metadata, citation and similar papers at core.ac.uk

Accepted Manuscript

Title: What NMR can do in the biopharmaceutical industry

Authors: Róbert Kiss, Ádám Fizil, Csaba Szántay Jr.

PII: S0731-7085(17)31357-2
DOI: http://dx.doi.org/doi:10.1016/j.jpba.2017.07.004
Reference: PBA 11381

To appear in: *Journal of Pharmaceutical and Biomedical Analysis*

Received date: 26-5-2017
Revised date: 30-6-2017
Accepted date: 1-7-2017

Please cite this article as: Róbert Kiss, Ádám Fizil, Csaba Szántay, What NMR can do in the biopharmaceutical industry, Journal of Pharmaceutical and Biomedical Analysishttp://dx.doi.org/10.1016/j.jpba.2017.07.004

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

What NMR can do in the biopharmaceutical industry

Róbert Kiss[1a], Ádám Fizil[1b], Csaba Szántay Jr.[1a*]

[1]Chemical Works of Gedeon Richter Plc.,
[1a]Spectroscopic Research Department
[1b] Analytical Department of Biotechnology

*To whom correspondence should be addressed:
Chemical Works of Gedeon Richter Plc., Hungary, Budapest, 10. P.O.B. 27. H-1475.
e-mail: cs.szantay@richter.hu. Fax: 36-1-432-6003. Phone: 36-1-431-4952

Highlights:

- NMR; biopharmaceuticals; therapeutic proteins, biosimilarity, bioanalytics

Abstract

Nuclear magnetic resonance (NMR) spectroscopy has a unique capability to probe the primary and higher order molecular structure and the structural dynamics of biomolecules at an atomic resolution, and this capability has been greatly fortified over the last five decades by an astonishing NMR instrumental and methodological development. Because of these factors, NMR has become a primary tool for the structure investigation of biomolecules, spawning a whole scientific subfield dedicated to the subject. This role of NMR is by now well established and broadly appreciated, especially in the context of academic research dealing with proteins that are purified and isotope-labeled in order to facilitate the necessary sophisticated multidimensional NMR measurements. However, the more recent industrial development, manufacturing, and quality control of biopharmaceuticals provide a different framework for NMR. For example, protein drug substances are not isotope-labeled and are present in a medium of excipients, which make structural NMR measurements much more difficult. On the other hand, biotechnology involves many other analytical requirements that can be efficiently addressed by NMR. In this respect the scope and limitations of NMR are less well understood. Having the non-expert reader in mind, herein we wish to highlight the ways in which modern NMR can effectively support biotechnological developments. Our focus will be on biosimilar proteins, pointing out certain cases where its use is probably essential. Based partly on literature data, and partly on our own hands-on experience, this paper is intended to be a guide for choosing the proper NMR approach for analytical questions concerning the structural comparability of therapeutic proteins, monitoring technology-related impurities, protein quantification, analysis of spent media, identification of extractable and leachable components, etc. Also, we focus on critical considerations, particularly those coming from drug authority guidelines, which limit the use of the well-established NMR tools in everyday practice.

Keywords: NMR; biopharmaceutics; therapeutic proteins, higher order structure, biosimilarity, bioanalytics

1. Introduction

In the last two decades the global biopharmaceuticals market has increased dramatically. Accordingly, several new companies specializing entirely in the research, analytics, or production of biopharmaceuticals have appeared, and most of the big traditional pharmaceutical companies have opted for introducing biopharmaceuticals at a growing ratio in their product portfolio.

For small-molecule drug substances and drug products the analytical techniques that are obligatory for their detailed structural characterization and quality control are well consolidated and are generally well known (at least at a conceptual level) by the non-analytical experts. However, in the field of industrial biotherapeutics the analogous demands on bioanalytics are not only rapidly changing, requiring a constant adaptation to the regulatory requirements [1], but form a certain "gray zone" within which companies may interpret what kinds of analytical methodologies, what level of instrumentation, and what level of experimental detail they want to, or are willing to employ. Making the right corporate-level decisions in that regard is certainly not easy. One important aspect of this problem is that the strategies for the structure investigation of biological macromolecules differ from those used for small molecules. This is not only because of the difference in molecule size and structural complexity, but also because the concept of a "structure" is less well-defined in biomolecules. For example, in large proteins the higher-order structure is a dynamical feature of the molecule that can change subtly depending on the circumstances. Also, because the production and formulation of biologics is very different from traditional drugs, there are small average differences between batches, and specific differences between individual molecules, in terms of the post-translational modifications of a protein. This basic difference between small and large molecules is also reflected in the terms "generic" and "biosimilar", as used for the respective follow-on drug substances. While for generic small molecules the original and the follow-on drug substance must be analytically demonstrated to have exactly the same structure, for biosimilars the follow-on substance must be shown to be sufficiently "similar" to the original, and the latter task calls for new analytical approaches. Furthermore, the function of bioanalytics covers not only the structural characterization of macromolecules, but quite often the application of a state-of-the-art analytical background is required to support the midstream and downstream biotechnological processes of drug production as well. Within this new, dynamic and diverse landscape of biopharmaceuticals the role of nuclear magnetic resonance (NMR) spectroscopy is also diverse, differs in many respects from its common applications, and is not well understood in general.

