



**HAL**  
open science

# 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

## ► To cite this version:

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. *Talanta*, 2021, 233, pp.122525. 10.1016/j.talanta.2021.122525 . hal-03245219

**HAL Id: hal-03245219**

**<https://hal.inrae.fr/hal-03245219>**

Submitted on 13 Jun 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

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

3

4 Silvia Mas Garcia<sup>1,2\*</sup>, Jean-Michel Roger<sup>1,2</sup>, Mireille Cambert<sup>3</sup>, Corinne Rondeau-Mouro<sup>3</sup>

5

6 <sup>1</sup> ITAP, INRAE, Institut Agro, University Montpellier, 34196 Montpellier, France

7 <sup>2</sup>ChemHouse Research Group, 34196 Montpellier, France

8 <sup>3</sup>INRAE, UR1466 OPAALE, 17 Avenue de Cucillé, CS 64427, F-35044 Rennes, France

9

10

11

12 \* Author for correspondence:

13 E-mail: [silvia.mas-garcia@inrae.fr](mailto:silvia.mas-garcia@inrae.fr)

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

1 **Abstract**

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

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

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

19

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

23

24

## 1 **1. Introduction**

2 *Arabidopsis* (*Arabidopsis thaliana*) seeds belong to species known to release a gel from the  
3 seed coat on imbibition in water. This gel, made of polysaccharides, forms a sticky, gelatinous  
4 halo around the seed termed mucilage. *Arabidopsis* mucilage is composed of two distinct  
5 layers, a water-soluble non-adherent outer layer and an adherent inner layer, which differ in  
6 their polysaccharide composition and structure [1]. The formation of these layers and the  
7 relation and interaction between them are unclear. Some natural mutants of *Arabidopsis*  
8 release water-soluble mucilage while others display only a thin layer of adherent mucilage.  
9 Indeed, many proposals were claimed about the biological and ecological functions of this  
10 trait. As suggested by several studies, it may have a number of roles, including seed  
11 germination, seedling growth and seed dispersal through adhesion to soil or animal vectors  
12 [2]. Mucilage of *Arabidopsis* seeds seems to be influenced by environmental factors as  
13 temperature and water availability, as suggested by their natural native distribution and  
14 biodiversity from Europe to central Asia [3].

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

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

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

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

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

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

## 17 **2. Material and Methods**

### 18 **2.1 Materials and sample preparation**

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

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

4 ILT on the relaxation decay curves obtained by MCR and HS-MCR analyses (see section 2.2)  
5 was performed using the homemade EMILIO-FID<sup>®</sup> software.

## 6 **2.2 Multivariate curve resolution-alternating least squares (MCR-ALS)**

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

$$9 \quad \mathbf{D}_k = \mathbf{C}_k \mathbf{S}^T + \mathbf{E}_k \quad \text{Equation 1}$$

10 where the rows of matrix  $\mathbf{D}_k$  are the TD-NMR spectra collected at different imbibition times  
11 and the columns are the kinetic traces recorded at different relaxation times.  $\mathbf{C}_k$  is the matrix  
12 of the concentration profiles of the compounds involved in the *k*th kinetic process and  $\mathbf{S}^T$  is  
13 the matrix of their related pure TD-NMR spectra. Finally,  $\mathbf{E}_k$  is the residual matrix with the  
14 unmodelled data variance.

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

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

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

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



$$1 \quad \mathbf{D}_{\text{aug}}=[\mathbf{D}_1;\mathbf{D}_2;\dots;\mathbf{D}_n]=[\mathbf{C}_1;\mathbf{C}_2;\dots;\mathbf{C}_n]\mathbf{S}^T+[\mathbf{E}_1;\mathbf{E}_2;\dots;\mathbf{E}_n]=\mathbf{C}_{\text{aug}}\mathbf{S}^T+\mathbf{E}_{\text{aug}} \quad \text{Equation}$$

2 2

3

4 where  $\mathbf{C}_{\text{aug}}$  is a column-wise augmented matrix formed by the  $\mathbf{C}_k$  submatrices that contain the  
 5 resolved kinetic profiles in the different imbibition experiments, and  $\mathbf{S}^T$  is a single data matrix  
 6 of pure TD-RMN spectra, assumed to be common and valid for all experiments (See Figure  
 7 1b).

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

$$11 \quad \% \text{ LOF} = 100 \times \sqrt{\frac{\sum e_{ij}^2}{\sum d_{ij}^2}} \quad \text{Equation 3}$$

$$12 \quad \% r^2 = 100 \times \left(1 - \frac{\sum e_{ij}^2}{\sum d_{ij}^2}\right) \quad \text{Equation 4}$$

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

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

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

3 **[Insert Figure 1]**

### 4 **3. RESULTS AND DISCUSSION**

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

#### 9 **3.1 Individual MCR analysis**

10 An exploratory analysis of the imbibition evolution by means of classical MCR-ALS has been  
11 carried out. As a way of example, Figure 2a shows the MCR-ALS resolved concentration  
12 profiles (**C**), the pure signals (**S<sup>T</sup>**) obtained for the wild genotype 186 over imbibition time.  
13 Resolution of three species was achieved with the lack of fit (LOF %) equal to 0.97% and the  
14 explained variance ( $r^2$  %) equal to 99.9 %. The inclusion of a different number of species  
15 gave solutions worse mathematically or unreliable spectra or concentration profiles. The blue,  
16 green and red curves correspond respectively to short, transitional and long imbibition times.  
17 The shape of the resolved concentration profiles suggested a kinetic model based on two  
18 consecutive first-order reactions (  $A \xrightarrow{k_1} B \xrightarrow{k_2} C$  ). Therefore, HS-MCR was applied to the data  
19 and the postulated model was introduced as a hard-modeling constraint. Concentration  
20 profiles (**C**), the pure signals (**S<sup>T</sup>**) obtained from HS-MCR analysis resemble in shape to soft-  
21 modelling results (see Figure 2b) and similar good data fit was achieved (LOF %= 0.98 and  $r^2$   
22 %=99.9), indicating that the hypothesis of the selected kinetic model was correct. Moreover,

