UPPSALA
UNIVERSITET

UPTEC X 20016
Examensarbete 30 hp
Juni 2020

Characterizing the pore structure of porous matrices using SEQ-NMR spectroscopy

Ella Strömberg

UPPSALA
UNIVERSITET

Teknisk- naturvetenskaplig fakultet
UTH-enheten

Besöksadress:
Ångströmlaboratoriet
Lägerhyddsvägen 1
Hus 4, Plan 0

Postadress:
Box 536
751 21 Uppsala

Telefon:
018 – 471 30 03

Telefax:
018 – 471 30 00

Hemsida:
http://www.teknat.uu.se/student

Abstract

Characterizing the pore structure of porous matrices using SEQ-NMR spectroscopy

Ella Strömberg

Characterization of the pore structure is a crucial part in the manufacturing of porous media used for purification of biological pharmaceuticals. This project took place at Cytiva in Uppsala and aimed at optimizing a newly developed method in pore structure characterization called size-exclusion quantification NMR (SEQ-NMR). By measuring with diffusion NMR on a polymer solution before and after equilibration with a material of interest the pore structure of the material can be determined. This project aimed at reducing the duration of a SEQ-NMR experiment while examining the performance of the method during different conditions with the goal of making the method applicable for quality control procedures. The method was optimized both by simulations and by experimental diffusion NMR measurements. It was discovered that the performance of the method could be improved by having an optimal mixture of the polymer solution and during experiments distributing ten measurement points with linear spacing. With these parameters optimized the duration of the method could be reduced with 22 hours landing on a total duration of 8 hours. The duration combined with the complexity of the method still makes the method unsuitable for use in quality control of porous media. Despite the small possibility of SEQ-NMR being a quality control method this project has proven the method to be both reproducible and sensitive.

Handledare: Fredrik Elwinger
Ämnesgranskare: Katarina Edwards
Examinator: Erik Holmqvist
ISSN: 1401-2138, UPTEC X 20016

Populärvetenskaplig sammanfattning

Kanske hade du precis som jag en klosslåda när du var liten. I de olika hålen på lådan fick klossarna plats om de hade rätt form och storlek. För ett företag som ska tillverka och sälja en klosslåda krävs väldigt precis kunskap om vilka mått, vilken struktur, hålen i lådan har. Det här projektet fokuserar kring optimeringen av en metod för att kunna bestämma just storleken på lådans håligheter där själva lådan egentligen är porösa geler som Cytiva i Uppsala tillverkar. Klossarna som passar eller inte passar i lådan motsvarar biologiska läkemedel som renas fram med hjälp av den porösa gelen. Metoden i projektet heter size-exclusion quantification nuclear magnetic resonance (SEQ-NMR) och bygger på mätningar av en lösnings koncentration innan och efter den varit i kontakt med en porös gel. Om en samling av klossar i olika storlekar och former hälls över klosslådan så kommer vissa klossar gå ner i lådan medan andra hamnar utanför. Genom att mäta klosshögens koncentration, hur många av varje storlek och form det finns i förhållande till hela högen, och jämföra den med koncentrationen av klossar som hamnade utanför lådan kan strukturen på lådans ihåligheter bestämmas.

Koncentrationen mäts indirekt genom diffusions-NMR där molekylers rörelse mäts med hjälp av ett magnetfält och applicering av magnetiska pulser. Pulserna appliceras i par och om en molekyl förflyttar sig, diffunderar, mellan pulserna kommer det synas som en försvagning av den signal som fås av mätningen. Ju starkare den applicerade magnetiska pulsen är desto större blir försvagningen i signal. Den försvagade signalen följer en avtagande kurva och kan översättas till en koefficient som karaktäriserar diffusionen av den molekylen. Då pulsstyrkan är noll är signalen direkt proportionell, lika med, koncentrationen av molekylen i lösningen. Det här kan sedan matematiskt översättas till vilken porstorlek den undersökta porösa gelen har.

Det är viktigt att strukturen på de porösa material som Cytiva producerar är karakteriserade på ett korrekt sätt för att framreningen av biologiska läkemedel som sedan distribueras till patienter håller hög kvalité. För att uppfylla efterfrågan på biologiska läkemedel krävs det också att efterfrågan på porösa material produceras effektivt där hög noggrannhet upprätthålls och produktionen sker på ett reproducerbart sätt. Alla geler som produceras testas därför för att kontrollera att de upprätthåller den kvalité som krävs för produktion av läkemedel. De biologiska läkemedlen kan sorteras på olika egenskaper och för geler som separerar med avseende på storlek är det porstorleken hos gelen som specificerar produkten. Porstorleken motsvarar alltså hålen i klosslådan och genom att veta dess storlek kan produkten specificeras för vilka storlekar på läkemedelsprodukter den kan rena fram.

Den metod Cytiva använder idag för att bestämma porstrukturen tar cirka 15 timmar vilket inte är optimalt i en process där man vill producera stora mängder gel. Lösningen som är i kontakt med gelen vid SEQ-NMR består av dextran, en stor grenad molekyl som kan ha olika storlekar.

En aspekt av optimeringen var att hitta den perfekta blandningen av dextranstorlekar, fördelningen av klossarna i samlingens storlekar. Vid starten av detta projekt tog SEQ-NMR 15 timmar per diffusionsmätning vilket ger en total tid på 30 timmar då mätningar ska göras på lösningen både före och efter jämvikt. När den optimala blandningen hittats genom simulering och mätningsmetoden optimerats hade totala experimenttiden förkortats till 8 timmar. Det motsvarar en förbättring hos utförandet av metoden men den perfekta blandningen av storlekar på dextran visade även från simuleringar att samma samling klossar inte fungerar för att bestämma storleken på hålen hos samtliga lådor. Cytiva producerar en mängd olika porösa geler bara för storleksseparation och att behöva en optimal storleksfördelning av dextran för varje gel gör metoden svårare att tillämpa inom till exempel kvalitetskontroll vilket var det tänkta användningsområdet för metoden.

Table of Contents

Abbreviations

E	NMR signal attenuation
D	diffusion coefficient
δ	pulse duration
Δ	diffusion time
g	gradient strength
ISEC	inverse size exclusion chromatography
K_{eq}	distribution coefficient
M	molecular weight
P	distribution
PGSTE	pulsed field gradient stimulated echo
R^2	coefficient of determination
r_H	hydrodynamic radius
r_P	pore radius of resin
SEQ-NMR	size exclusion quantification nuclear magnetic resonance
SEC	size exclusion chromatography
SNR	signal-to-noise ratio
$V_{accesible}$	accessible pore volume for a certain molecule
V_{pore}	total pore volume
V_{tot}	total volume in column
V_{void}	volume between resin beads

1 Introduction

Chromatography is a common method for separating molecules by letting them pass through a porous material called resin. The resin can have different characteristics regarding size and chemical properties. What resin to use in chromatography depends on the characteristics of the molecule of interest. The most important characteristic of the resin when separating molecules with respect to size is the pore structure, which is difficult to both specify and asses. The main aim of this project is to optimize a newly developed method for pore characterization. This project is performed as a master thesis at Cytiva in Uppsala. The current method for pore structure characterization used by Cytiva is inverse size exclusion chromatography (ISEC) which uses defined polymer standards to characterize the pore structure of resins. ISEC is a reversed form of size exclusion chromatography (SEC) where instead the pore structure is known, and molecules are separated with respect to that structure. SEC is a common method for determining molecular weight distribution and in practice is the experimental procedure the same for both methods. ISEC is a simple method in terms of execution but it is also time consuming and requires packing of a column making it less optimal for quality control measurements (Guo *et al.* 2017). Cytiva produces a range of materials used to purify biological pharmaceuticals and a routine procedure for characterizing these materials in a reproducible way is required.

The method to be optimized in this project was developed in 2018 by Elwinger *et al*., where a novel and promising approach for pore structure characterization was presented. The method is called size-exclusion quantification nuclear magnetic resonance (SEQ-NMR), and the principle of the method is based on a solution of polymers with a wide size range equilibrating with the porous material to be examined. Smaller polymer fragments within the solution have access to a greater part of the total pore volume compared to larger polymers and their concentration will therefore be reduced in the surrounding solution. The focus of this project is to optimize SEQ-NMR to make it more time efficient and, in the future, suitable for quality control analysis. Throughout the optimization the method will also be examined regarding its sensitivity and robustness together with other aspects of its performance.

In SEQ-NMR a solution of polymers is analysed with diffusion NMR before and after equilibration with a resin. Through a multiexponential fit to the received data the change in size distribution is obtained making it possible to determine the pore structure of the resin. In SEQ-NMR no column packing is needed, and the pore structure can be determined by using a solution with broad size distribution of polymers, i.e. no monodisperse polymers are needed. This makes SEQ-NMR more advantageous compared to the ISEC method (Elwinger *et al.* 2018). ISEC takes 15 hours which is equivalent to one diffusion measurement of SEQ-NMR. Two measurements are needed giving SEQ-NMR a total duration of 30 hours. The aim of this master thesis is to reduce the duration of SEQ-NMR to less than one hour per diffusion measurement while making an evaluation of the method.

2 Background

2.1 NMR

NMR is based on nuclei having properties as angular momentum and magnetic moment, referred to as spin (Hore 2015). The spins of the nuclei in a sample to be analysed are at first randomly oriented. When the sample is put in a static magnetic field the magnetic moments will take the direction with or opposite the direction of the magnetic field. The distinct orientations relative to the magnetic field will exhibit slightly different energies and will be populated according to the Boltzmann distribution. The difference in energy between spin directions is what makes all NMR measurements possible (Bruice 2010). Further insight to the basics of NMR can be found in P.J. Hore's book *Nuclear Magnetic Resonance.*

2.2 Diffusion NMR

Diffusion NMR is being used in medical, biological and material science applications (Guo *et al.* 2017). By measuring the self-diffusion of a molecule, it becomes possible to study its size and shape. Diffusional movement is driven by thermodynamic energy and can be quantitatively described by the self-diffusion coefficient *D*. This coefficient is a measure of with what rate a molecule moves in the unit m^2s^{-1}. From a diffusion NMR measurement information on the relative size of a molecule can be received from the relationship between *D* and hydrodynamic radius r_H,

$$D = \frac{k_B T}{6\pi\eta r_H} \quad . \tag{1}$$

This is known as the Stokes-Einstein equation where k_B is the Boltzmann constant, T is the absolute temperature and η is the solution viscosity. Hence, the diffusion coefficient is inversely proportional to the radius of the diffusing molecule. How well Eq. (1) provides an accurate estimation depends on the shape of the studied compound, the more sphere like a molecule is, the better is the estimation. (Claridge 2009).

When relating the diffusion coefficient to size via the Stokes-Einstein equation the temperature is often known from calibration, but the solution viscosity can be harder to find or determine. At low concentrations, in the millimolar range, where interactions between the diffusing entities can be neglected the viscosity of the sample can be assumed to be the same as the solvent viscosity. Other ways to determine the viscosity is by using a molecule with known hydrodynamic radius in the solution and calculate the solution viscosity from the measured diffusion coefficient or by measuring the viscosity of the mixed solution. The requirement for using a molecule with known hydrodynamic radius is that there should be no interference between the signal of the reference molecule and the other components in solution (Claridge 2009).

2.2.1 Pulsed field gradient stimulated echo

The pulse program used for diffusion NMR experiments in this project is pulsed field gradient stimulated echo (PGSTE). Here, two gradient pulses with strength g and duration δ are applied to the sample separated by the diffusion time Δ (Claridge 2009). The applied pulses can have different shapes and rectangular pulses are used throughout this project. It is the simplest pulse shape and can therefore suffer some drawbacks when using high gradients (Willis *et al.* 2016). The first gradient pulse gives a spatial dependence of the magnetization in the sample providing information on the position of the nuclei. The second pulse reverses this dependence by refocusing the dephased magnetization. The refocusing is only perfect if the nuclei are in exactly the same physical location at the time of the first and second pulses. If diffusion occurs the refocusing will not be complete and an attenuated signal is obtained. The attenuated signal is both dependent on the length and strength of the gradient pulse and on how far the molecules diffuse during Δ which in turn depends on the diffusion coefficient of the molecule. To measure the diffusion coefficient of a molecule, g can be applied in a gradient of increasing strength, higher g means more dephasing of the magnetization in the sample leading to less signal being refocused. With increasing g there will be more attenuation of the signal (Claridge 2009).

The data analysed from PGSTE experiments is the attenuation (E) of the signal due to diffusion. By integrating the peak area as a function of g the attenuation can be described by following the equation,

$$E = e^{-bD} \tag{2}$$

where b in the exponent is given by,

$$b = \gamma^2 g^2 \delta^2 \left(\Delta - \frac{\delta}{3}\right) \tag{3}$$

where γ is the gyromagnetic ratio of the nuclear spin (Price 1997). In diffusion measurements it is common to use protons, ^{1}H, since it is stable and the nucleus with the highest γ (Hore 2015).

As illustrated by Eq. (1) above is the diffusion coefficient not only dependent on the size and shape of the analysed molecule but also the viscosity of the solution, the temperature of the sample as well as concentration. This means that the parameters used in a diffusion experiment often need to be optimized for each new sample. The principal parameters δ, g and Δ can be determined once the physical parameters of the sample have been set. When deciding the principal parameters, the goal is to receive a considerable attenuation, so that the subsequent fitting to the data yields reliable results. If the attenuation is too rapid the late recorded data points, having high b values, will not contribute to meaningful fitting to the data and if the attenuation is too slow it will not give an accurate determination of D (Claridge 2009). Common values for Δ are between milliseconds and hundreds of milliseconds making it possible for macromolecules to diffuse a distance much longer than their own radii during Δ. The study of larger molecules having small values of D require higher values of Δ which can lead to a

decrease in the signal-to-noise ratio (SNR). Small sample volumes can also give a reduction in SNR. Common values for δ are 1-10 milliseconds (Claridge 2009, Stilbs 2019).

2.2.2 Data processing

The data obtained in a diffusion NMR experiment is a measure of the attenuated signal as either peak height in the spectra or integrated peak area, both as a function of *g*. From this can *D* be derived either by plotting *E* against *b* and make an exponential fit, or by plotting *E* on a logarithmic scale against *b* where a straight line will be given with *-D* as the slope. This is possible since all parameters within *b* are constant except *g* (Claridge 2009).

2.3 Sensitivity and SNR

The signal strength of NMR as a technique is considered low. SNR is a well-established concept of signal processing and in NMR is the SNR defined as the height, or as in this report, the area of the NMR peak divided by the root mean square of the noise (Hyberts *et al.* 2013). The noise is obtained by integrating ten equal areas far away from peaks in the spectrum and calculating the standard deviation for these integrals. The signal is deterministic, constant, and the noise is randomly fluctuating, making SNR increase by the square root of the number of scans (Hore 2015).

To receive a high-resolution spectrum, it is important that the magnetic field is homogeneous. To increase the homogeneity of the field the currents applied to specially designed assisting coils can be adjusted, a process called shimming (Topgaard *et al.* 2004). The shimming is performed prior to all measurements and, when using short sample heights, the process can be very difficult. The sample must moreover be placed in the sensitive region of the signal receiving coil which is approximately 1 cm along the sample height (Price 2009).