With the above considerations in mind, in this review we wish to highlight the unique role that NMR spectroscopy can play regarding the quality control and regulatory aspects of biotechnological products, with special focus placed on the emerging analytical questions that can be answered by NMR during the development of protein biosimilars. In addition to reviewing the most relevant recent literature, our discussion is guided by our own hands-on experience gained in bioanalytics over the last decade in the region's prominent pharmaceutical company, Gedeon Richter Plc.

To clarify the scope of this review, we first need to point out the circumstances that can limit the range of applicable NMR tools in the biopharmaceutical industry. In the discovery phase of protein drug substances, NMR spectroscopy is one of the major tools for the characterization of solution-phase macromolecular structure and dynamics. This is because at this stage isotope-labeled (i.e., enriched in the NMR-active nuclei ^{13}C and/or ^{15}N),

proteins are used, which is a prerequisite for the large number of sophisticated multidimensional NMR techniques that may be employed for probing a structure at atomic resolution [2]. Under these circumstances the researchers are mostly free to choose the scheme of isotopic enrichment, the solution conditions that are the most advantageous for the studied molecular system, and thus the optimal NMR method for addressing the pertinent structural question. In contrast, this methodological repertoire becomes much more limited in the industrial development phase. At this stage biologics must to be analyzed with natural isotopic abundance and very often in the presence of other components that may hamper the NMR investigations. For biosimilars the initial characterization of the drug substance and the drug product, followed by the comparability studies involving their originator counterpart, must be performed without any modification of the solution. Other types of analytical tasks that NMR can provide at certain steps of the biotechnological procedures (e.g. spent broth analysis, or the structural characterization of unknown contaminants, or protein quantification) are also preferably carried out with minimal modification of the sample. Due to these special conditions, the biggest challenge for the NMR spectroscopist in the biopharmaceutical industry is, on the one hand, to recognize the analytical question that can be answered by NMR, and on the other hand to creatively chose among the applicable techniques from the vast set of the NMR toolbox.

There are some concerns that may be raised against using NMR for macromolecular characterization in an industrial environment, such as its low sensitivity compared to other techniques, most typically mass spectrometry (MS), whereby much more substance must be used for the measurements which can also take excessively long times, and the requirement for using costly state-of-the-art instrumentation. Herein we wish to demonstrate that although these drawbacks of NMR are true, the information that NMR can provide is in many cases inaccessible by any other method. Moreover, the constant improvements in NMR technology keep alleviating – at least in partly – these limiting attributes. Although NMR is not the dominant analytical technique in the biopharmaceutical industry, its use as a complementary method is either useful, or sometimes inevitable. The growing number of scientific papers in this field also indicates that NMR is rapidly gaining an increasingly solid foothold in the arena of biopharmaceuticals.

2. Structure investigation of therapeutic proteins

2.1 NMR in the context of the regulatory requirements for biosimilars

As already noted in the Introduction, analogously to the generic drug substances, biologics also have their follow-on versions, usually referred to as biosimilars. Due to the high structural complexity, and in some respects inherent structural diversity of therapeutic proteins, the exact identity of the follow-on and the originator's drug substance (which is itself a structurally diverse entity with some batch-to-batch variations) cannot be achieved in practice. Moreover, apart from the fact that the primary structures of the relevant proteins must be identical, with respect to other structural features the criteria for "similarity" are more difficult to define, and the pertinent regulatory requirements are changing as the field evolves. According to the FDA, biosimilars are "highly similar to the reference product they were compared with, but have no clinically meaningful differences in terms of safety, purity and potency from the reference product"[1]. The criteria for what differences between the originator and the generic products are, or are not acceptable, cannot be uniformly defined, because not all of the molecular features are equally relevant regarding the biological activity of a biopharmaceutical drug substance. Thus, the critical "similarity-parameters" must be established for each individual case. Although this task is challenging and leaves a certain

space for judgment for the developers of a biosimilar molecule, if done wisely, meticulously and scientifically convincingly, it can forcefully demonstrate the quality of the biosimilar product.