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

### 3 **[Insert Figure 2]**

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

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

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

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

#### 4 **[Insert Figure 3]**

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

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

16

### 17 **3.2. Simultaneous MCR analysis**

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

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

4 Firstly, simultaneous analysis by classical soft-modelling MCR-ALS has been carried out  
5 (Figure not shown). A good description of the data set was obtained (LOF %  $\leq 1,27$  and  $r^2$   
6 % = 99.9). As in individual analysis, resolution of three species is achieved and suggested  
7 again the underlying kinetic model (  $A \xrightarrow{k_1} B \xrightarrow{k_2} C$  ). Therefore, simultaneous analysis of the  
8 different imbibition experiments by HS-MCR was performed, fitting each experiment to the  
9 mentioned kinetic model. Resolved spectral profiles from the multi-experiment MCR-ALS  
10 analysis were used as initial estimates. Figure 4 shows the resolved augmented concentration  
11 matrix ( $C_{aug}$ ) and the related pure signals ( $S^T$ ) of the simultaneous analysis. Application of  
12 ILT on the related pure signals ( $S^T$ ) of the simultaneous analysis was also carried out (see  
13 Figure 4c). The resolved concentration profiles, the pure signals and the ILT of the pure  
14 signals are consistent with the reference kinetic model and with the pure signals obtained in  
15 both individual and simultaneous analysis. A rather similar fit to simultaneous soft-modelling  
16 MCR-ALS was obtained (LOF % = 3.66 and  $r^2$  % = 99.6), strongly supporting the previously  
17 proposed kinetic model (  $A \xrightarrow{k_1} B \xrightarrow{k_2} C$  ). Moreover, imbibition rate constants,  $k_1$  and  $k_2$  (see  
18 table S1 in supplementary information), for each genotype were obtained as additional  
19 information.

20 **[Insert Figure 4]**

21  
22 In order to better evaluate the imbibition trends of all genotypes, a log-transformation of  
23 obtained  $k_1$  and  $k_2$  was carried out. Figure 5 displays the 19 genotypes according to their  
24 values of  $\log k_1$  and  $\log k_2$ . Genotypes 13, 167 and 517 (in red square) seem to follow

1 imbibition kinetics similar to the wild genotype 186, where absorption of imbibition water in  
2 the adherent mucilage layer ( $\log k_1$ ) is slightly faster than the hydration of the internal seed  
3 tissues ( $\log k_2$ ). HS-MCR imbibition kinetic profiles of these genotypes (see figure 4a) show  
4 also their kinetics similarity with the wild genotype 186. Genotypes 254, 257, 259, 335 and  
5 472 (in black circle) absorb imbibition water rapidly (higher  $\log k_1$ ) but slow down the  
6 progression of water through the internal tissues (lower  $\log k_2$ ), especially the genotype 259  
7 which presents the highest  $\log k_1$ . Their kinetic profiles in Figure 4a present faster decay of  
8 the first profile (blue) and faster emergence of the intermediate state (green) than the genotype  
9 186. This difference of imbibition kinetics between 254, 257 and 259 genotypes compared to  
10 the wild genotype 186 could be due to their lower content of soluble mucilage that facilitate  
11 the hydration of the adherent layer of mucilage [8,23]. On the other hand 335 and 472  
12 genotypes show a higher soluble mucilage content but with lower intrinsic viscosity relative  
13 to 189. As mentioned before, the first ILT spectra (see blue spectra in Figure 4c), with a  
14 longer  $T_2(5)$  component, should be interpreted as water uptake by the outer soluble mucilage  
15 polysaccharides. Previous results obtained on 189 (Col-0) and two natural mutants (Mum2  
16 and Myb5) have shown that the outer soluble mucilage did not improve imbibition of internal  
17 seed tissues and, on the contrary, tend to retain water around the seeds [4]. Therefore, it is not  
18 surprising that genotypes that contain less soluble mucilage than the wild type present faster  
19 decay of this first profile (see blue profiles in Figure 4a) and, hence, higher  $\log k_1$ . In case of  
20 335 and 472 genotypes that release a higher outer mucilage content, the water uptake kinetic  
21 seems to be influenced by other physiological and/or physico-chemical characteristics. It may  
22 be due to the difference in the outer mucilage intrinsic viscosity instead of its quantity [23].  
23 For genotypes 166, 301, 397 and 549 (in blue square) the absorption of imbibition water into  
24 the adherent layer of mucilage ( $\log k_1$ ) is slower than the hydration of the internal seed tissues  
25 ( $\log k_2$ ), especially in the genotype 301. The kinetics profiles shows in Figure 4a slower decay

1 of the first profile (blue) and almost negligible occurrence of intermediate species (green) in  
2 comparison with the wild genotype. These genotypes are also characterized by less soluble  
3 mucilage but with comparable or a higher adherent mucilage content compared to 189 [8].  
4 The higher value of  $\log k_2$  ( $k_2$  is multiplied by 5 to 20 compared to 189) of these genotypes is  
5 supposed to be due to higher adherent mucilage content. It could be suggested that the  
6 adherent mucilage speeds up the transfer of water into internal tissues for 166, 301, 397 and  
7 549 seeds.

8 In genotypes 77, 136, 178, 258, 261 and 456 (in green square) the water uptake by mucilage  
9 polysaccharides ( $\log k_1$ ) and water transfer to internal tissues of seeds ( $\log k_2$ ) present similar  
10 speed. Their kinetic profiles in Figure 4a shows slower decay of the first profile (blue), and  
11 slower emergence and decay evolution of the intermediate species (green) than the genotype  
12 186 (smaller  $\log k_1$  and  $\log k_2$ ). Except for 178 and 258, this low kinetics might be correlated  
13 to a lower adherent mucilage content [8]. Compared to 189, the content of soluble mucilage  
14 varies but it is generally also less important (except for 258 and 456). .

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

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

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

#### 6 **4. Conclusions**

7 Results obtained in this work have shown the power of combining TD-NMR with  
8 multivariate curve resolution methods to study the kinetic processes of imbibed-water seeds.