2.4 Polydispersity and dextrans

The polymers used in the SEQ-NMR measurements in this project are dextrans, which are common in biotechnological and pharmaceutical applications. A dextran is a branched glucan polymer with branched chains that can differ in both length and weight. This means that a dextran is polydisperse and the branches may often consist of one or two glucose molecules. The molecular weight of a polymer is therefore a mean value that is dependent on its actual molecular weight distribution. Three variables are commonly used for describing the weight of a dextran polymer. M_n is the number-average molecular weight, M_w is the weight-average molecular weight and M_p is the peak molecular weight. The molecular weights are often modeled by lognormal distribution with the relationship between the weights as $M_p = \sqrt{M_w \times M_n}$ (Kuz'mina *et al.* 2014). Throughout the report the dextran weights will be given as M_p values. An NMR diffusion measurement on a polydisperse solution will give data as the integral of the signal from all polymers in the solution (Guo *et al.* 2017). Consequently, the diffusional attenuation measured is affected by polydispersity present in the sample.

2.5 SEQ-NMR

As described in the introduction, SEQ-NMR uses a solution of polymers which is analysed before and after equilibration with a resin. The polymers used in this project are dextrans as stated in Section 2.4, the solution is made by dissolving the dextrans in heavy water, D_2O. The distribution of each dextran follows lognormal distribution and is computed as

$$P_i = \frac{1}{\sqrt{2\pi}\sigma_i} e^{\left(-\frac{1}{2}\left(\frac{ln(M)-\mu_i}{\sigma_i}\right)^2\right)}, \quad (4)$$

where M is the molecular weight, μ_i is the expected value and σ_i is the standard deviation of dextran i (Chang 2015). M is connected to r_H through the Mark Houwink equation,

$$r_H = aM^{\alpha}, \quad (5)$$

where a and α are molecule specific parameters. The distribution coefficient K_{eq} and r_H gives the selecitivity curve further described in Section 2.5.1.

2.5.1 Selectivity curve

The theory behind the selectivity curve as a result of SEQ-NMR experiments is shared with the theory of ISEC experiments. The selectivity curve is the distribution coefficient K_{eq} as a function of molecular size expressed in r_H, for a one-pore model are they connected by

$$K_{eq} = \left(1 - \frac{r_H}{r_p}\right)^2, \quad (6)$$

where r_p is the pore radius of the resin. The one-pore model assumes all pores of the resin bead to have the same size and structure. Regardless of the pore model is K_{eq} also connected to the volumes describing the functionality of a resin by the relationship

$$K_{eq} = \frac{V_{accessible}}{V_{pore}}. \quad (7)$$

$V_{accessible}$ is the accessible pore volume for a certain molecule, and V_{pore} is the total pore volume. In a packed column of porous beads, V_{pore} is given by V_{tot} minus V_{void} where V_{tot} corresponds to the total available volume of the column and V_{void} corresponds to the volume in between resin beads (Knox & Scott 1984). In practice, V_{void} can be determined with the help of a large molecule that cannot access any pores in the column while V_{tot} is assessed with the help of a small molecule that can access all pores (Knox & Ritchie 1987). In this project V_{tot} is determined by measurements using D_2O, and V_{void} by measurements with a large polyethylene oxide polymer (PEO). These are not measured with diffusion NMR, instead the volumes are calculated by the NMR signal being proportional to the concentration c. The dilution equation $c_1V_1 = c_2V_2$ where the concentrations c_1 and c_2 are from the NMR signals before and after

equilibrium and V_1 is the volume added to the resin making V_2 the sum of V_1 and the searched for volume. The NMR signal is hence, turned into volume by comparing the signals from before and after equilibrium (Bruice 2010).

The solution of polymers will after equilibrium with a resin be diluted compared to the stock solution added to the resin. This dilution *d* is described by

$$d = \frac{V_0}{V_0 + V_{void} + V_{accesible}} \,, \tag{8}$$

where V_0 corresponds to the volume of stock solution added to the resin before equilibration. The value of *d* will be specific for each polymer length in the solution (Elwinger *et al.* 2018). The change in concentration will give information on the pore size distribution in terms of the selectivity curve.

2.5.2 Attenuation curves

When a solution contains multiple compounds giving rise to the attenuated signal the attenuation becomes a summation of Eq. (2) for each individual component. The attenuation before equilibrium is calculated by

$$E = \sum_{i=1}^{N} P_i e^{-bD_i} \,, \tag{9}$$

where the distributions and diffusion coefficients are specific for each dextran *i*. The attenuation after equilibrium is then given by combining Eq. (8) and Eq. (9) as

$$E = \sum_{i=1}^{N} d_i P_i e^{-bD_i} \,. \tag{10}$$

2.5.3 Evaluation of fitted models

The model fit can be expressed in terms of R^2 known as the coefficient of determination. R^2 is the sum of fit residuals squared relative to the sum of the mean square deviations from the average value of the data,

$$R^2 = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \,. \tag{11}$$

In equation 11, y_i are the observed values, $\hat{y}_i$ the fitted curve and $\bar{y}$ is the mean *y*-value. When R^2 is close to one the data overlaps well with the fitted model (Smith 2015).

2.6 Experimental errors in diffusion NMR

The most common factors giving errors in the diffusional attenuation are convection, non-uniform gradient pulses and eddy currents (Kuz'mina *et al.* 2014).

2.6.1 Convective flow

One common reason for receiving unreliable diffusional data from a modern NMR spectrometer is convection within the sample. Convection arises from temperature gradients in the sample caused by the temperature regulation in the NMR spectrometer. The regulation is often performed by a flow of gas passing over the sample tube. To ensure the sample is exposed to constant temperature the gas is often heated before entering the probe. The sample temperature is regulated via a feedback mechanism controlled by a sensor placed in the probe at the base of the sample tube. As a result, the overall temperature of the sample will be stable. However, if extensive heating is required thermal gradients may appear within the sample. This causes convective flow which displaces molecules leading to a faster signal attenuation than what self-diffusion would generate. Hence, the data provide larger inaccurate diffusion coefficients (Claridge 2009). Since convection is present over the whole sample volume, every molecule in the solution is affected by convection in the same way regardless of their size (Price 2009). It is important to check for convective flow before performing diffusion measurements (Claridge 2009).

One way to reduce the temperature gradients leading to convective flow is to remove temperature regulation, having no gas pass over the sample and make sure that the temperature is equilibrated. This method is both limiting and impractical. Before all diffusion measurements the sample should be allowed to equilibrate for a period of approximately 30 minutes. The temperature gradients can also be reduced by having a high flow rate of the gas passing over the sample. This can limit convection but may instead cause vibrations in the sample. A way to test if convective flow is present is by doing the same experiment with different values of Δ and then compare the resulting diffusion coefficients. If no convection is present, the diffusion coefficient will be the same for the different experiments (Claridge 2009).

2.6.2 Non-uniform gradient pulses

The attenuated signal as described by Eq. (2) will not be completely accurate if the gradient applied to the sample is not perfectly uniform. The more attenuated the signal is, with a non-uniform gradient, the deviation from Eq. (2) will increase. This in turn results in inaccurate diffusion coefficients from the Eq. (2) fitting. Non-uniform gradient pulses also increase the error estimate in the data fitting. This will lower the resolution of diffusion in the diffusion experiments. All NMR probes have non-uniform gradients to some extent. It is common that the gradient is strongest at the middle of sample and decreases on each side. (Connell *et al.* 2009).

2.6.3 Eddy currents

When using rectangular gradient pulses, the steep increase in the local magnetic fields can generate eddy currents in the conducting materials around the sample. The effect of eddy currents increases with the pulse intensity and speed of the pulse rise and fall. If eddy currents are present, they can lead to changes of phase in the spectra together with irregular attenuation changes and spectral broadening. The minimum time needed after a pulse before signal recording may be initiated is the time needed to lose eddy currents. Modern shielded gradient coils normally only produce negligible eddy currents, but sometimes additional actions are needed. One way to reduce eddy currents is to use pre-emphasis where addition of a small exponential correction at the leading and closing edges of the pulse can compensate for the effect of the eddy currents. The risk with pre-emphasis is that new eddy currents can be induced by the actions made to avoid it (Price 2009).

2.7 Porous matrices

Porous matrices, resins, are the key materials in chromatography. Different types of resins are needed for different separations and one characterization aspect are their pore structure or pore size. Sephacryl High Resolution (HR) resin beads are common in SEC and were used for the SEQ-NMR characterization in this project. Since SEC separates molecules with regard to size the components in solution will not specifically bind to the beads (GE Healthcare Life Sciences 2018). The K_{eq} curves of a selection of resins produced and distributed by Cytiva are presented in Figure 1 and show the range of pore sizes that exists in the collection of resins. Different resins differ from each other in pore structure but lot to lot variation of the same resin product also exists.

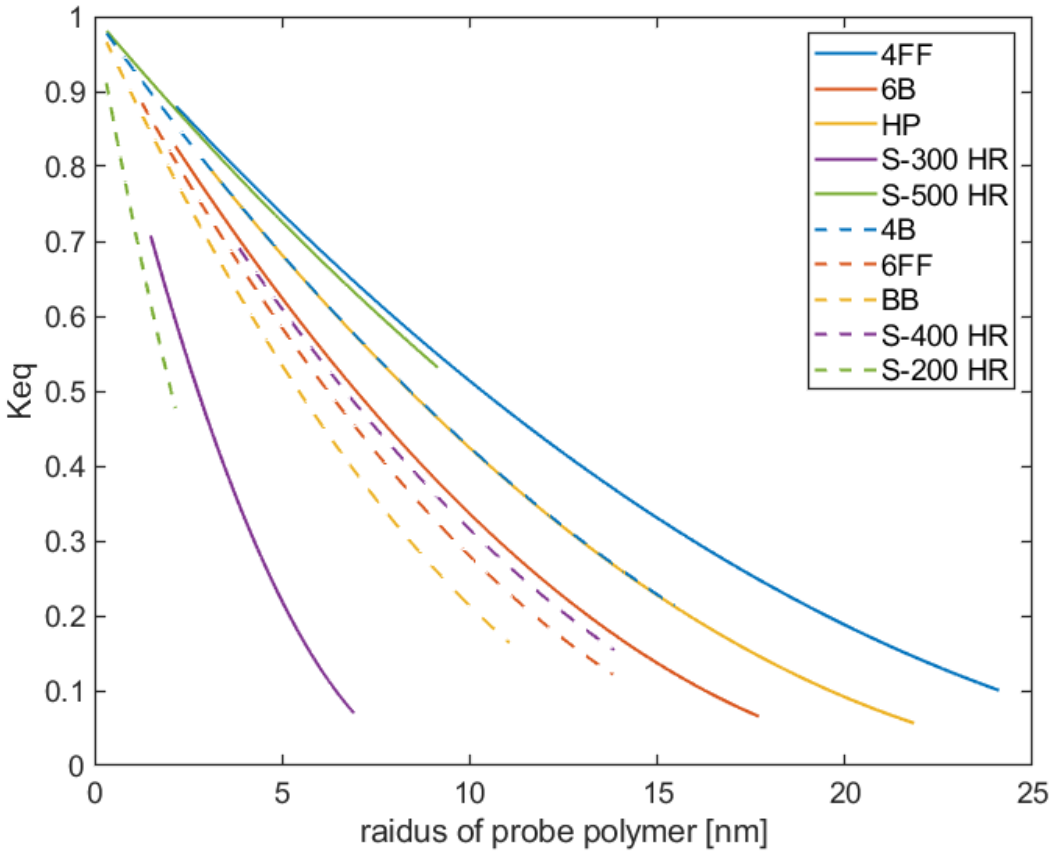

Figure 1. K_{eq} curves of different resins manufactured by Cytiva. The curves are one-pore model fits to data from ISEC measurements, which were provided by Jonny Wernersson at Cytiva R&D.

When investigating the data on the resins 4FF and 6FF in Figure 1 the average pore radius of 4FF and 6FF is 38 and 21 nm, respectively. The 4FF resin had a standard deviation of 2.9 nm while the 6FF resin had a standard deviation of 1.5 nm when comparing different lots of the same resin. The Sephacryl S-200 HR used in this project is to the far left in Figure 1 meaning it has a small pore radius compared to other resins produced by Cytiva. Data on the S-200 HR resin was inadequate since only data from two lots were used but with the numbers available it got an average pore size of 6.6 nm and a standard deviation of 0.53 nm between lots. Figure 1 and the standard deviations indicate that a method suited for pore characterization of all these resins needs to be robust in a wide range of pore sizes. To be able to detect variations of pore structure over lots, the characterization needs to be robust over the possible size range of that resin.

3 Materials and methods

To investigate the performance of SEQ-NMR experiments and to assess their ability to provide reliable results while minimizing experimental time, both simulations and experiments were performed. The simulations focused on how the experimental time could be shortened, and then the results from simulations were tested experimentally. The effects of experimental errors that can occur were also examined through simulations as well as experiments.

3.1 Simulations

The simulations were performed in Matlab. First, the general method of SEQ-NMR simulations is described, and then specific simulations for this project are explained in more detail. The base of all simulations is the general setup from Elwinger *et al*. (2018) and the parameters used are listed Table 1.

Table 1. Parameters used for SEQ-NMR simulations in Matlab. Marked values (*) differ with experiment and instrument, the values used during simulations are the same as in Elwinger *et al.* (2018). The pre-calibrated gradient is the max gradient strength of the instrument in simulation. Gmin and Gmax are what percentage of the maximum gradient strength that is used. V_0 is the added volume of dextran solution, V_{void} is the volume between resin beads and V_{tot} is the total volume in the column.

Parameter	Value
Δ	0.114 s
δ	0.010 s
maximum gradient strength	0.5649 T/m*
g_{min} (%)	1
g_{max} (%)	75.533
gradient points	15
SNR	1000
V_0	1593 µl*
r_p	6 nm
V_{void}	69.5 µl*
V_{tot}	1517.8 µl*

The first step in the simulation was to define a molecular weight vector M needed to calculate the dextran distributions. The molecular weight vector ranges from 180 g/mol, the weight of one monomer, to 5 times the size of the biggest dextran in the mixture. The lognormal distribution of each dextran was calculated by Eq. (4). The value of µ was approximated to $\ln(M_p)$. In previous simulations of this method, see Elwinger *et al.* (2018), SEC information regarding σ of the used dextrans have been available. Since there was no SEC data available for all dextrans used in the simulations in this project, an average of σ was determined from the

previous SEC data. This gave an estimate of σ to be 0.5 except for glucose where σ is equal to zero. All simulations have used this value of σ if not stated otherwise.