A question often raised in connection with the application of NMR in comparability studies is whether the employed method should be validated. NMR methods are considered as state-of-the-art structure-elucidation tools which, according to recent recommendations [3], should not necessarily be validated, but must be based on sound science. If an NMR method has already been described in a monograph, one only needs to verify its suitability for the employed conditions; validation is required only when an NMR method is used as an alternative to the official procedure. [3,4]

One of the most difficult tasks of biosimilar comparability studies is the monitoring of the protein higher order structure (HOS) during drug development so as to prove that any batch-to-batch variability in the product is within acceptable limits (just as with the orginator's molecule). The most commonly applied analytical methods in comparability studies, such as liquid chromatography, different types of electrophoresis, bioassays, MS and peptide mapping give detailed information mainly about the covalent structure of the protein (i.e., the amino acid sequence, posttranslational modifications such as disulfide bridges, glycosylation, etc.), together with other related structural features such as the intact molecular mass, charge, and aggregation, but tell us nothing or very little about the HOS. Further complementary methods like proton-deuterium exchange mass spectrometry (HDX-MS), Fourier transform infrared spectroscopy (FT-IR), circular dichroism (CD), and differential scanning calorimetry (DSC) are capable of providing some insight into the higher order protein structure as well, but this information is indirect or sensitive to global differences in the secondary and tertiary structure (CD, FT-IR, DSC), therefore structural differences cannot be localized. Only HDX-MS gives sufficiently detailed information about the local changes in protein conformation and conformational dynamics [5–7], but at the expense of labor-intensive sample preparation and with some difficulties in terms of reproducibility. Thus, when it comes to HOS, NMR is in principle a unique player on the scene, with its special capability of providing highly exact and detailed information about the three-dimensional (3D) structure at atomic resolution.

However, the detailed NMR characterization of the 3D structure of formulated therapeutic molecules is quite challenging. This is because without isotopic labeling, and in the presence of the buffer and stabilizers used for the formulation of a protein drug substance, only significantly fewer NMR experiments can be used, making resonance assignment and *de novo* structure determination practically unattainable. Fortunately, for the purposes of the comparability and quality-control studies in question such detailed NMR characterization of the solution structure is not imperative, and quite a lot of useful information can be acquired from experiments that use the most common element in organic compounds, which also happens to be the most sensitively detectable nucleus in NMR, namely ^{1}H. Unlike in CD and IR spectroscopy, in the ^{1}H NMR spectrum of a protein the signals originate from the individual atoms, all reflecting their own local chemical environment by their chemical shift. The chemical shift is a highly sensitive function of the constitution, configuration and conformation of the molecule, and the ^{1}H chemical shifts differ slightly, but usually measurably (depending on the spectral resolution) for each individual ^{1}H nucleus within a protein. Thus, envisioning an idealistic situation where each ^{1}H NMR signal of a protein is fully resolved, the spectrum is essentially an atomic-level spectral fingerprint which uniquely encodes the constitution together with the HOS of the protein. In reality this of course is not the case, and signal overlaps often "smear" certain regions of the "fingerprint" (Figure 1.) Nevertheless, even in such cases the ^{1}H NMR spectrum typically provides enough atomic-resolution detail that will allow one to make a fast and very reliable comparison of the

constitutional and HOS-level likeness of two "similar" protein drug substances. Should a one-dimensional (1D) ^{1}H NMR spectrum prove not to be sufficiently detailed for that purpose, two-dimensional (2D) NMR spectra may be recorded in which the resonances are represented as cross-peaks that carry information about various through-bond or through-space atomic connections within a molecule. These 2D spectra are generated via a sophisticated sequence of radio-frequency pulses, the theory and application of which is a subfield of NMR in its own right, and is often referred to as "spin gymnastics". In the present context, the most important such 2D experiments are as follows. a) The ^{1}H-^{1}H NOESY (nuclear Overhauser effect spectroscopy) experiment shows the ^{1}H-^{1}H pairs that are in close spatial proximity (only a few Ångströms apart); b) the ^{1}H-^{13}C and ^{1}H-^{15}N HSQC (heteronuclear single quantum correlation spectroscopy) or HMQC (heteronuclear multiple quantum correlation spectroscopy) experiments reveal the one-bond H-C and H-N moieties in the molecule. Each experiment gives a specific pattern of cross-peaks (a 2D fingerprint) which is, again, unique for the given protein in terms of its primary and higher order structure. These 2D fingerprints have a much higher resolution (particularly the heteronuclear versions) than the 1D ^{1}H NMR fingerprint. It should be noted that ^{13}C and ^{15}N can be found in a relative natural abundance of 1.3 and 0.4 %, respectively, which still allows the recording of heteronuclear 2D experiments, albeit at a sensitivity which is approximately two orders of magnitude less than for uniformly isotope-labeled proteins.