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

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



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

3

#### 4 **5. Acknowledgements**

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

#### 12 **6. References**

- 13 [1] A. Macquet, M.C. Ralet, J. Kronenberger, A. Marion-Poll, H.M. North, In situ,  
14 chemical and macromolecular study of the composition of Arabidopsis thaliana seed  
15 coat mucilage, *Plant Cell Physiol.* (2007). doi:10.1093/pcp/pcm068.
- 16 [2] X. Yang, J.M. Baskin, C.C. Baskin, Z. Huang, More than just a coating: Ecological  
17 importance, taxonomic occurrence and phylogenetic relationships of seed coat  
18 mucilage, *Perspect. Plant Ecol. Evol. Syst.* 14 (2012) 434–442.  
19 doi:10.1016/j.ppees.2012.09.002.
- 20 [3] M.A. Koch, The plant model system Arabidopsis set in an evolutionary, systematic,  
21 and spatio-temporal context, *J. Exp. Bot.* (2019). doi:10.1093/jxb/ery340.
- 22 [4] S. Saez-Aguayo, C. Rondeau-Mouro, A. Macquet, I. Kronholm, M.C. Ralet, A. Berger,  
23 C. Sallé, D. Poulain, F. Granier, L. Botran, O. Loudet, J. de Meaux, A. Marion-Poll,  
24 H.M. North, Local Evolution of Seed Flotation in Arabidopsis, *PLoS Genet.* 10 (2014)  
25 13–15. doi:10.1371/journal.pgen.1004221.
- 26 [5] L.A. Colnago, Z. Wiesman, G. Pages, M. Musse, T. Monaretto, C.W. Windt, C.  
27 Rondeau-Mouro, Low field, time domain NMR in the agriculture and agrifood sectors:  
28 An overview of applications in plants, foods and biofuels, *J. Magn. Reson.* (2021).  
29 doi:10.1016/j.jmr.2020.106899.
- 30 [6] L.A. Colnago, M. Engelsberg, A.A. Souza, L.L. Barbosa, High-throughput, non-  
31 destructive determination of oil content in intact seeds by continuous wave-free  
32 precession NMR, *Anal. Chem.* 79 (2007) 1271–1274. doi:10.1021/ac062091+.
- 33 [7] R.A. Prestes, L.A. Colnago, L.A. Forato, L. Vizzotto, E.H. Novotny, E. Carrilho, A  
34 rapid and automated low resolution NMR method to analyze oil quality in intact  
35 oilseeds, *Anal. Chim. Acta.* 596 (2007) 325–329.  
36 doi:https://doi.org/10.1016/j.aca.2007.06.022.
- 37 [8] M. Cambert, A. Berger, C. Sallé, S. Esling, D. Charif, T. Cadoret, M. Ralet, H.M.  
38 North, C. Rondeau-Mouro, Datasets of seed mucilage traits for Arabidopsis thaliana

- 1 natural accessions with atypical outer mucilage, *Sci. Data.* 8 (2021) 79.  
2 doi:10.1038/s41597-021-00857-3.
- 3 [9] I. Fabrissin, G. Cueff, A. Berger, F. Granier, C. Sallé, D. Poulain, M.C. Ralet, H.M.  
4 North, Natural variation reveals a key role for rhamnogalacturonan I in seed outer  
5 mucilage and underlying Genes, *Plant Physiol.* (2019). doi:10.1104/pp.19.00763.
- 6 [10] R.F. Ling, C.L. Lawson, R.J. Hanson, Solving Least Squares Problems., *J. Am. Stat.*  
7 *Assoc.* (1977). doi:10.2307/2286501.
- 8 [11] R.L. Parker, Y.Q. Song, Assigning uncertainties in the inversion of NMR relaxation  
9 data, *J. Magn. Reson.* (2005). doi:10.1016/j.jmr.2005.03.002.
- 10 [12] R. Tauler, Multivariate curve resolution applied to second order data, *Chemom. Intell.*  
11 *Lab. Syst.* 30 (1995) 133–146. doi:10.1016/0169-7439(95)00047-X.
- 12 [13] R. de Juan, Anna; Rutan, S. and Tauler, Two-Way Data Analysis: Multivariate Curve  
13 Resolution – Iterative Resolution Methods, in: B. Brown, S. D. ; Tauler, R. and  
14 Walczak (Ed.), *Compr. Chemom.*, Elsevier, 2010: pp. 325–344.
- 15 [14] A. De Juan, M. Maeder, M. Martínez, R. Tauler, Combining hard- and soft-modelling  
16 to solve kinetic problems, *Chemom. Intell. Lab. Syst.* (2000). doi:10.1016/S0169-  
17 7439(00)00112-X.
- 18 [15] S. Mas, A. Carbó, S. Lacorte, A. De Juan, R. Tauler, Comprehensive description of the  
19 photodegradation of bromophenols using chromatographic monitoring and  
20 chemometric tools, *Talanta.* 83 (2011) 1134–1146. doi:10.1016/j.talanta.2010.06.042.
- 21 [16] A. Jayaraman, S. Mas, R. Tauler, A. de Juan, Study of the photodegradation of 2-  
22 bromophenol under UV and sunlight by spectroscopic, chromatographic and  
23 chemometric techniques, *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* (2012).  
24 doi:10.1016/j.jchromb.2012.03.038.
- 25 [17] N. Lord, G.H. Golub, C.F. Van Loan, *Matrix Computations*, *Math. Gaz.* (1999).  
26 doi:10.2307/3621013.
- 27 [18] W. Windig, C.E. Heckler, F.A. Agblevor, R.J. Evans, Self-modeling mixture analysis  
28 of categorized pyrolysis mass spectral data with the SIMPLISMA approach, *Chemom.*  
29 *Intell. Lab. Syst.* (1992). doi:10.1016/0169-7439(92)80104-C.
- 30 [19] R. Tauler, A. Smilde, B. Kowalski, Selectivity, local rank, three-way data analysis and  
31 ambiguity in multivariate curve resolution, *J. Chemom.* 9 (1995) 31–58.  
32 doi:10.1002/cem.1180090105.
- 33 [20] M. Maeder, A.D. Zuberbühler, Nonlinear Least-Squares Fitting of Multivariate  
34 Absorption Data, *Anal. Chem.* (1990). doi:10.1021/ac00219a013.
- 35 [21] *Maximum Entropy and Bayesian Methods*, 1989. doi:10.1007/978-94-015-7860-8.
- 36 [22] J. Jaumot, A. de Juan, R. Tauler, MCR-ALS GUI 2.0: New features and applications,  
37 *Chemom. Intell. Lab. Syst.* 140 (2015) 1–12. doi:10.1016/j.chemolab.2014.10.003.
- 38 [23] D. Poulain, L. Botran, H.M. North, M.C. Ralet, Composition and physicochemical  
39 properties of outer mucilage from seeds of *Arabidopsis* natural accessions, *AoB Plants.*  
40 (2019). doi:10.1093/aobpla/plz031.
- 41  
42  
43  
44  
45  
46  
47  
48  
49  
50

1 **Figure Captions**

2

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

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

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

11 **Figure 4.** HS-MCR results for the multiset analysis of 19 genotypes: a) kinetic profiles of  
12 each genotype from the augmented concentration matrix (**C<sub>aug</sub>**), b) related pure signals, c) the  
13 ILT processing of pure signals.