M was then translated into a vector of r_H through Eq. (5) where a= 0.029 and α=0.46 from calibrations by Elwinger *et al.* (2018). The r_H vector was then translated into D by Eq. (1) and K_{eq} values were calculated for a one-pore model according to Eq. (6). The next step was to calculate $V_{accesible}$, the volume seen by each component in the mixture. This was done by a combination of Eq. (4) and the relationships between V_{pore}, V_{tot} and V_{void} explained in Section 2.5.1. $V_{accesible}$, V_0 and V_{void} was then turned into a dilution d by Eq. (8), for parameters see Table 1. The dextran distributions were then normalized resulting in the distribution before equilibrium, by multiplying this distribution with d was the distribution after equilibrium obtained. With the calculated distributions were the attenuation curves obtained from Eq. (9) and Eq. (10). Prior to these equations was b calculated according to Eq. (3). By having discrete g values between g_{min} and g_{max} of the maximal gradient strength, see Table 1, was the range of b represented by a corresponding vector.

The points on the selectivity curve were set to nine as in Elwinger *et al.* (2018) distributed on a logarithmic scale of D and translated into hydrodynamic radii through Eq. (1). The fitting for receiving a selectivity curve was then done with 1000 Monte Carlo iterations (MC). At the beginning of each MC loop noise was added to the attenuation curves before and after equilibrium with the resin. The noise follows normal distribution with standard deviation equal to the first attenuation value before equilibrium divided with SNR. The attenuation curves with noise added were then fitted with the Matlab function *fmincon* that finds the minimum of a nonlinear multivariate function using preset constraints (MATLAB 2019). These constraints were, as defined by Elwinger *et al*. (2018), that the larger a molecule is the smaller will $V_{accesible}$ be. The largest molecule in solution cannot be excluded from a larger volume than the pore volume of the resin and the smallest molecule in solution cannot have access to a volume larger than the pore volume of the resin (Elwinger *et al.* 2018).

From fmincon a dilution for each dextran was given as output making it possible to calculate $V_{accesible}$ from Eq. (8) which was then used to calculate K_{eq} from the fitted data with Eq. (7). In each MC loop R^2 was calculated according to Eq. (11) and then a mean value of all calculated R^2 values was given. A 68.3 % confidence interval was calculated for the K_{eq} values according to Alper & Gelb (1990). The confidence interval was later used for plotting error bars in the selectivity curve where K_{eq} is plotted against the nine hydrodynamic radii.

3.1.1 Optimal mixture of dextrans

The optimal dextran mixture was simulated by having six different dextrans in each solution. At Cytiva there were ten dextrans available with M_p values of 1080, 2800, 4440, 9890, 21400, 43500, 66700, 123600, 196300 and 401300 g/mol. Additionally, glucose with M_p value 180 g/mol was available. To enable experimental testing of simulated results the simulations were performed with these dextrans and glucose. With r_p 6 nm, M_p = 123600 g/mol was considered the biggest needed dextran for the mixture with its hydrodynamic radius of approximately

6 nm. This left the simulations to consider eight dextrans plus glucose. All possible combinations of six components in the mixture were simulated, a total of 84 mixtures.

To justify the exclusion of the two largest dextrans, the performance was investigated by simulations with all ten dextrans in solution and with all ten dextrans and glucose. The performance of the method depending on r_p in the range of 4 to 8 nm was also simulated by varying r_p for the optimal mixture found above. The optimal mixture had been obtained with r_p 6 nm and the performance was then simulated between 4 and 8 nm in 20 steps.

3.1.2 The number of gradient points and their distribution

The distribution of *g* in an NMR diffusion experiment determines the distance between measurement points and thereby affects the distribution of *b*. The possibility of reducing the number of measurement points was investigated by simulation together with the distribution in *b*. The different distributions tested were linear with equal spacing between *b* points, squared with a decrease in spacing between points giving equal spacing in b^2 and reversed squared with an increase in spacing between points being the inverse of the squared distribution. These distribution profiles of *b* were simulated for 5 to 30 measurement points giving R^2 as output. With the best performing distribution of *b* the same procedure was simulated but with the SNR varying from 300 to 1500 for determination of the minimum SNR required for a SEQ-NMR measurement.

3.1.3 The robustness of method

The sensitivity of the results to different types of errors and uncertainties during an experiment decides how robust a method is. This was analysed by simulations examining how the performance of the method was affected by introducing errors of 1-3% to important parameters. When using short samples, see Section 3.2.1, a variation in the diameter between NMR tubes can lead to a difference in amount of sample volume in the region of measurement in the instrument. This would affect the attenuated signal and was tested by introducing errors to *E*. Errors due to misplacement of the sample in the probe, see Section 2.3, or an imperfect gradient, see Section 2.6.2, was investigated by introducing a difference between the *b* used before and after equilibrium.

In the above simulations the nine points, nine hydrodynamic radii, have been distributed logarithmically within a limited interval. An alternative would be to manually set the radii of the simulation points to be the radius corresponding to the top of each dextran distribution in the mixture since these points should contain maximum information. This was investigated by simulation using the optimal dextran mixture found in Section 3.1.1. The estimation of σ, the width of the individual dextrans, equaling 0.5 could contain errors and therefore was σ equaling 0.2 and 0.8 evaluated through simulation to analyse the effects on R^2.

3.2 Experiments

All experiments were performed on a Bruker 500 MHz Avance III-HD spectrometer with a 5 mm TXI probe if nothing else is stated. Two sample heights were used, for long samples the NMR tube contained 1 mL solution while short samples contained 160 µL solution. All diffusion experiments used the pulse program *stegp1s* and rectangular pulse shapes. Diffusion measurements were done with the gradient found in Section 3.1.2 with a strength of 1 to 100% if nothing else is stated. All dextran standards were from Pharmacosmos A/S except the one with M_p= 2800 g/ml which came from American Polymer Standards Corporation. The PEO standard of 4×10^6 g/mol was from Scientific Polymer Products, the $CuSO_4$ from Merck Eurolab and the D_2O from Cambridge Isotope laboratories. All NMR experiments had the temperature set to 20 °C.

3.2.1 Evaluation of diffusion NMR measurements

The comparison of diffusion experiments with long or short samples was performed with 5 mM $CuSO_4$ in D_2O. The diffusion measurement had 15 gradient points with 16 scans each, Δ was set to 20 ms and δ to 3.4 ms. As stated in Section 2.3 must the sample be placed in the constant region of the probe corresponding to 1 cm of the sample height. To investigate what the results would be of a misplaced short sample, a measurement was made where the sample was intentionally misplaced by approximately 4 mm. Since the long sample covers much more of the tube height than 1 cm there is no problem with misplacement of the sample.

To know the actual maximum value of the gradient, it needs to be calibrated. This was done with the result from the long sample of 5 mM $CuSO_4$ in D_2O where the observed attenuation had *b* values containing the known diffusion coefficient of HDO molecules, semi-heavy water, in D_2O at 20 °C being 1.621×10^{-9} m^2s^{-1} (Mills 1973).

To check for convection, a short sample of 5 mg/mL PEO of molecular weight 4×10^6 g/mol in D_2O filtered with a 2 µm filter was measured with Δ set as 200, 300 or 400 ms in three separate measurements. To maintain the same *b* in each measurement δ was also altered (see Table 3). The measurements were performed with 32 scans and 10 gradient points for each Δ and δ combination. Diffusion measurements of macromolecules like dextrans require large Δ values. Since Δ cannot be set to infinitely large numbers, δ also had to be increased to maintain the value of *b* and provide a sizeable attenuation. To validate that a δ of 10 ms would work seven experiments of the same PEO in D_2O sample were performed with δ varying between 4 and 10 ms, keeping *b* unaltered by also changing Δ. This was performed with a short sample and 15 gradient points for 8 scans each.

3.2.2 The robustness of method

How experimentally reproducible the method is was examined by repeating five identical measurements in a row. This was done with short samples of 5 mM $CuSO_4$ in D_2O and 5 mg/mL PEO in D_2O. For the PEO sample δ was set to 7.5 ms and Δ to 200 ms. The measurement was done with 10 gradient points for 32 scans each. For the $CuSO_4$ in D_2O, δ was set to 2 ms and Δ

to 50 ms. Here, the diffusion measurement was done twice with the relaxation delay between scans (*D1*) set to 1.5 s and 10 s respectively. For these two measurements the gradient values were set between 1 and 80% of the maximum gradient with 10 gradient points with 16 scans each.

The concentration dependency on self-diffusion was measured with short samples of the optimal dextran mixture obtained from simulations at the concentrations 4, 6, 8 and 10 mg/mL. The diffusion measurements were done with 256 scans, 10 gradient points and with Δ set to 200 ms and δ to 7.5 ms.

3.2.3 Sample preparation and the optimal experiment

With the optimal parameters from simulations and experiments a final SEQ-NMR experiment was made to evaluate the partly new experimental protocol. The sample preparation began with mounting of a PD10 column from Cytiva with lid, filter and stopper. When assembled, 1.6 mL of Milli-Q (MQ) water was added to the column to saturate the filter with liquid. The excess liquid was removed by spin-out centrifugation at 1000g for 1 minute. The column was then weighted to get the mass of the empty column. The Sephacryl S-200 HR resin was also provided by Cytiva and was first washed with MQ water on a glass filter with pore size 4 µm before filling the column with 6 mL of 50 % slurry containing Sephacryl and MQ water. The column was again centrifuged at 1000 g for 1 minute and weighted to establish the mass of the slurry and the agarose resin.

To the assembled column with washed Sephacryl, 1.6 mL 5 % D_2O in MQ water was added. The column was then weighted again to get the exact volume of the added solution. The resin and D_2O solution were then left to equilibrate for 15 minutes on a shaking table at 1100 rpm. The column was then centrifuged as above, and the excess solution was collected. 1 mL of stock solution (before equilibrium) and collected solution (after equilibrium) was added to separate NMR tubes. The resin was then washed in the column with 30 mL of MQ water and then 3 washes with 3 mL of D_2O with vortexing in between each wash. At the last wash with D_2O the solution was removed by centrifuge as above and the column was again weighted. A solution of 5 mg/ml PEO in D_2O was mixed and filtered with a 2 µm filter. 1.6 mL of this solution was added to the column which then was weighed as above to calculate the exact volume of solution added. The procedure for equilibration, centrifugation, weighting and filling of NMR tubes was then performed in the same way as described above.

The dextran mixture was prepared by making separate 3 mg/mL solutions of each dextran in the from simulation optimal mixture with M_p values of 1080, 2800, 9890, 21400, 66700 and 123600 g/mol in D_2O. Then 1 mL from each solution was mixed together to form the test solution. The resin was again washed with MQ water, approximately 60 mL to remove any PEO sample left in the column, and then washed 3 times with 3 mL of D_2O with vortexing in between each wash. The last D_2O wash was removed by centrifugation as described above and the column was weighed. Then 1.6 mL of the prepared test solution was added to the column and it was let equilibrating with the resin for 1 hour on a shaking table at 1100 rpm. The

equilibrated solution was again collected by centrifugation as above. 1 mL of both the stock solution and the equilibrated collected solution were added to separate NMR tubes. After the sample preparations six samples were ready for NMR measurements. All NMR tubes were stored with parafilm wrapped around the lids to avoid evaporation.

Spectra needed for calculating V_{tot} and V_{void} according to section 2.5.1 were recorded on a Bruker 300 MHz Avance III-HD spectrometer with a 5 mm QNP probe. The ^{2}H NMR experiments were performed on the 5 % D_2O in MQ samples and the spectral intensities from before and after equilibrium were used to calculate V_{tot}. Spectral intensities in the ^{1}H spectra of the 5 mg/mL PEO in D_2O samples before and after equilibrium were used to calculate V_{void}. The final diffusion measurements on the test solution before and after equilibrium were as all earlier diffusion experiments recorded on the Bruker 500 MHz with 256 scans and with Δ set to 114 ms and δ to 10 ms.

4 Results

4.1 Simulated results

The results obtained from simulations are divided into three parts where the two first parts consider the optimal mixture and the number of gradient points aiming at improving the method and decreasing the duration of the experiment. The last part present results on how robust the method is.

4.1.1 Optimal mixture of dextrans

The results from simulations regarding the optimal dextran mixture were reviewed both by their visual appearance and performance in terms of R^2. Of 84 possible combinations, 13 performed with an R^2 above 0.99 and were visibly similar. Solution mixture number 69 with the M_p values 1080, 2800, 9890, 21400, 66700 and 123600 g/mol, had the highest R^2 of 0.9941 and was therefore chosen as the optimal mixture. Figure 2 illustrates the simulated performance of mixture 69, the K_{eq} values obtained from SEQ-NMR simulation closely follows the theoretically calculated data. Figures corresponding to all 84 combinations and a table of R^2 values are presented in Appendix A.

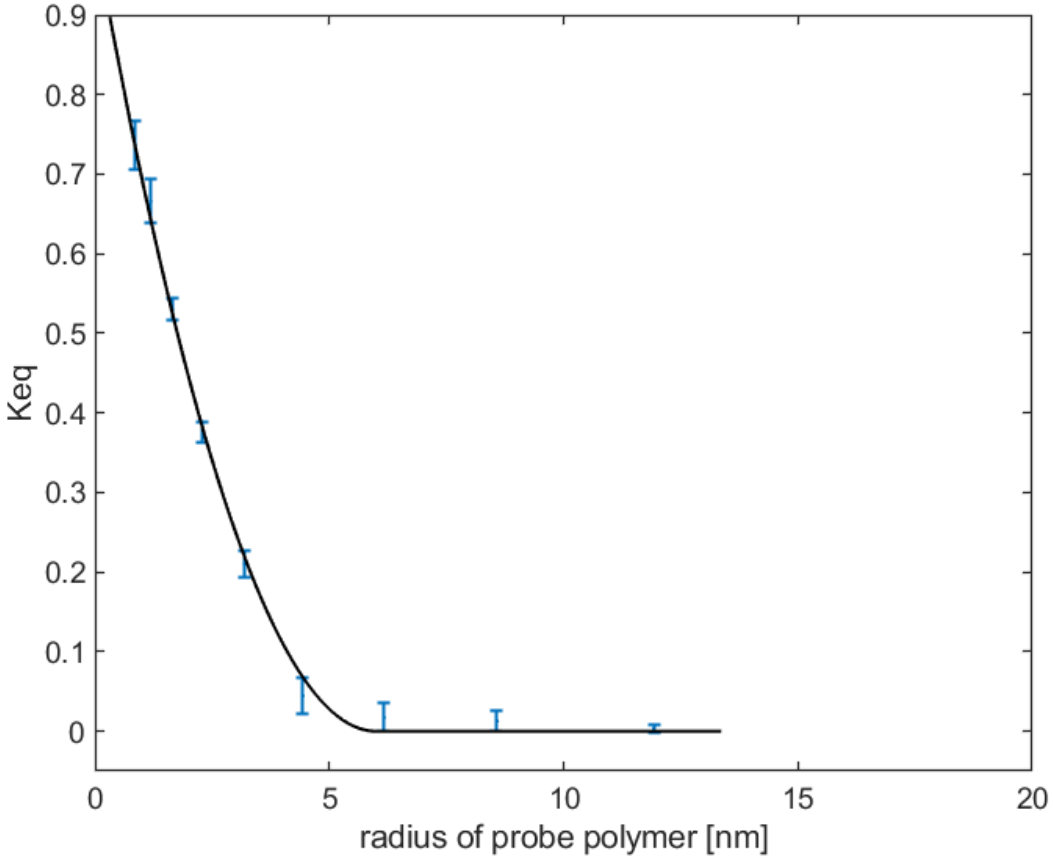

Figure 2. The simulated SEQ-NMR result obtained by the optimal dextran mixture applied to a system with pore radius 6 nm, the known pore radius of a Sephacryl S-200 HR resin. K_{eq} values were sampled in the simulations at nine hydrodynamic radii up to five times the biggest dextran in mixture. The error bars correspond to 68.3 % confidence intervals as provided by MC statistics. The solid line corresponds to theoretical data and the optimal dextran mixture contains *M*p values of 1080, 2800, 9890, 21400, 66700 and 123600.