In the scientific literature, publications describing the above concepts started to appear only relatively recently. For the biosimilar version of the protein drug substances filgrastim and rituximab, 1D ^{1}H NMR spectra were compared visually with those of the respective original products as a part of a comparability exercise. [8,9] Freedberg [10] suggested the use of ^{1}H-^{1}H NOESY as an additional method for fingerprinting. The NOESY experiment was applied to prove the structural identity of biosimilar proteins to the reference product of filgrastim [11] and rituximab [12].

As mentioned above, the ^{1}H NMR structural fingerprinting of formulated therapeutic proteins can be difficult because of the extra signals due to the excipients. Such additives are usually small organic molecules and their molar concentration is significantly higher than that of the protein drug substance, thus producing intensive signals which often overlap with important regions of the protein signals in the 1D ^{1}H or homonuclear 2D spectra, limiting the available fingerprint information. A possible way to work around this difficulty is to do a complete buffer exchange for both the original and the "similar" protein before NMR measurement. This act may induce some minor alterations in the 3D structure of the drug substance, but since this change is expected to affect the reference and the follow-on proteins identically, technically the obtained NMR fingerprints will be relevant as a basis of the comparison. Care must be taken however with this approach, because the fact that a modified version of the reference product is used for the comparison may be questionable from the point of view of the authorities.

Another solution for the suppression of additive signals was suggested by Poppe and coworkers [13], who exploited the difference in the diffusion properties of the small excipient molecules and the large protein. By implementing the principles of the so-called diffusion ordered spectroscopy (DOSY) NMR experiment which can separate signals as a function of their translational diffusion coefficient, they could successfully eliminate the intensive excipient signals from 1D ^{1}H NMR spectra. Their method was employed for different monoclonal antibody (mAb) samples, which are the largest therapeutic proteins, although it works well for smaller proteins. Feng and coworkers [14] used so-called first-increment NOESY-based experiments in NMR comparability studies involving the protein drug substances daptomycin and trastuzumab. Franks and coworkers [15] proposed an interesting

1D ^{1}H fingerprint-enhancing method that can be used to remove the small-molecule excipient signals, together with signals that are due to the highly flexible regions of the protein, from the signals associated with the more rigid parts of the protein. The method exploits the fact that small molecules, as well as the highly flexible parts of large molecules, have very different relaxation characteristics from the more rigid parts of large molecules. One important aspect of this difference is that in the latter case a phenomenon known as spin diffusion is far more efficient than in the former cases. Due to this efficient spin diffusion, the saturation (i.e., the quenching of a signal by using a suitably selective and long irradiation) of any part of the spectrum that is associated with the rigid region of the protein will saturate all other signals as well that originate from that region (note that small molecules and the flexible protein parts do not show this effect). Thus, if one subtracts a spectrum obtained by using such saturation from a control spectrum recorded without saturation, one obtains a spectrum that will give a fingerprint of the more rigid parts of the protein, essentially eliminating the excipient signals. By another approach which was suggested by Skidmore et al. [16], the protein signals can be filtered out from the spectra using the so-called Carr-Purcell-Meiboom-Gill (CPMG) pulse sequence, which works on the basis of the large difference between how fast a protein and a small molecule rotates in solution. This method [16] proved to be more useful for the characterization of small molecules in the presence of macromolecules and will be further discussed in Chapter 2.1.

With increasing molecular weight, the ^{1}H NMR spectrum not only becomes more crowded, but the resonances also become broader, and beyond a certain protein size the two effects start to limit the use of the method as a diagnostic molecular fingerprint. The higher the magnetic field, the larger the so-called Larmor frequency with which NMR-active spins precess within the magnetic field. A magnet's strength is commonly defined as the approximate ^{1}H Larmor frequencies associated with that given magnet. With increasing magnetic field strengths, the resonance frequencies will spread over a larger frequency span, yielding increased spectral resolution. At the currently available high-end magnetic fields (700MHz - 1000MHz), the practical protein size limit for ^{1}H based NMR experiments is approximately 200kDa.