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

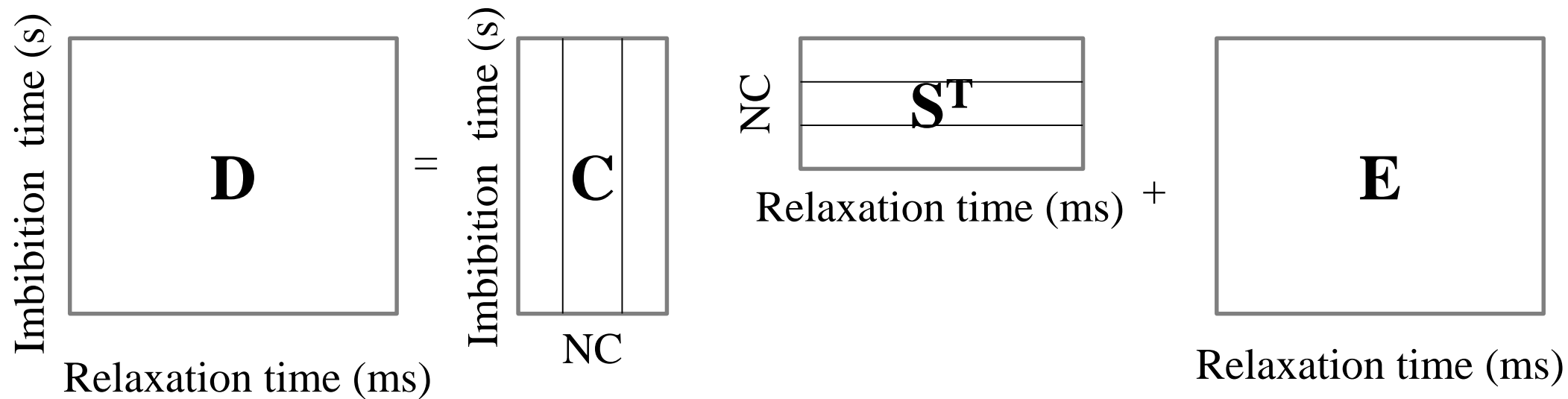
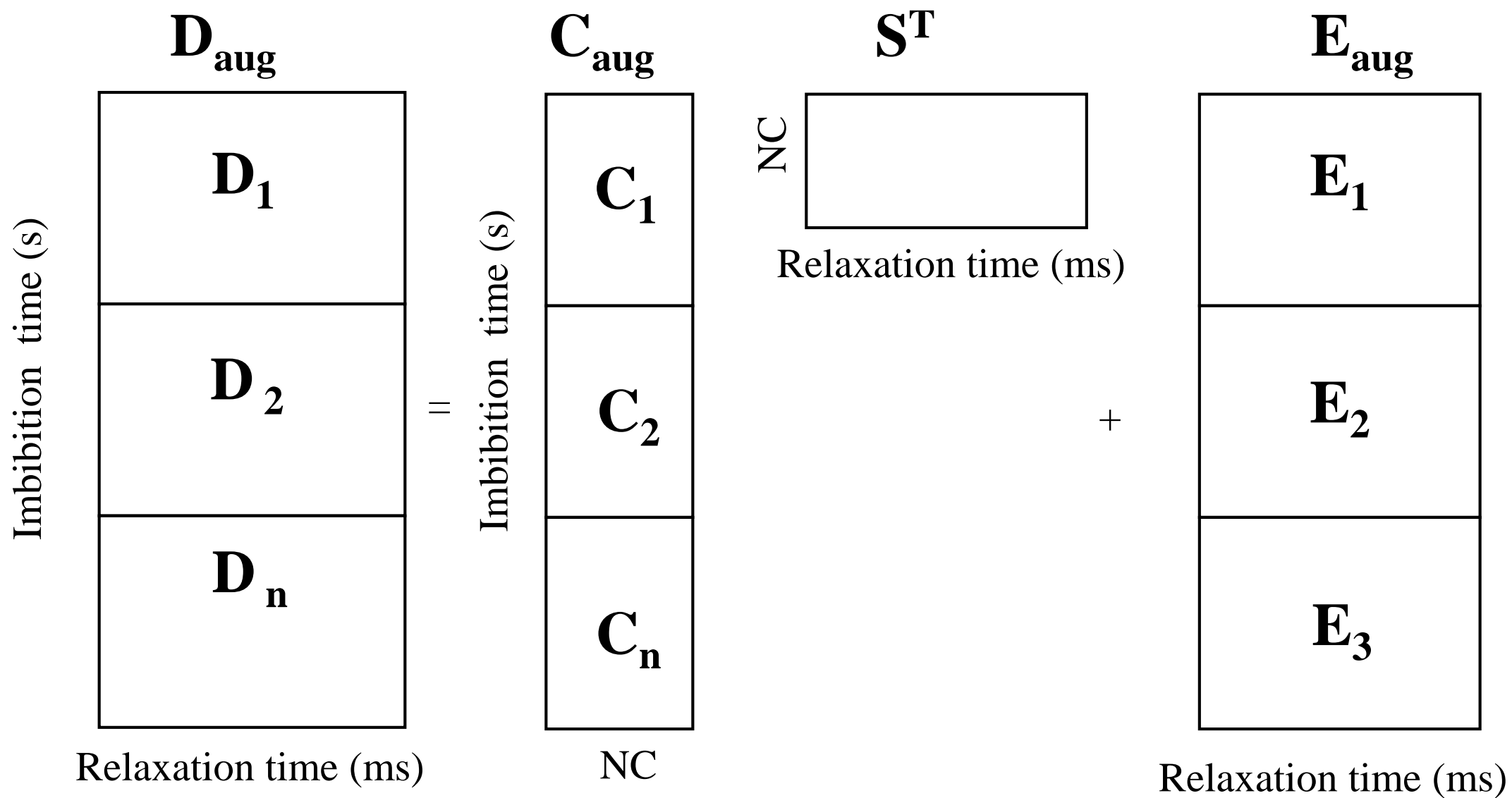
17