If all eight dextrans were added to the mixture the R^2 received from simulation was 0.9916 and if glucose was included too, R^2 became 0.9928. Since the last digit in R^2 may vary due to the added noise and therefore be insignificant no improvements could be achieved by adding larger or smaller test molecules to the mixture.

As stated in Section 2.5 it would be desirable for the method to be robust over a range of pore radii making it possible to use the same dextran solution of characterization of a wider range of resins. This was investigated by analyzing the performance of the method using the optimal dextran mixture with r_p ranging from 4 to 8 nm. That range correspond to ±2 nm of the $r_p$used for simulation of the optimal mixture and is more than the lot-to-lot variation for the Sephacryl S-200 HR resin as stated in Section 2.7. The results are presented in Figure 3 as performance depending on r_p and the method clearly performs well with the optimal dextran solution in the range of 4.5-7 nm but quickly worsens beyond that range.

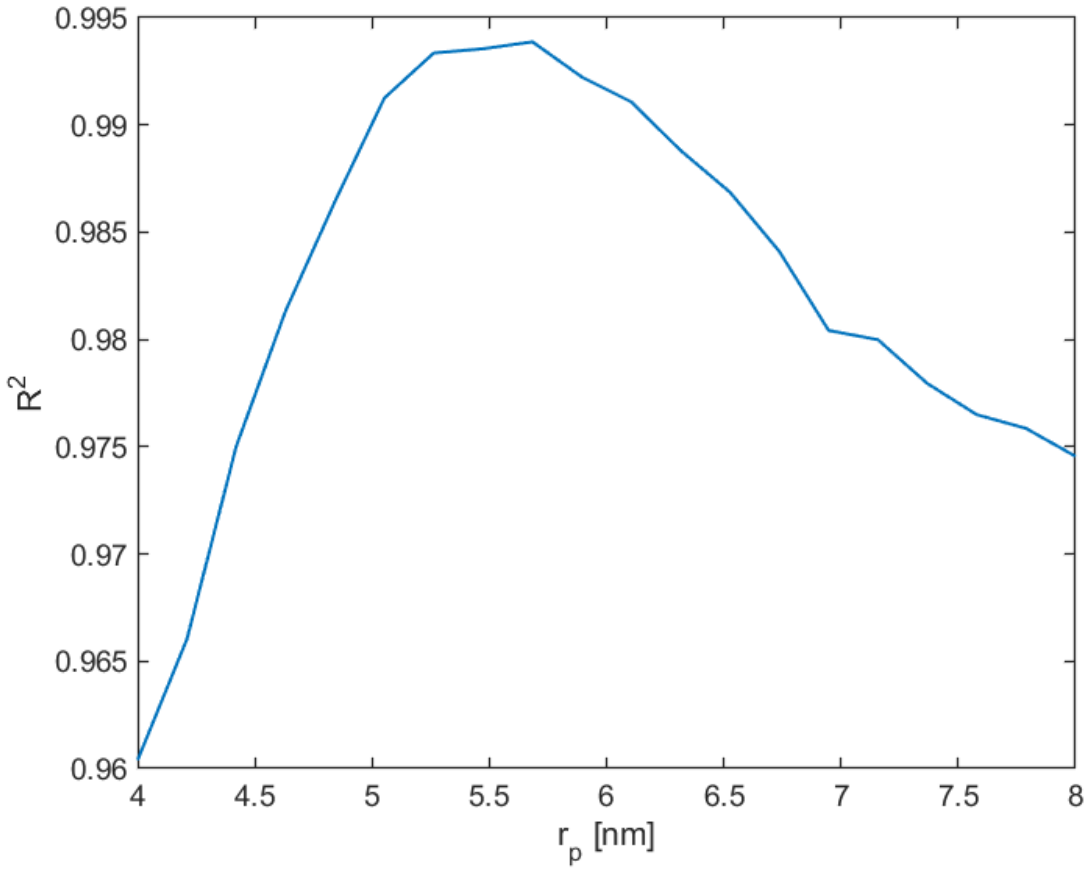

Figure 3. The simulated performance of the method using the optimal dextran mixture for resins with pore radii ranging from 4 to 8 nm. The performance is given in terms of R^2.

4.1.2 The number of gradient points and their distribution

Regarding the distribution of *b* values along the attenuation curve, the squared distribution performed much worse compared to the linear and reversed squared distributions that performed equally good. This is illustrated by Figure 4. Since the reversed squared distribution showed more uncertainty at a low number of gradient points and the linear distribution was available as instrumental default, the linear distribution was chosen for the subsequent simulations and experiments.

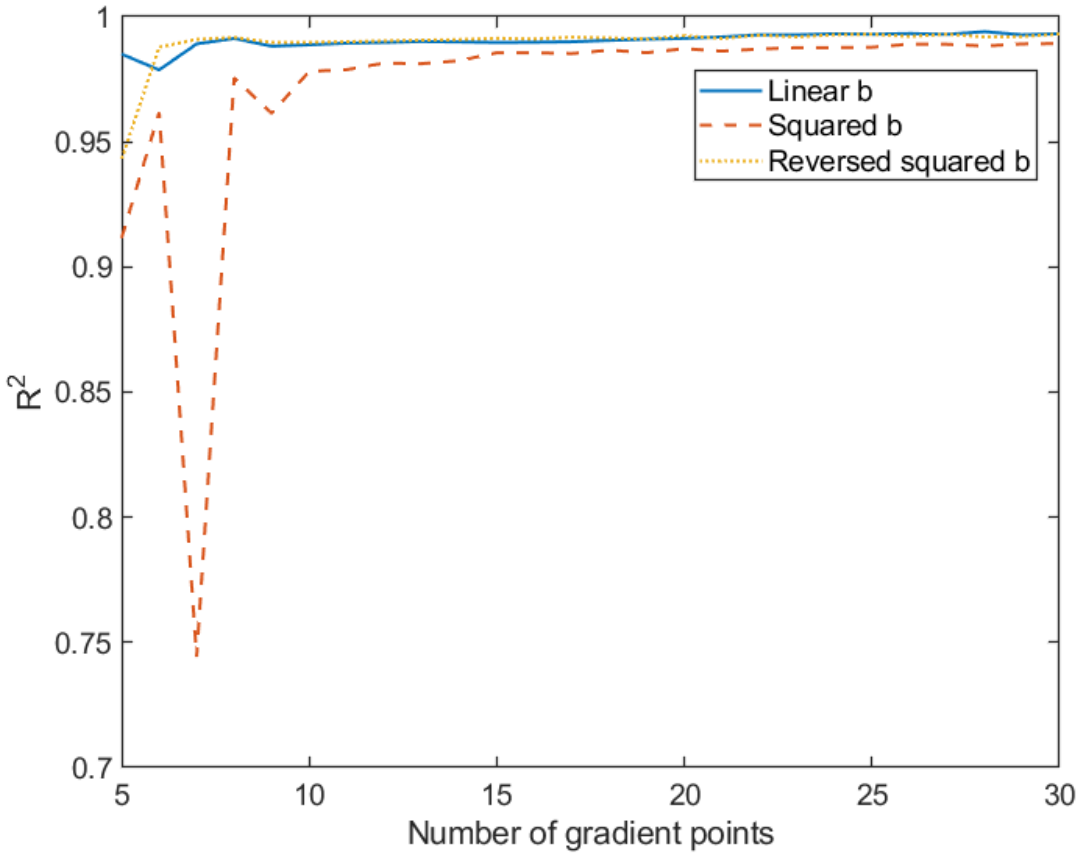

Figure 4. The performance of the method given in R^2 depending on both the number of gradient points used and their distribution along the attenuation curve, see text.

With the best performing linear distribution the simulation in Figure 4 was repeated with varying SNR. The results from this are shown in Figure 5 and contributes with two conclusions, one being that above a given SNR, approximately 1000, there is no longer a rapid gain of performance when increasing the SNR, and second, for lower SNR there is a need for 15 gradient points to reliably provide a high R^2. The following simulations and experiments therefore used 15 gradient points except the optimal experiment in Section 4.2.3 having 10 gradient points. This was chosen because 10 points were shown by Figure 5 to be sufficient when having high SNR. 10 points was also the lowest possible number of gradient points and Figure 4 illustrates that the performance is roughly the same at 10 and 30 points when SNR is 1000. Fewer number of gradient points than 10 would lead to overfitting since the simulations have nine hydrodynamic radii used in the making of the selectivity curve.

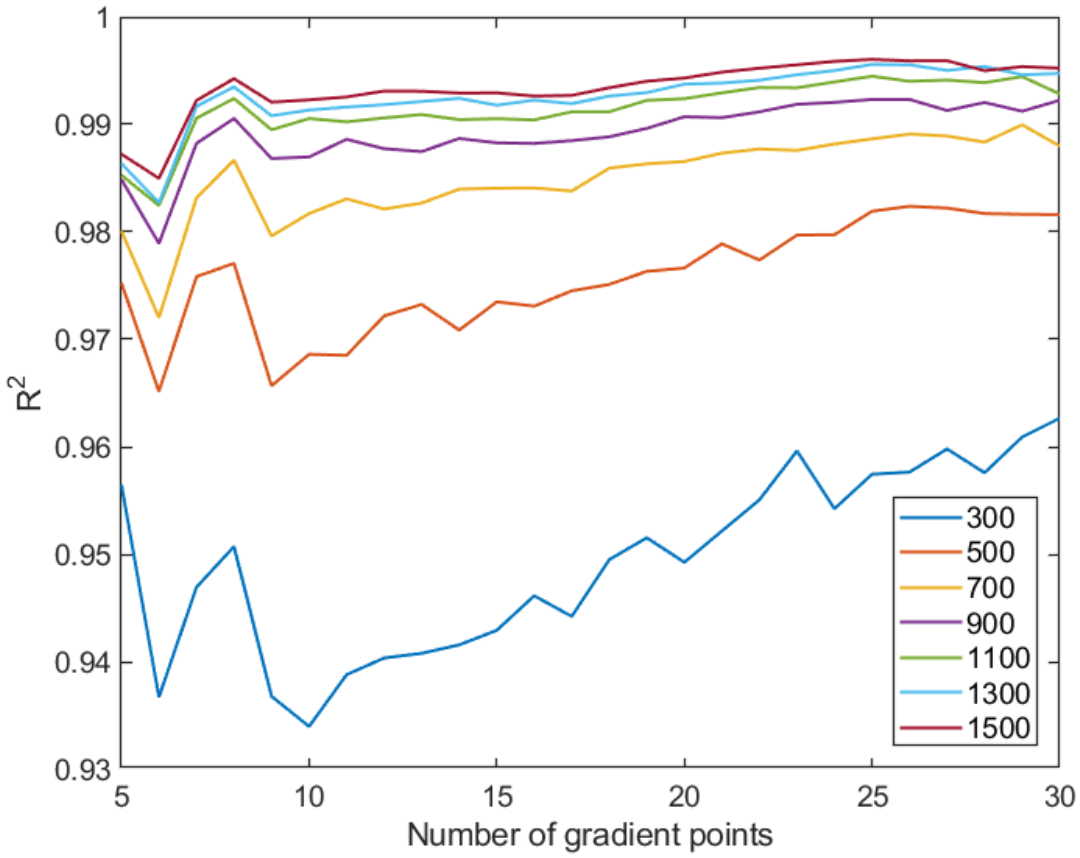

Figure 5. Simulated results of the performance of the method using the linear distribution of *b* with different values of SNR listed in the legend of the figure. The simulations were made with number of gradient points ranging from 5 to 30.

4.1.3 The robustness of method

A robust method is insensitive to bias data which can occur in various ways during experiments. The robustness of the NMR measurements was investigated by introducing plausible errors in important parameters and examining the effect of the biased data on the method through simulations. Additionally the robustness was tested by changes in distribution of the hydrodynamic radii used in the selectivity curve and changes in the width of the size distribution of individual dextran sizes.

The effects of a variation between NMR tube diameters leading to a difference of the sample volume within the measuring region of the probe was analysed by introducing signal attenuation errors, errors in *E*. Gradient mismatch or misplacement of the position of a short sample was introduced as errors in *b* values after equilibrium. Errors in the attenuation signal, *E*, influence the performance at 2% and errors in *b* gives effects on the performance at 1%, see Table 2.

Table 2. Performance of the method in terms of R^2 for the optimal dextran mixture when by simulation introducing errors to the parameters *E* and/or *b*. Errors in *E* corresponds to differences in the diameter of the NMR tubes and errors in *b* corresponds to an imperfect gradient and is dependent on how well the sample is placed in the instrument.

Error in *E* (%)	Error in *b* (%)	R^2
0	0	0.9941
1	0	0.9919
2	0	0.9850
3	0	0.9706
0	1	0.9890
0	2	0.9785
0	3	0.9632
1	1	0.9792
2	1	0.9637
1	2	0.9459

Intuitively the method would perform better if instead of having nine randomly distributed hydrodynamic radii used for making the selectivity curve, the radii corresponded to the top of each individual dextran distribution in the mixture. The result was the opposite, for this new way of expressing the *x*-axis of hydrodynamic radii the R^2 obtained was 0.6688, for figure see Appendix B. If the estimation of σ was incorrect a smaller distribution would mean a worse performance of the method whereas a wider distribution would improve the performance, see Table 3.

Table 3. Performance of the SEQ-NMR method expressed in R^2 from simulation of different sigma determining the size distribution of each dextran in the optimal mixture.

σ	R^2
0.2	0.9859
0.5	0.9941
0.8	0.9950

4.2 Experimental results

The experimental results are divided into three parts where the first part contains experimental tests of the instrument, the second part contain testing of robustness and the third part presents result from what the simulations indicated would be an optimal SEQ-NMR experiment.

4.2.1 Evaluation of diffusion NMR measurements

For the SEQ-NMR measurements, high values of both Δ and δ are needed to have a sizeable attenuation for diffusion of all components in the dextran mixture. To investigate if a δ value of 10 ms would work in a reliable manner with the instrument available, diffusion measurements with 4×10^6 g/mol PEO in D_2O were made with δ varying from 4 to 10 ms. Figure 6 shows that δ values in the range from 5-10 ms give equivalent results but the lowest δ of 4 ms deviates from the other measurements. This is probably because of an imperfect pulse shape caused by the short pulse length, the rise and fall of the pulse were too rapid. *D* was calculated according to Eq. (1) from measurements with the same PEO sample as above but from new measurements with different combinations of δ and Δ. The result is presented in Table 4 proving that with the set conditions there is no detectable convection in the sample.

