaxation time (ms)
(c)
Intensity (u.a)
Relaxation time (ms)


301
397
549
166
261
136
456
258
178
77
167
186
13
517
259
257
335
254
472
-5.5
-6
-6.5
-7
-7.5
-8
-8.5
-9
-9.5
-9
-8.5
-8
-7.5
-7
-6.5
-6
-5.5
-5
-4.5
Log k2
Log k1

TD-NMR signal




MCR-ALS

$A \xrightarrow{k_1} B \xrightarrow{k_2} C$