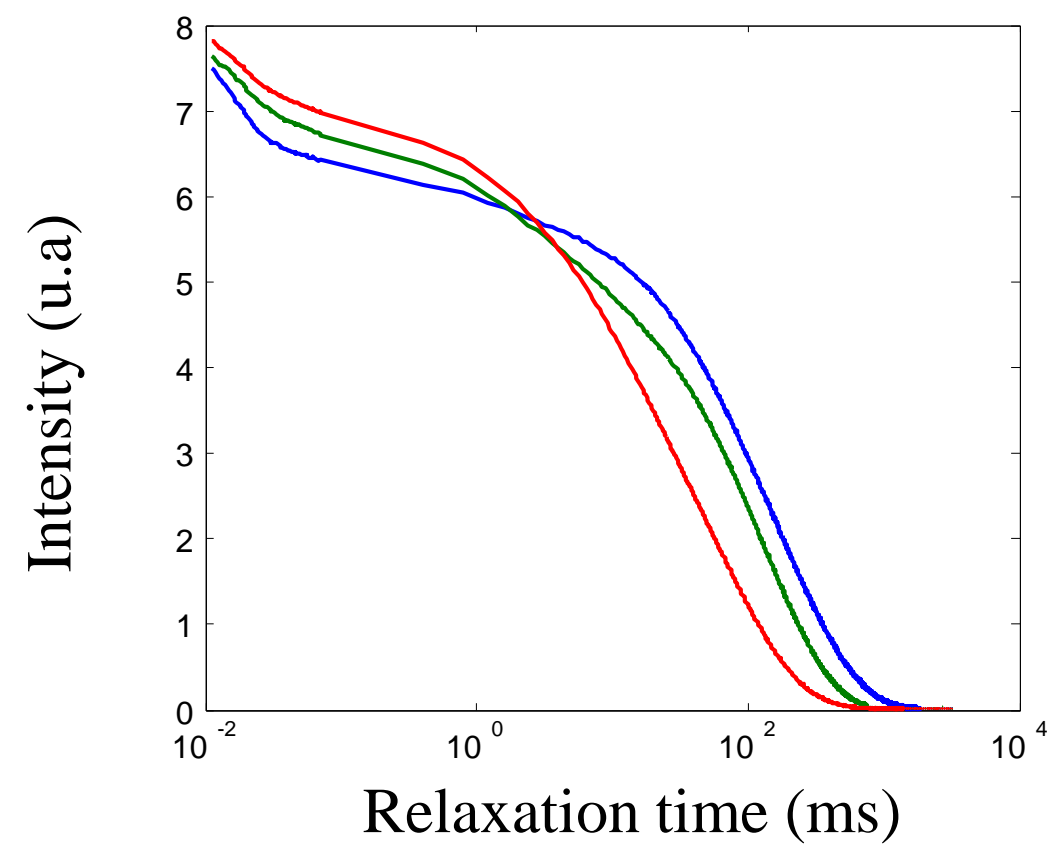
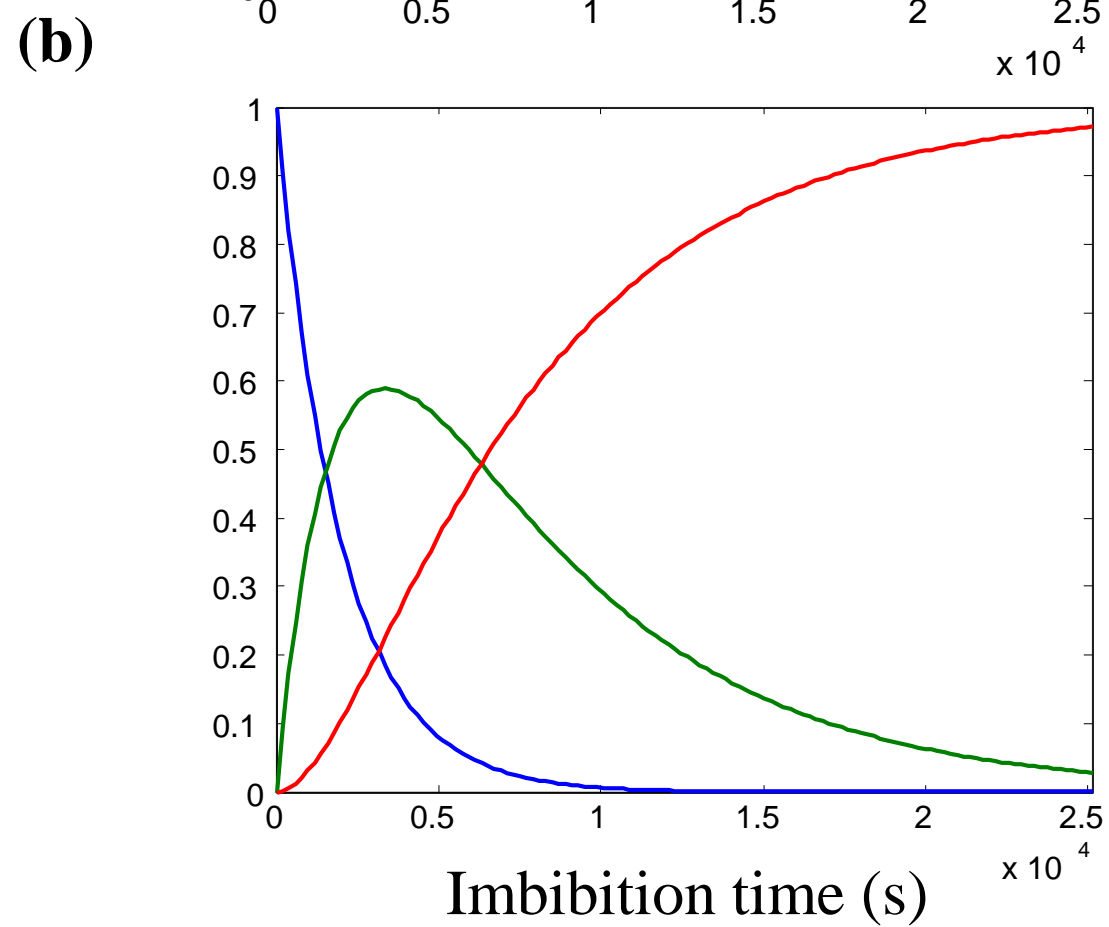