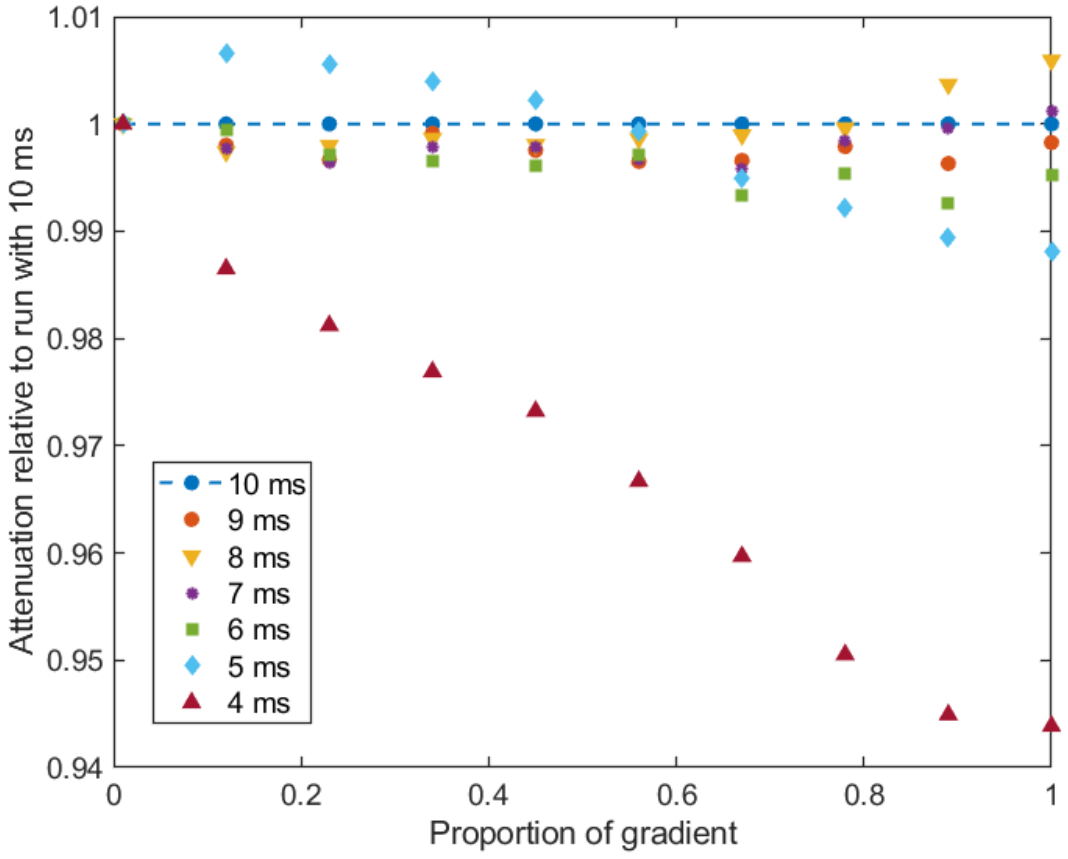

Figure 6. Experimental result of filtered 5 mg/ml PEO in D_2O with different values of pulse duration δ but same value of *b*. Results are presented relative to the run with δ 10 ms for comparison.

Table 4. Diffusion coefficients, *D*, obtained from experimental results of a 5 mg/ml PEO in D_2O sample with different combinations of diffusion time Δ and pulse duration δ.

Δ [ms]	*δ* [ms]	D [10^{-12} m^2s^{-1}]
200	7.5	1.168
300	6.106	1.179
400	5.282	1.190

The performance of the experiments with long and short samples were compared to see potential differences caused by gradient non-linearity. The residual plot in Figure 7 shows that the gradient linearity is similar for both samples and no trends can be distinguished. As stated in Section 2.3 short samples are sensitive to misplacements in the probe. This was tested by misplacing the sample by 4 mm introducing an error in *D* of approximately 3.3%.

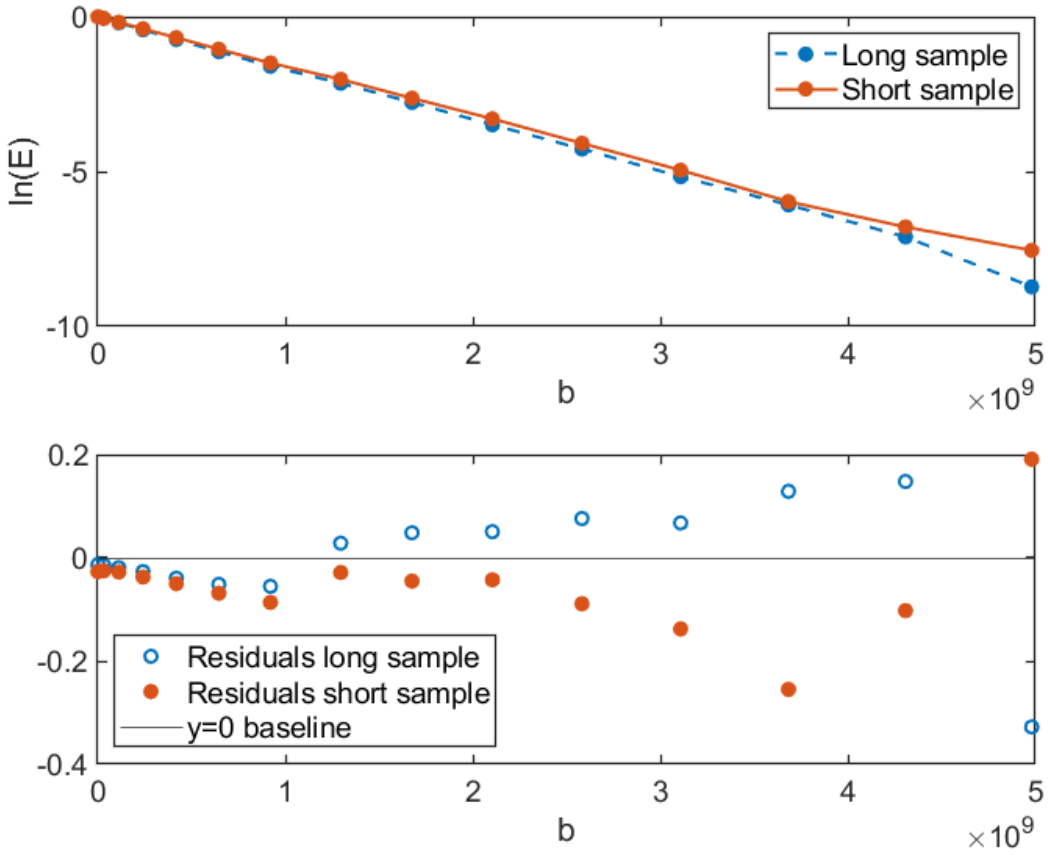

Figure 7. Logarithmic representation of integral values from two experimental runs with 5 mM $CuSO_4$ in D_2O. One tube contained 1000 µl (long sample) and another 160 µl (short sample).

The maximum gradient strength of the instrument was calibrated from diffusion measurements on the long water sample in Figure 7. With *g* as the only unknown variable the slope of the curve in Figure 8 corresponds to $-g^2$ giving the instrument a maximum gradient strength of 0.5765 Tm^{-1}. This calibrated gradient strength is used in the plotting in Section 4.2.3, other results have used the maximum strength from Table 1.

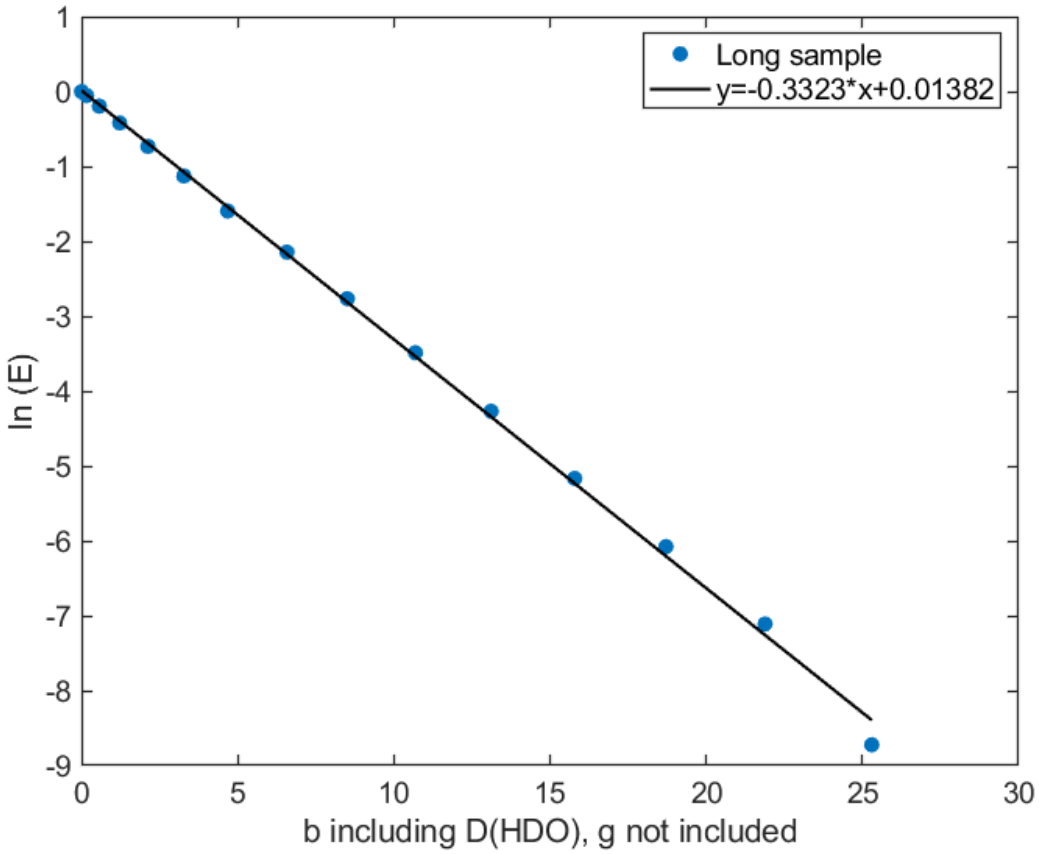

Figure 8. Linear fit to the logarithmic signal attenuation obtained through a diffusional ^{1}H NMR experiment with a long sample (1000 µl) containing D_2O (with trace amounts of $CuSO_4$) for calibration of the maximum gradient of the Bruker 500 MHz Avance III-HD spectrometer with 5 mm TXI probe.

4.2.2 The robustness of method

The reproducibility of the diffusion measurements was tested by repeating five identical experiments of two different samples and the results are presented in Figure 9 as the standard deviation in each measurement point of the repeated experiments. The standard deviation does not increase with increasing *b* indicating the standard deviation is mostly due to the SNR and not an imperfect gradient. This holds true both for the sample with $CuSO_4$ in D_2O and PEO, the upper figure also shows that there is no palpable difference between the runs with *D1* 10 and 1.5 s. The size of the standard deviations is in milli-scale indicating a high reproducibility of the measurements both with $CuSO_4$ having a small δ value as for PEO having a large δ value. Since the standard deviation is correlated to the SNR of the experiment, see Section 2.3, Table 5 shows SNR from Figure 9 and the manually obtained values of SNR. The manual values are higher since the gradient variation is excluded there but included in the SNR given from Figure 9. From Table 5 it is also clear that the PEO sample have much lower SNR compared to $CuSO_4$ in D_2O no matter if the SNR is taken manually or from the standard deviation.

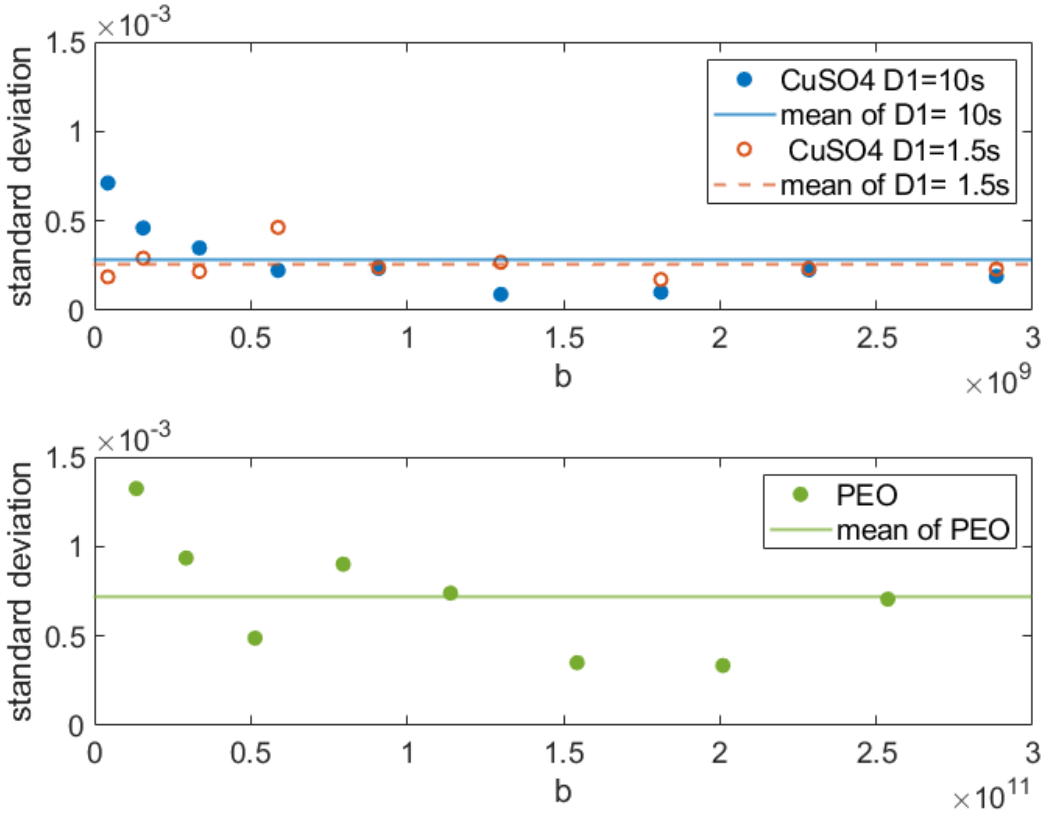

Figure 9. Standard deviation in each measurement point for repeated experiments of two short samples of 5 mg/ml filtered PEO in D_2O and 5 mM $CuSO_4$ in D_2O. Each experiment was repeated five times in a row. The experiment with $CuSO_4$ in D_2O was repeated in two ways, with D1 equaling 10 or 1.5 seconds. The top solid line corresponds to the mean std value of 2.87×10^{-4}, the dashed line 2.56×10^{-4} and the bottom solid line 7.22×10^{-4}.

Table 5. SNR from Figure 9 according to Section 2.3 and manually from NMR spectra for three measurements with two short samples, 5 mM $CuSO_4$ in D_2O and 5 mg/ml PEO in D_2O.

Experiment	SNR (Figure 9)	SNR (manually)
5 mM $CuSO_4$ in D_2O, *D1*= 1.5 s	3904	6658
5 mM $CuSO_4$ in D_2O, *D1*= 10 s	3479	4491
5 mg/ml PEO in D_2O	1384	3177

Diffusion measurements of the dextran mixture at the different concentrations, 4, 6, 8 and 10 mg/ml were made to test if the dilution could be increased without having molecular size influence the intermolecular interactions. From each experiment *D* was calculated according to Eq. (1) and was plotted at corresponding concentration in Figure 10.

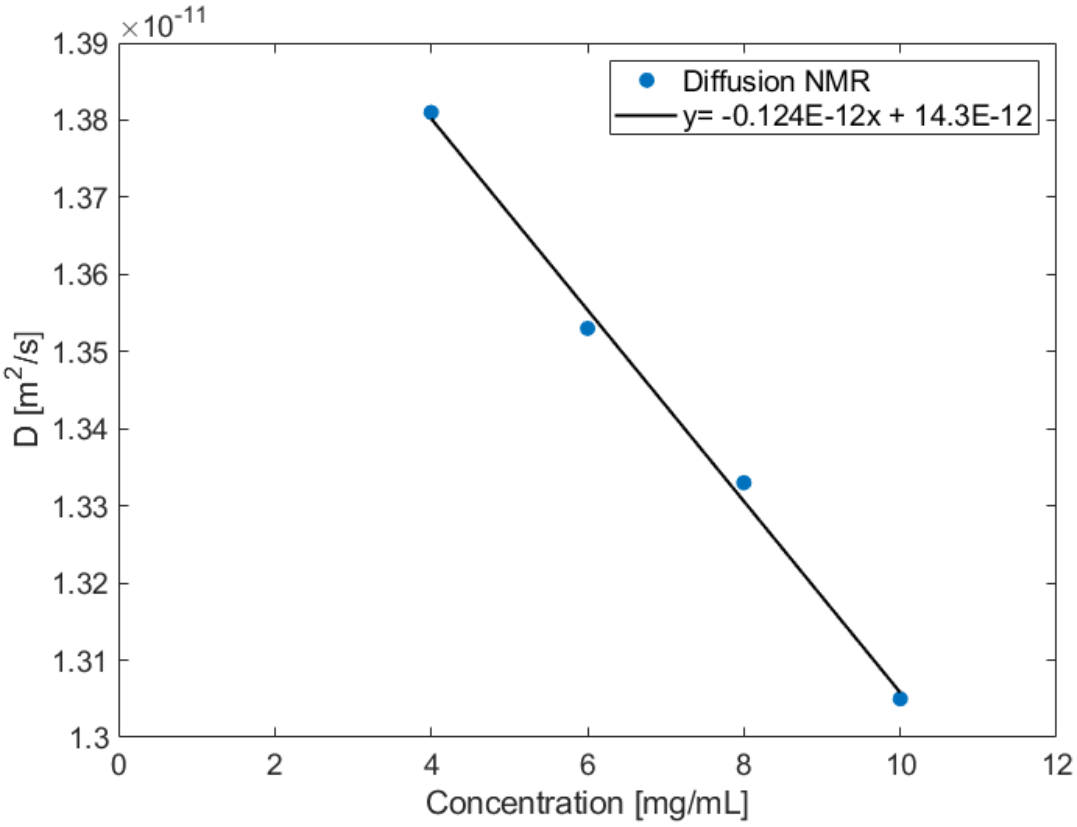

Figure 10. Diffusion measurement for four concentrations of the optimal dextran mixture. *D* for the concentrations (solid dots) have been fitted with a straight solid line.

The straight line in Figure 10 corresponds to the linear approximation $D = D_0(1 - kc)$, where D_0 represent the diffusion coefficient of a molecule in nothing but solvent, c is the concentration and k is a coefficient of dilution (Furukawa *et al.* 1991). From Figure 10, k is given as 0.00867 Lg^{-1}, Furukawa *et al.* (1991) got a value of 0.0186 Lg^{-1} for a single dextran with a molecular weight of 150 kDa. When the sample is diluted during equilibrium in a SEQ-NMR experiment the concentration after equilibrium is approximately 60 % of the stock solution. Since the solution before equilibration has a concentration of 3 mg/mL the concentration after equilibration will be 1.8 mg/mL. The concentration after equilibrium multiplied by k from Figure 10 equals 0.0156 meaning that D will change by approximately 1.6 % from before to after equilibration. It was shown in Table 2, Section 4.1.3, that 1 % errors in the attenuation will not affect the final performance of the method in a noticeable way. When the errors exceed 2 % the effect on performance becomes more visible. With increasing concentration will the change in percentage of D also increase.

4.2.3 Optimal parameters

The volumes calculated from measurements with PEO and D_2O gave a V_{tot} of 1779 µl and a Vvoid of 185.8 µl accordingly to Section 2.5.1. The obtained selectivity curve relative to ISEC data is presented in Figure 11, the SEQ-NMR data points follow the ISEC data up to approximately 5 nm of the probe polymer and then starts to deviate. SNR of the optimal measurement was manually estimated as described in Section 2.3 and was approximately 3600.

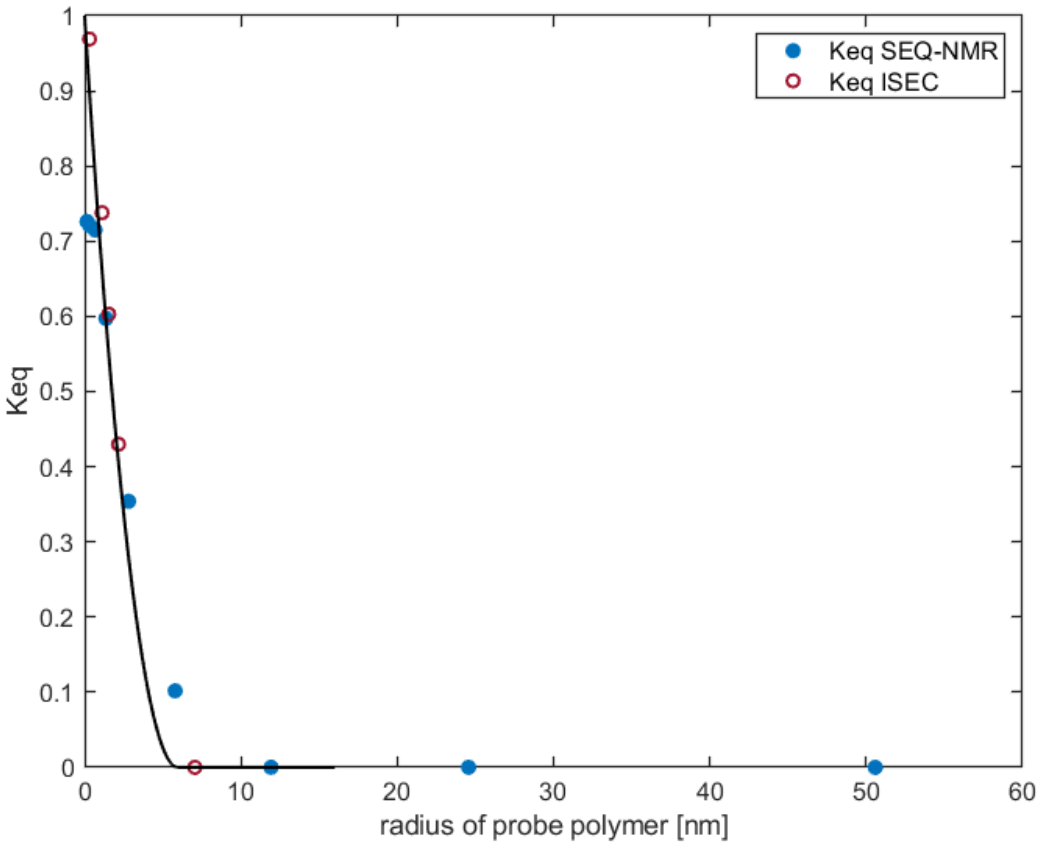

Figure 11. Experimental SEQ-NMR results with optimal parameters shown as blue dots and ISEC data from the same type of resin but from a different lot are shown as red dots. The one-pore model fit to the SEQ-NMR data corresponds to the solid line.

5 Discussion

The SEQ-NMR simulations led to insights on how to improve the experimental performance of the method and returned valuable information on the sensitivity of the results to errors. The first result from simulations were that different mixtures of dextrans perform with varying outcome. If limited to the dextrans available at Cytiva an optimal mixture could be determined having the highest R^2 of all possible combinations, see Figure 2. This optimal mixture was used in all later simulations and diffusion experiments. All simulations ran with 1000 Monte Carlo

iterations to avoid variations in results from identical simulations. Since normally distributed noise is added to the attenuation curves in the simulations there is still a possibility of variations of the fourth significant decimal between runs with 1000 Monte Carlo iterations. Within the spread of values 1000 Monte Carlo iterations permits the optimal dextran mixture to perform nearly equivalent to more than 10 other mixtures. The components of the best performing mixtures indicate that the size range of the included components is more important than the actual components of the mixture, see Table A1 in Appendix A.

The comparison of long and short NMR samples showed a similar gradient linearity for both samples, see Figure 7. The intuition was that short samples would experience a more linear gradient, and even though the SNR is less for short samples it was decided to perform the following experiments with a short sample. Measurements with short samples of PEO and the dextran mixture showed a large overlap with the water peak at low gradient strength. The overlap was due to bad shimming, and made the data processing a difficult process. With the shimming of short samples being troublesome, and the long samples having less overlap with water the decision fell on using long samples in the final experiment with optimal parameters. This decision was made even though short samples were used in many of the preparatory measurements. The residual plot in Figure 7 supports this choice showing that the gradient was linear also for long samples. Since the shimming of short samples is a time-consuming process that decreases the applicability of the method in quality control procedures, the conclusion is that long sample formats are most promising for the future work with SEQ-NMR.

Previous results by the SEQ-NMR method presented in Elwinger *et al*. (2018) showed a very good overlap between SEQ-NMR and ISEC data. The final SEQ-NMR measurement had all optimized parameters, the optimal dextran mixture (Figure 2), 10 gradient points with linear *b* (Figure 4) and a long sample. Despite that simulations indicated improvements of the method, the optimal measurement did not provide the expected result, see Figure 11. Instead, under the assumption that the previous results with good overlap between SEQ-NMR and ISEC data was not accidental, the method slightly underperforms. The results by Elwinger *et al.* (2018) were recorded at the Royal Institute of Technology (KTH) with another NMR instrument having a by far better probe for diffusion measurements and a higher maximum gradient strength. Thus the instrument at Cytiva will inevitably have a poorer instrumental performance. As was discussed in Section 2.6 there are several factors giving rise to errors in diffusional attenuation which have been ruled out by experiments. There is no thermal convection (Table 4), only negligible eddy currents due to the modern instrument and no imperfect gradient pulses (Figure 6 and Figure 8). Since no experimental errors could be found the instrument at Cytiva may be functional for SEQ-NMR measurements. It is possible that the unexpected result in Figure 11 are due to a combination of several small errors, making the recorded diffusional attenuations have slightly worse quality compared to those recorded at KTH.

The reproducibility was proven to be high since the standard deviations at each measurement point for repeated experiments in Figure 9 show no trend as a function of *b*. This indicates that the effect of an imperfect gradient is not larger than the SNR, but since the values of the SNR

deviates from the manual value, see Table 5, no conclusion regarding the gradient independence can be made. An interesting way to continue the development of the method would be to repeat the optimal measurement on the instrument at KTH, since it has proven to work well for SEQ-NMR measurements in the past.

The reason for the slightly worse performance of SEQ-NMR discussed above can also be due to other errors while implementing the method. The uncertainty of the diameter of the NMR tubes and potential differences in temperature inside the sample is not normal error sources of diffusion NMR but they arise in SEQ-NMR, since intensities before and after equilibrium are compared. For a reliable comparison the environment of the sample must be the same when measuring before and after equilibrium. Small changes affecting the attenuation of one sample will affect the resulting selectivity curve making SEQ-NMR very sensitive to external and internal errors. If the estimation of σ involved any errors it would not affect the method enough to alone explain the result of the optimal measurement, see Table 3.

To investigate if the optimal measurement was affected by differences in the diameter between the two NMR tubes used before and after equilibrium, the samples could be measured in the same tube with careful and precise washing and drying of the tube in between measurements. Alternatively, a molecule of known reference peak and concentration could be added to both samples. Then the difference in peak area of the reference molecule in the samples before and after equilibrium could be used to compensate for the difference in diameter between the tubes. The NMR tubes used in this project were from Wilmad-LabGlass and the specification of their inner diameter is 4mm ±0.026 mm which will give the cross sectional area, and hence the volume in the tube, a variation of 1.3%. Up to 1% errors in the sample volume, tested as 1% error in the attenuation in Table 2, there was no noticeable change in the performance of the method. Since errors of 1.3% can occur is it not possible to rule out the uncertainty of the NMR tubes as the reason for the non-reliable result in Figure 11, but it is unlikely the only reason contributing to the results.

One major drawback with the method as it works today is its robustness regarding the pore size of the resin examined. The simulations used a pore radius of 6 nm which is a good estimation for the Sephacryl S-200 HR resin used in this project. In Section 2.7 it is stated that for Sephacryl S-200 HR resin lots the standard deviation is 0.5 nm. Figure 3 shows the optimal mixture in Figure 2 to be useful for all lots of this resin but if the pore size starts to change more than ±2 nm the performance of the method quickly worsens. This indicates that for the method to be able to characterize multiple resin types there will be a need of having one optimal dextran mixture for each pore radius. All optimal mixtures would also have to be tested for their robustness, as in Figure 3, since the 4FF resin in Section 2.7 have a standard deviation of 2.9 nm, which is a range that the optimal mixture of S-200 HR would not manage. From the investigation of the optimal mixture shown in Figure 2, the conclusion can be drawn that the included dextrans in an optimal mixture must range from very small up to approximately two times the radius of the pore of the resin. If including the tails of the individual dextran distributions. By scaling the optimal mixture according to these results it can be predicted what

mixture to use for a resin with known mean value of its pore size. The limiting factor of this strategy is that very large dextran standards are needed for characterization of the resins with the biggest pore radius and that kind of material is not currently available at Cytiva.

There could be improvements made regarding the method of data fitting if first a pore-model is fitted for normally distributed pore radii giving mean value and standard deviation as output parameters. Thus constraining the fit to only two parameters. Then all dilutions can be calculated from that fitting, and when having obtained the dilutions, starting values that will be more similar to the real solution can be calculated. These starting values can then be used to initialize a subsequent unconstrained fitting that is model-free. This could give more reliable values and it could possibly also improve the robustness dependent on changes in pore radius demonstrated in Figure 3.

Choosing to have long samples in the optimal measurements increased the SNR. As stated in Section 4.2.3, the SNR of the measurement was by manual calculation 3600. The manually calculated SNR does not include the gradient variation as illustrated by Table 5. An estimate of the real SNR from the information in Table 5 would be that the optimal measurement had a SNR of approximately 2000. Figure 5 shows that the performance of the method stops increasing notably at a SNR of approximately 1000. Since SNR is scaled by the square root of number of scans, the number of scans could be lowered from 256 to 64 reducing the experimental time while maintaining a SNR of 1000. The experimental time could also be shortened by an increase in concentration of the dextran solution, but this has proven to not be permissible for SEQ-NMR. As indicated by Figure 10, increasing the concentration from 3 mg/mL to 4 mg/mL could change *D* by more than 1%, which clearly will affect the outcome of the selectivity curve.