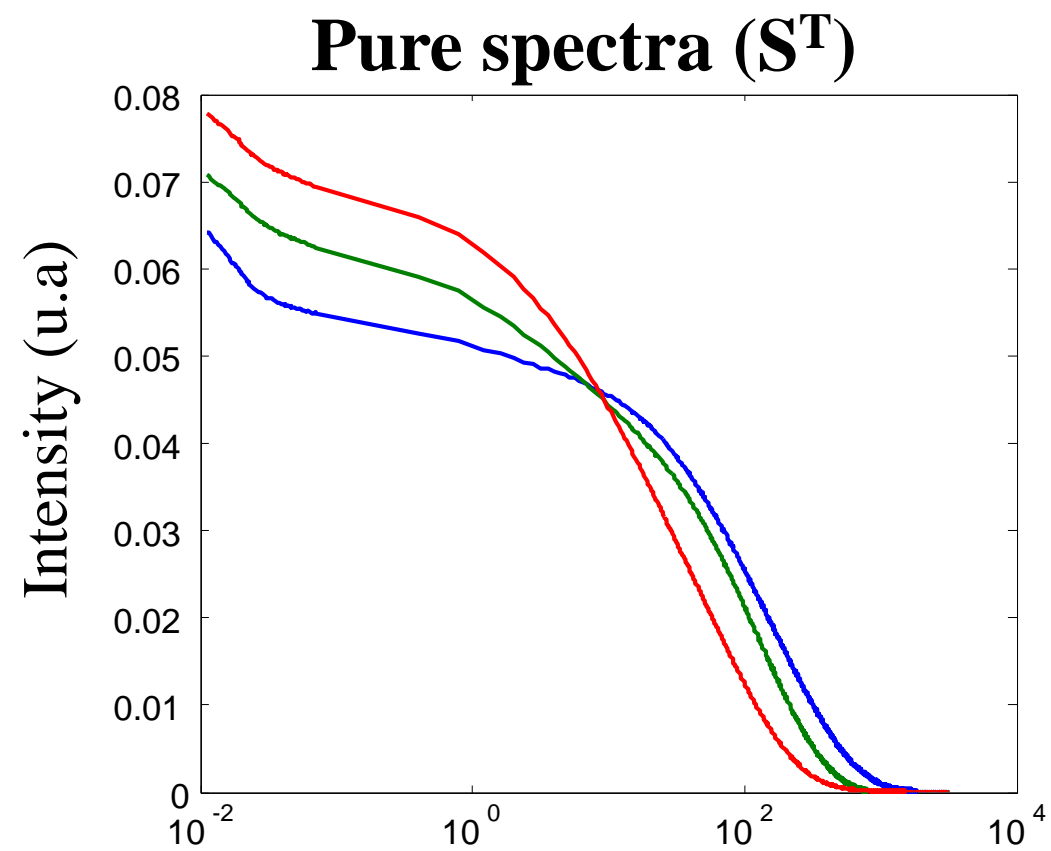
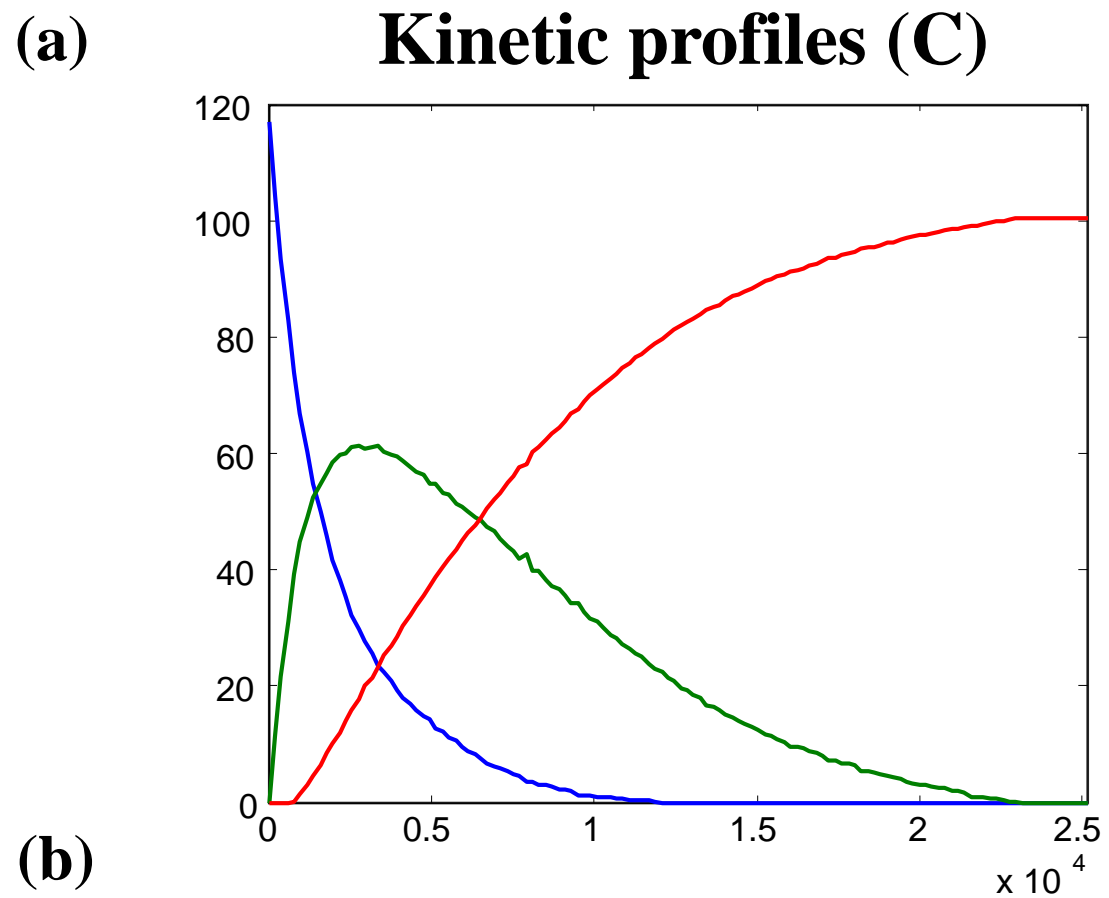
18