One of the main goals of this master thesis project was to reduce the duration of a SEQ-NMR experiment to make it suitable for quality control analysis. With the optimized number of gradients and dextran mixture the method takes approximately four hours for one diffusion measurement. Including the measurements of V_{tot} and V_{void}, the method takes just above eight hours, which is an improvement of 22 hours. Despite the big reduction in time, SEQ-NMR is still not the right choice if the Quality Control department at Cytiva is in need of a pore characterization method that takes less than one hour.

Improvements on the duration of the method are still possible. With both SNR and concentration being set parameters for the method, further decrease in experimental time can be achieved by modification of the relaxation time. This modification could be done by adding a relaxation agent, which would mean shorter durations of each scan and make it possible to increase the SNR by increasing the number of scans. Relaxation agents are paramagnetic ions that are added to the sample to get a reduction in relaxation time (Stilbs 2019). $CuSO_4$ used for some of the control experiments is such an agent. If the results regarding limited choices in concentration and SNR had been known at an early stage of this project, the work would most probably been directed more towards such paramagnetic solutions. This makes the decrease of

diffusion NMR experimental time to one hour still possible. An alternative method for pore size characterization in quality control could be Dynamic Light Scattering (DLS). With the same sample preparations as for SEQ-NMR the diffusion coefficients can be calculated from the time autocorrelation given by the intensity of the scattered light. DLS could be a much faster method but it measures mutual diffusion and not self-diffusion. By using dilute solutions can the mutual diffusion be approximated to self-diffusion.(Stilbs 2019).

6 Conclusion

The overall examination of SEQ-NMR has proven the method to have high reproducibility in measurements. The method is also sensitive to surrounding errors and its robustness needs further investigation. It is unfortunate that the final result with optimal parameters gives the method a less bright future for application in quality control procedures. The duration of a SEQ-NMR experiment has been reduced by 22 hours but still takes 8 hours of diffusion measurements to perform. With the possibility to lower the number of scans from 256 to 64 together with a relaxation agent and reference molecule for the NMR tube diameter, the method would most likely be able perform in less than one hour. Even without being a one-hour method, SEQ-NMR can with further improvements be suitable for pore size characterization if the required time and knowledge is provided.

7 Acknowledgements

This project would not have been possible without the support and knowledge of my supervisor Fredrik Elwinger and co-supervisor Istvan Furo. You have guided me through this project and I am grateful for your contributions. I would also like to thank Katarina Edwards for proofreading my report and supporting me throughout the project. At last I would like to thank the whole group of Analytical Technologies and the rest of R&D at Cytiva for welcoming me and making these months both fun and inspiring.

References

Alper JS, Gelb RI. 1990. Standard errors and confidence intervals in nonlinear regression: comparison of Monte Carlo and parametric statistics. The Journal of Physical Chemistry 94: 4747–4751.

Bruice PaulaY. 2010. Essential Organic Chemistry, 2nd ed. Pearson Education, Inc., United States of America.

Chang K-H. 2015. Chapter 10 - Reliability Analysis. In: Chang K-H (ed.). e-Design, pp. 523–595. Academic Press, Boston.

Claridge TDW. 2009. Diffusion NMR spectroscopy. Tetrahedron Organic Chemistry Series, pp. 303–334. Elsevier.

Connell MA, Bowyer PJ, Adam Bone P, Davis AL, Swanson AG, Nilsson M, Morris GA. 2009. Improving the accuracy of pulsed field gradient NMR diffusion experiments: Correction for gradient non-uniformity. Journal of Magnetic Resonance 198: 121–131.

Elwinger F, Wernersson J, Furó I. 2018. Quantifying Size Exclusion by Diffusion NMR: A Versatile Method to Measure Pore Access and Pore Size. Analytical Chemistry 90: 11431–11438.

Furukawa R, Arauz-Lara JL, Ware BR. 1991. Self-diffusion and probe diffusion in dilute and semidilute aqueous solutions of dextran. Macromolecules 24: 599–605.

GE Healtchcare Life Sciences. 2018. Size Exclusion Chromatography -Principles and Methods.

Guo X, Laryea E, Wilhelm M, Luy B, Nirschl H, Guthausen G. 2017. Diffusion in Polymer Solutions: Molecular Weight Distribution by PFG-NMR and Relation to SEC. Macromolecular Chemistry and Physics 218: 1600440.

Hore PeterJ. 2015. Nuclear Magnetic Resonance, 2nd ed. Oxford Chemistry Primers, Oxford.

Hyberts SG, Robson SA, Wagner G. 2013. Exploring signal-to-noise ratio and sensitivity in non-uniformly sampled multi-dimensional NMR spectra. Journal of Biomolecular NMR 55: 167–178.

Knox JH, Ritchie HJ. 1987. Determination of pore size distribution curves by size-exclusion chromatography. Journal of Chromatography A 387: 65–84.

Knox JH, Scott HP. 1984. Theoretical models for size-exclusion chromatography and calculation of pore size distribution from size-exclusion chromatography data. Journal of Chromatography A 316: 311–332.

Kuz'mina NE, Moiseev SV, Krylov VI, Yashkir VA, Merkulov VA. 2014. Quantitative determination of the average molecular weights of dextrans by diffusion ordered NMR spectroscopy. Journal of Analytical Chemistry 69: 953–959.

MATLAB. 2019.The MathWorks Inc., Natick, Massachusetts.

Mills R. 1973. Self-diffusion in normal and heavy water in the range 1-45.deg. The Journal of Physical Chemistry 77: 685–688.

Price WS. 1997. Pulsed-field gradient nuclear magnetic resonance as a tool for studying translational diffusion: Part 1. Basic theory. Concepts in Magnetic Resonance 9: 299–336.

Price WS. 2009. NMR Studies of Translational Motion: Principles and Applications. Cambridge Molecular Science, Cambridge.

Smith G. 2015. 10 - Multiple Regression. In: Smith G (ed.). Essential Statistics, Regression, and Econometrics (Second Edition), pp. 301–337. Academic Press, Boston.

Stilbs P. 2019. Diffusion and Electrophoretic NMR, 1st ed. Walter de Gruyter GmbH, Leck.

Topgaard D, Martin RW, Sakellariou D, Meriles CA, Pines A. 2004. 'Shim pulses' for NMR spectroscopy and imaging. Proceedings of the National Academy of Sciences 101: 17576–17581.

Willis SA, Stait-Gardner T, Torres AM, Price WS. 2016. Chapter 2. Fundamentals of Diffusion Measurements using NMR. In: Valiullin R (ed.). New Developments in NMR, pp. 16–51. Royal Society of Chemistry, Cambridge.

Appendix A

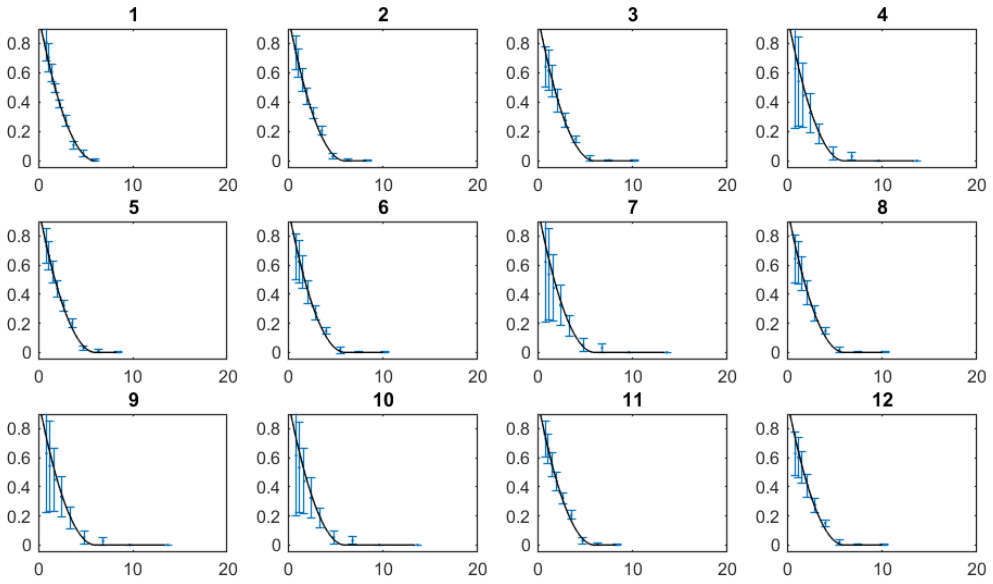

Figure A1. Simulation of dextran mixtures 1-12, performance and dextrans included in each mixture are found in Table A1. The error bars correspond to 68.3 % confidence intervals as provided by MC statistics and the solid line corresponds to theoretical data.

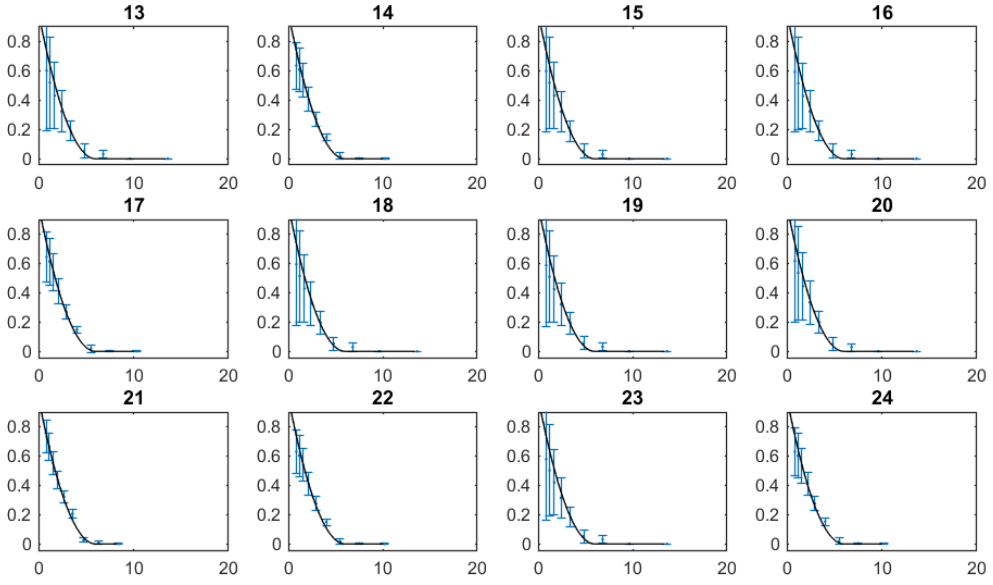

Figure A2. Simulation of dextran mixtures 13-24, performance and dextrans included in each mixture are found in Table A1. The error bars correspond to 68.3 % confidence intervals as provided by MC statistics and the solid line corresponds to theoretical data.

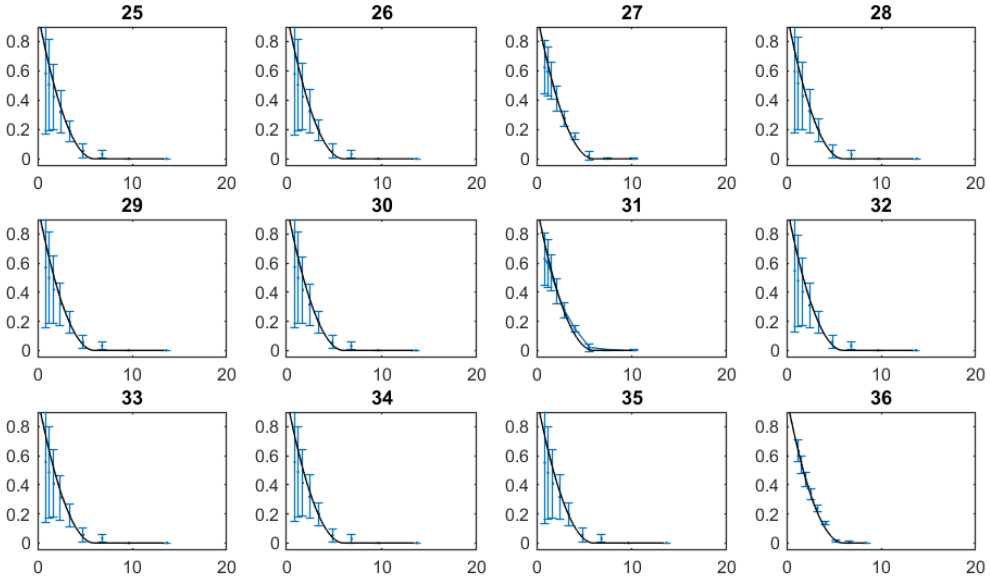

Figure A3. Simulation of dextran mixtures 25-36, performance and dextrans included in each mixture are found in Table A1. The error bars correspond to 68.3 % confidence intervals as provided by MC statistics and the solid line corresponds to theoretical data.

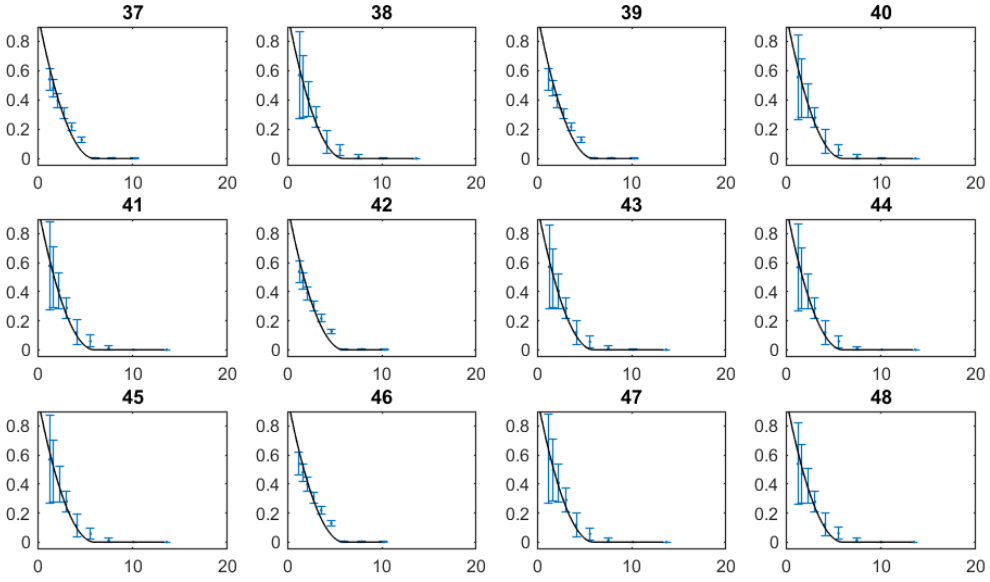

Figure A4. Simulation of dextran mixtures 37-48, performance and dextrans included in each mixture are found in Table A1. The error bars correspond to 68.3 % confidence intervals as provided by MC statistics and the solid line corresponds to theoretical data.

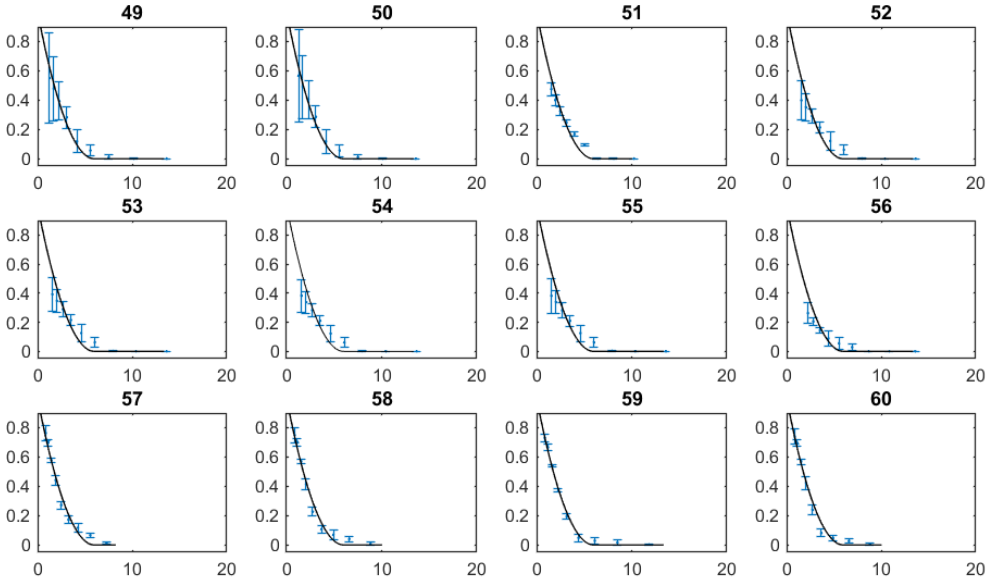

Figure A5. Simulation of dextran mixtures 49-60, performance and dextrans included in each mixture are found in Table A1. The error bars correspond to 68.3 % confidence intervals as provided by MC statistics and the solid line corresponds to theoretical data.

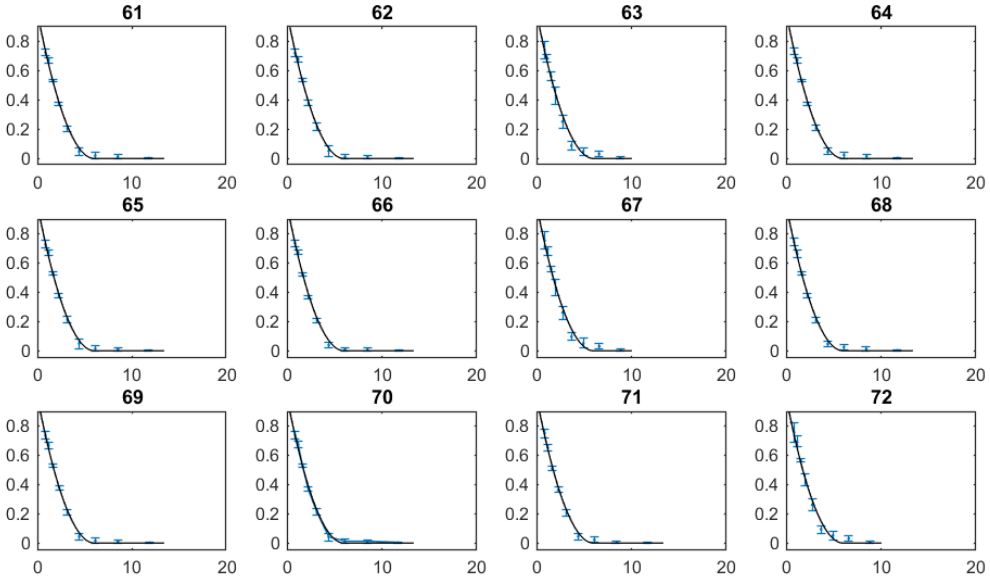

Figure A6. Simulation of dextran mixtures 61-72, performance and dextrans included in each mixture are found in Table A1. The error bars correspond to 68.3 % confidence intervals as provided by MC statistics and the solid line corresponds to theoretical data.

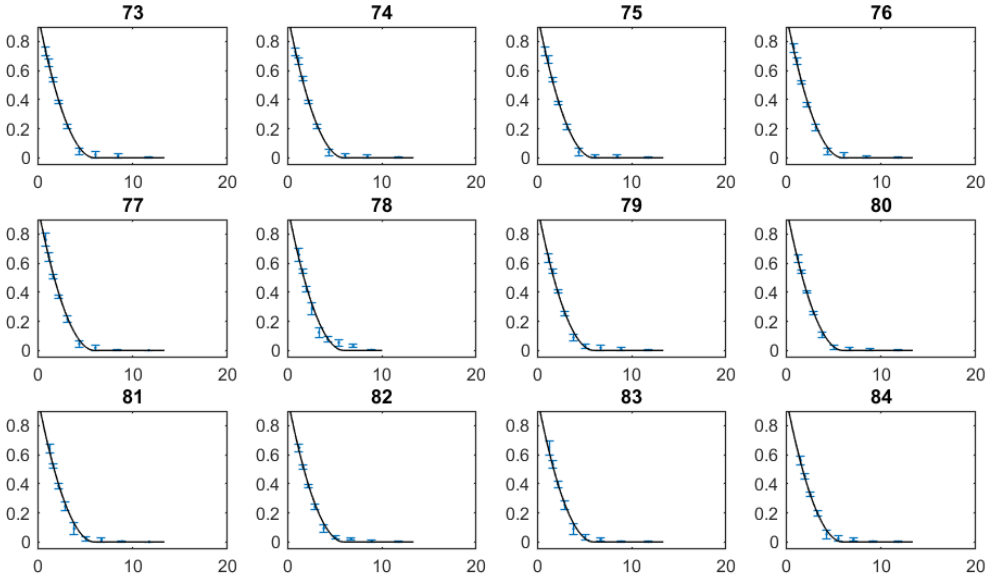

Figure A7. Simulation of dextran mixtures 73-84, performance and dextrans included in each mixture are found in Table A1. The error bars correspond to 68.3 % confidence intervals as provided by MC statistics and the solid line corresponds to theoretical data.

Table A1. R^2 values of all 84 dextran mixtures used in the simulation of the optimal mixture for SEQ-NMR experiments. The simulations were made in Matlab and the table is sorted with the highest R^2 at the top, the mixture number is for indexing and the included dextrans in each mixture is presented with their M_p values.

R^2	mixture number	M_p values of dextrans in mixture [g/mol]					
0.9923	69	1080	2800	9890	21400	66700	123600
0.9922	66	1080	2800	4440	43500	66700	123600
0.9921	73	1080	4440	9890	21400	43500	123600
0.9920	80	2800	4440	9890	21400	66700	123600
0.9919	64	1080	2800	4440	21400	43500	123600
0.9919	68	1080	2800	9890	21400	43500	123600
0.9919	70	1080	2800	9890	43500	66700	123600
0.9917	74	1080	4440	9890	21400	66700	123600
0.9914	76	1080	4440	21400	43500	66700	123600
0.9911	75	1080	4440	9890	43500	66700	123600
0.9909	65	1080	2800	4440	21400	66700	123600
0.9906	71	1080	2800	21400	43500	66700	123600
0.9904	61	1080	2800	4440	9890	43500	123600
0.9895	59	1080	2800	4440	9890	21400	123600
0.9883	62	1080	2800	4440	9890	66700	123600
0.9877	77	1080	9890	21400	43500	66700	123600

R^2	mixture number	M_p values of dextrans in mixture [g/mol]					
0.9865	79	2800	4440	9890	21400	43500	123600
0.9848	82	2800	4440	21400	43500	66700	123600
0.9841	81	2800	4440	9890	43500	66700	123600
0.9827	84	4440	9890	21400	43500	66700	123600
0.9788	83	2800	9890	21400	43500	66700	123600
0.9769	67	1080	2800	9890	21400	43500	66700
0.9758	72	1080	4440	9890	21400	43500	66700
0.9746	60	1080	2800	4440	9890	43500	66700
0.9746	63	1080	2800	4440	21400	43500	66700
0.9721	58	1080	2800	4440	9890	21400	66700
0.9654	57	1080	2800	4440	9890	21400	43500
0.9589	78	2800	4440	9890	21400	43500	66700
0.9424	36	180	2800	4440	9890	21400	43500
0.9325	1	180	1080	2800	4440	9890	21400
0.8978	51	180	4440	9890	21400	43500	66700
0.8960	39	180	2800	4440	9890	43500	66700
0.8887	37	180	2800	4440	9890	21400	66700
0.8837	21	180	1080	4440	9890	21400	43500
0.8790	42	180	2800	4440	21400	43500	66700
0.8732	11	180	1080	2800	9890	21400	43500
0.8731	46	180	2800	9890	21400	43500	66700
0.8724	5	180	1080	2800	4440	21400	43500
0.8676	2	180	1080	2800	4440	9890	43500
0.7598	3	180	1080	2800	4440	9890	66700
0.7464	6	180	1080	2800	4440	21400	66700
0.7368	12	180	1080	2800	9890	21400	66700
0.7353	14	180	1080	2800	9890	43500	66700
0.7170	22	180	1080	4440	9890	21400	66700
0.7125	8	180	1080	2800	4440	43500	66700
0.7008	24	180	1080	4440	9890	43500	66700
0.6918	17	180	1080	2800	21400	43500	66700
0.6507	27	180	1080	4440	21400	43500	66700
0.6502	31	180	1080	9890	21400	43500	66700

R^2	mixture number	M_p values of dextrans in mixture [g/mol]					
0.3957	56	180	9890	21400	43500	66700	123600
0.1742	38	180	2800	4440	9890	21400	123600
0.1727	52	180	4440	9890	21400	43500	123600
0.1508	54	180	4440	9890	43500	66700	123600
0.1113	41	180	2800	4440	9890	66700	123600
0.0802	43	180	2800	4440	21400	43500	123600
0.0722	53	180	4440	9890	21400	66700	123600
0.0522	44	180	2800	4440	21400	66700	123600
-0.0507	40	180	2800	4440	9890	43500	123600
-0.0537	45	180	2800	4440	43500	66700	123600
-0.0885	55	180	4440	21400	43500	66700	123600
-0.2346	50	180	2800	21400	43500	66700	123600
-0.2567	47	180	2800	9890	21400	43500	123600
-0.4570	49	180	2800	9890	43500	66700	123600
-0.6248	48	180	2800	9890	21400	66700	123600
-3.1948	4	180	1080	2800	4440	9890	123600
-3.4580	9	180	1080	2800	4440	43500	123600
-3.8687	10	180	1080	2800	4440	66700	123600
-4.0559	7	180	1080	2800	4440	21400	123600
-4.0561	13	180	1080	2800	9890	21400	123600
-4.2317	15	180	1080	2800	9890	43500	123600
-4.3754	16	180	1080	2800	9890	66700	123600
-4.4004	20	180	1080	2800	43500	66700	123600
-4.6481	25	180	1080	4440	9890	43500	123600
-4.7346	19	180	1080	2800	21400	66700	123600
-4.7446	18	180	1080	2800	21400	43500	123600
-4.7904	28	180	1080	4440	21400	43500	123600
-4.9013	23	180	1080	4440	9890	21400	123600
-5.0164	26	180	1080	4440	9890	66700	123600
-5.0486	29	180	1080	4440	21400	66700	123600
-5.2113	30	180	1080	4440	43500	66700	123600
-5.8449	34	180	1080	9890	43500	66700	123600
-5.9126	35	180	1080	21400	43500	66700	123600

R^2	mixture number	M_p values of dextrans in mixture [g/mol]					
-6.1030	32	180	1080	9890	21400	43500	123600
-6.3766	33	180	1080	9890	21400	66700	123600

Appendix B

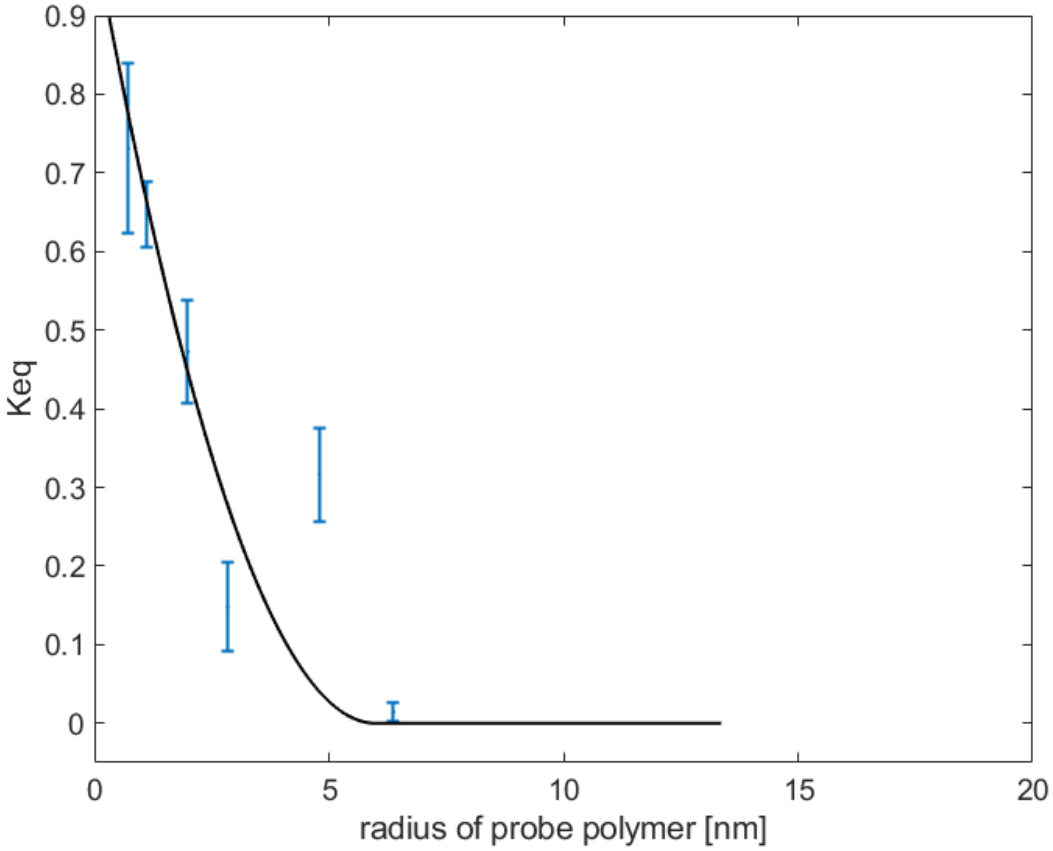

Figure B1. Performance of the SEQ-NMR method tested by simulation were the radii used for fitting are the radii on top of each distribution of dextran included in the optimal mixture. The optimal mixture contains M_p values of 1080, 2800, 9890, 21400, 66700 and 123600. The error bars correspond to simulated SEQ-NMR data and the solid line corresponds to the theoretical data. R^2 of the fitting is 0.6878.