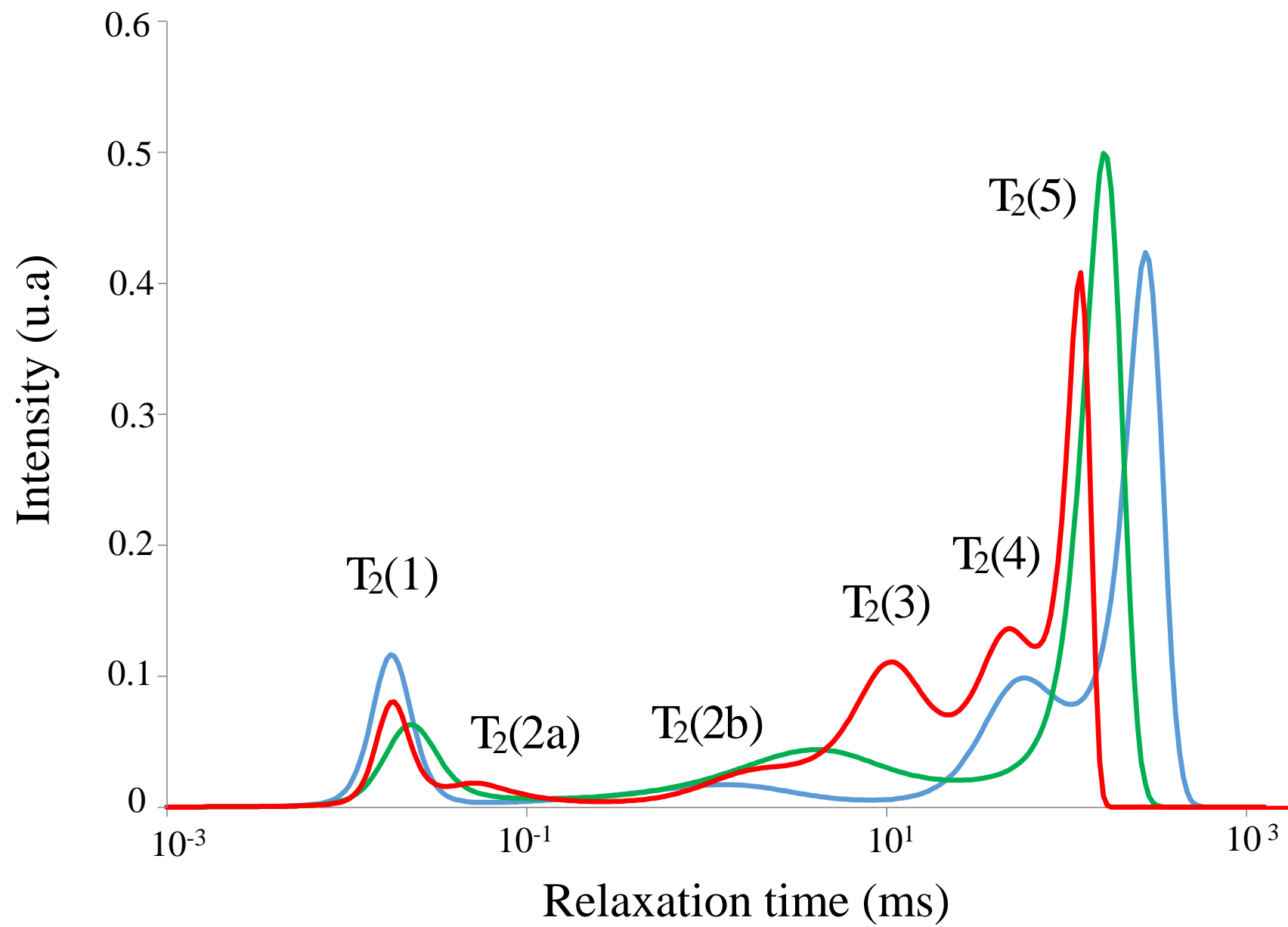
19

20

21

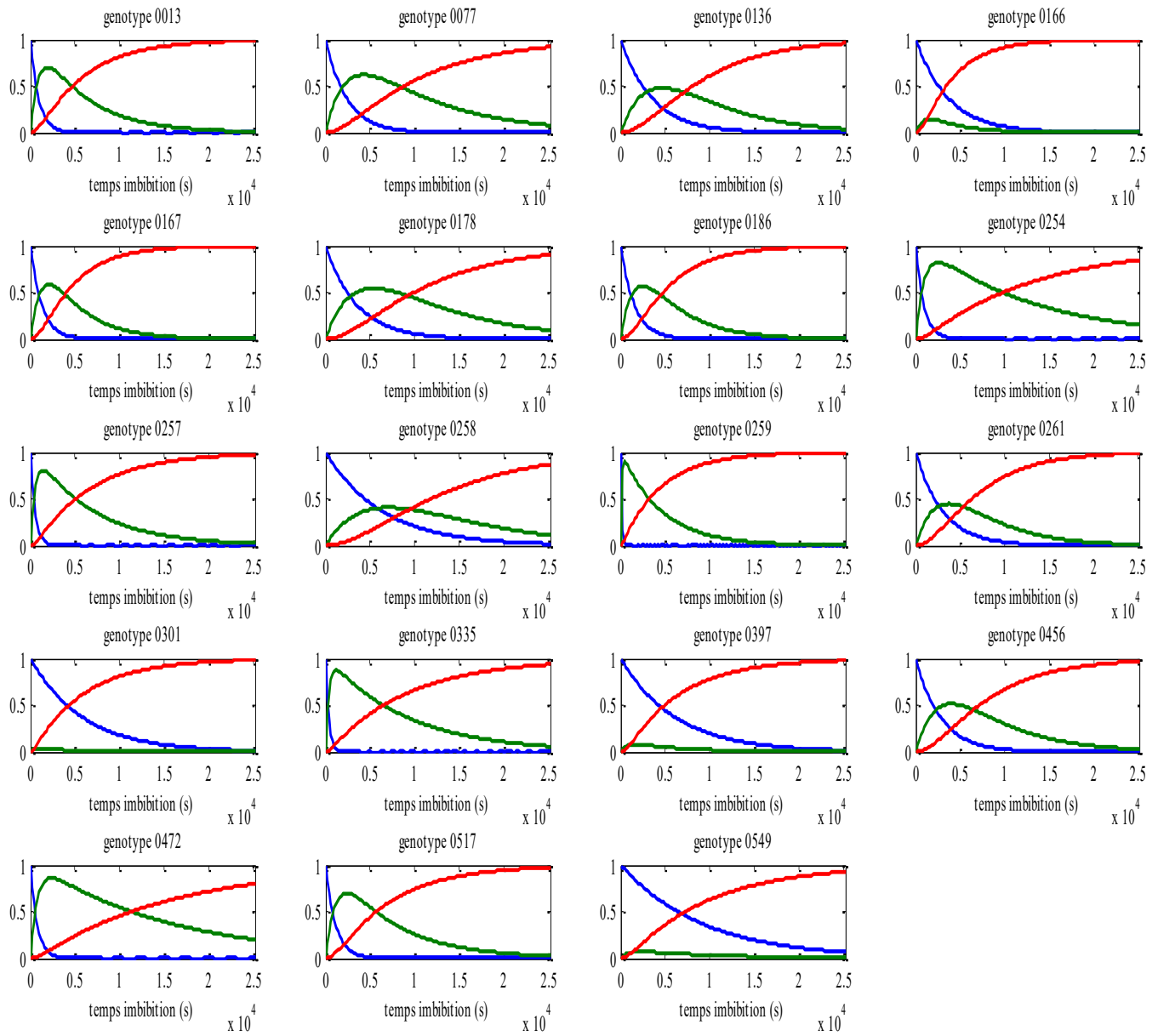
**(a)****(b)**



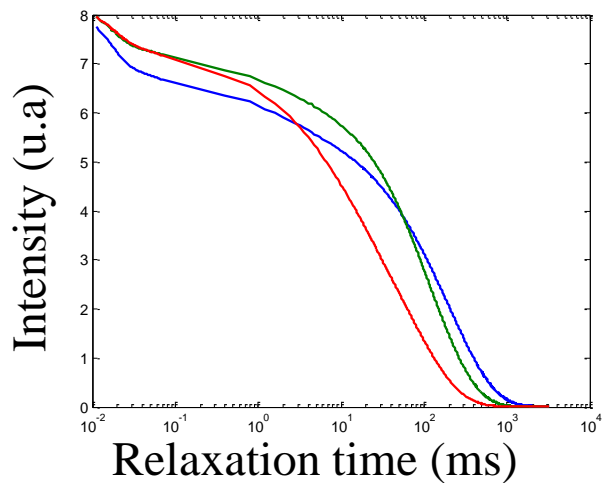


# Kinetic profiles (Caug)

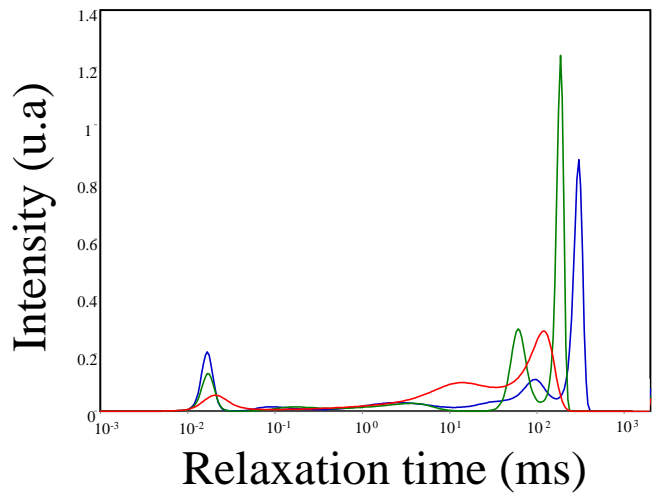
(a)

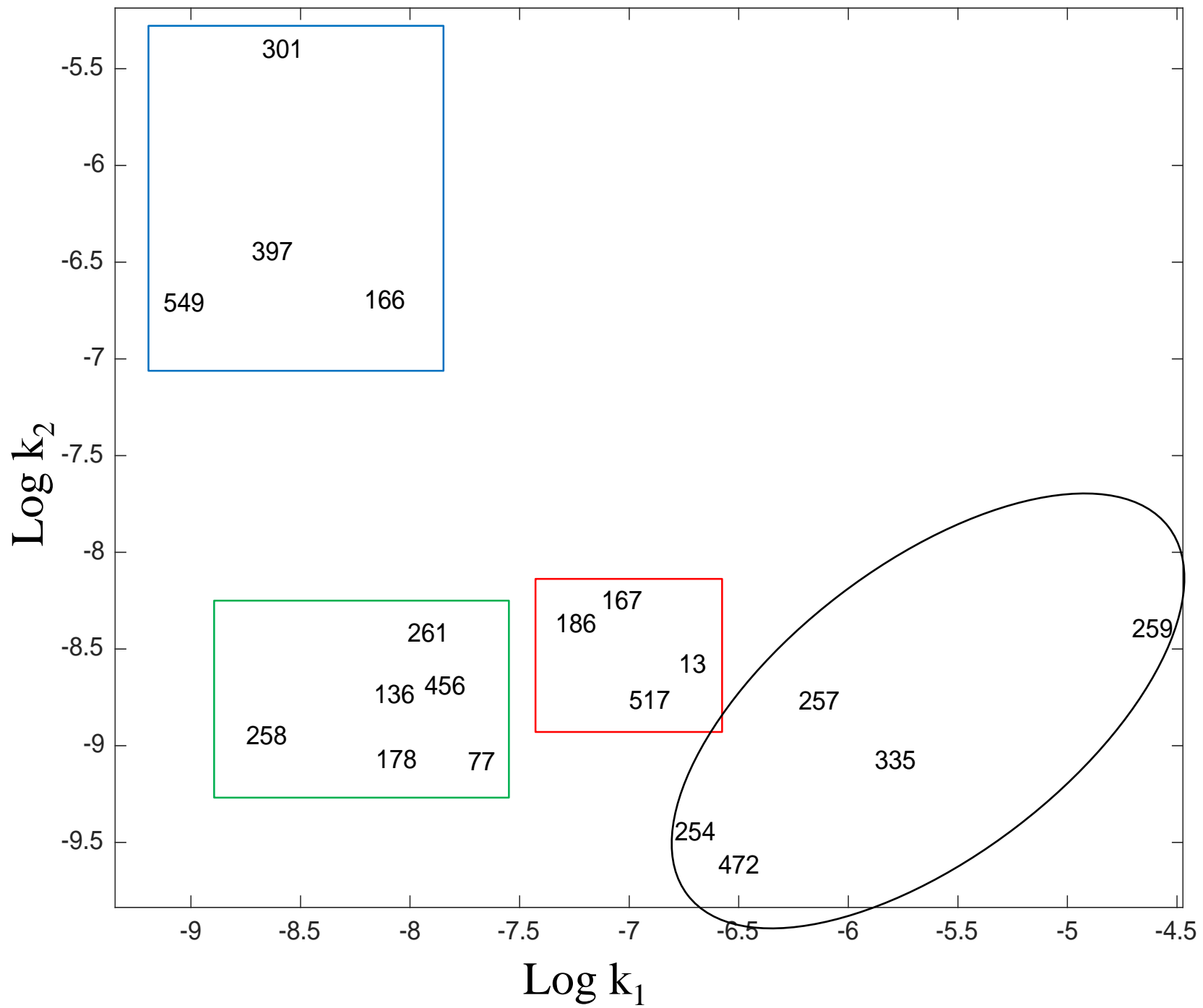


(b) Pure spectra ( $S^T$ )



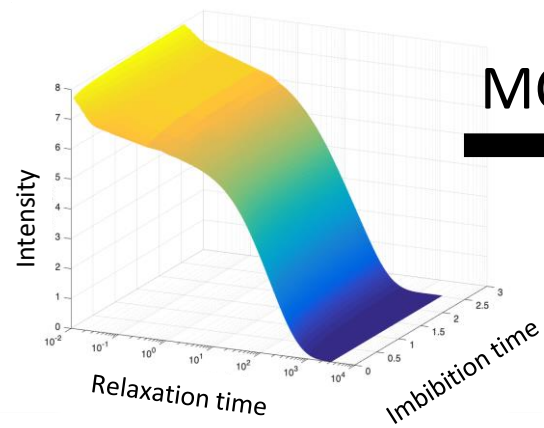
(c)







TD-NMR signal



MCR-ALS