As mentioned above, in order to resolve heavy signal overlapping, 2D heteronuclear correlation methods can be used. A 2D ^{1}H-^{15}N HSQC spectrum will be less crowded, because it only shows cross-peaks that represent the H-N units in the molecule, with each cross-peak belonging to one amino acid residue (Figures 3, 4). Additionally, because the excipient usually does not contain NH or NH_2 groups, it does not give cross-peaks in the ^{1}H-^{15}N HSQC spectrum. Since the amino acid residues observed in the ^{1}H-^{15}N HSQC experiment are situated in the protein backbone, and are usually involved in hydrogen bonding, their chemical shifts are extremely sensitive to any change in the H-bond pattern and backbone conformation. Thus the ^{1}H-^{15}N HSQC seems to be highly suitable choice for structural fingerprinting. However, a major drawback of this method is that the recording of a ^{1}H-^{15}N HSQC spectrum with adequate signal-to-noise ratio can be excessively time consuming because of the fact that the natural abundance of the NMR-active ^{15}N nucleus is only 0.4%. For example, using a non-isotope-enriched protein sample of 0.5-1mM concentration, on a high-field NMR instrument (700MHz and above) equipped with cryogenically cooled signal-reception coils (so as to enhance the signal-to-noise ratio) a ^{1}H-^{15}N-HSQC spectrum (Figure 3.) can be recorded in approximately 40 to 90 hours, depending on the protein size and the applied 2D pulse sequence. Methodological improvements concerning the fine details of pulse sequences were developed in the last decade in order to significantly reduce the time of data acquisition. As a result of these developments, the so called SOFAST (alternatively referred to as BEST) versions of HSQC and HMQC experiments became available. An

alternative acquisition method known as non-uniform sampling (NUS) resulted in a further reduction in the overall experimental time. In the case of 2D methods, if the acquisition parameters of NUS and its related data processing parameters are set properly, the experiment time can be reduced by half without disadvantageous artifacts on the spectra [17,18]. These rapid acquisition methods proved to be extremely useful for recording NMR fingerprints of therapeutic proteins [12,17,19,20].

The ^{1}H-^{15}N HSQC method was successfully applied to assess the identity of the bioactive conformation of recombinant insulin [21] , interferon alpha [22], and the most extensively investigated therapeutic protein, the human granulocyte macrophage-colony stimulation factor (GCSF) [23].

In practice, well-resolved ^{1}H-^{15}N HSQC spectra may be recorded at natural isotopic abundance from proteins weighing up to about 50kDa. Most of the therapeutic proteins are mAbs (**Figure 2**) whose average molecular size is above this limit (**Table 1**). Arbogast and coworkers [17] described a protocol in which enzymes were used to cleave the monoclonal antibodies (up to 150kDa) into measurable parts. They proved that the cleavage did not modify significantly the HOS of the truncated parts as compared to the intact molecule, and the resulting pieces of proteins are already small enough for HSQC measurement.

The aforementioned examples show that NMR fingerprinting can be an invaluable tool for assessing the structural similarity of biopharmaceuticals. If the NMR fingerprints are sufficiently detailed, due to their super-sensitivity to minor structural changes, an excellent match can be regarded as a strong experimental support for a very high degree of structural similarity between the compared proteins. This sensitivity however can also have drawbacks: slight differences in the solution conditions may cause minor, but from a practical viewpoint irrelevant differences in the 3D structure (such as those shown in Fig. 4 – see below), easily leading to observable spectral mismatches. Moreover, some spectral differences may stem purely from inevitable differences in the instrumental measurement conditions. Even when using the same NMR spectrometer in a comparability study, the NMR spectra will be recorded at different times, entailing slight differences in magnet homogeneity, pulse power, etc. The ensuing spectral differences may be indistinguishable from those originating from structural dissimilarities that are unrelated to the solution conditions. Which spectral difference is meaningful or irrelevant regarding the assessment of structural similarity is therefore a question that requires careful expert judgment.

Aubin and coworkers [23] addressed the following question. When assessing the formulation-dependent variations observed in the location of the amide-region cross-peaks in the ^{1}H-^{15}N HSQC spectra, to what extent can these changes be attributed purely to formulation-related protein conformational effects, and to what extent are they influenced by other solution-related physicochemical and biochemical effects (such as chemical exchange and H-bonding)? In their study the authors used a ^{13}C as well as ^{15}N labelled version of filgrastim, so that a full resonance assignment could be achieved. Having this atomic level information in hand, they could discriminate between conformation-related and non-conformation-related chemical shift changes. The effects of buffer composition were extensively investigated by titration studies, in which pH, ionic strength, and the concentration of excipients were changed gradually. It was concluded that the amide ^{1}H-^{15}N moieties can be too sensitive to pH effects, resulting in relatively large chemical shift differences in the ^{1}H-^{15}N HSQC spectra. Therefore, any observable spectral change of this kind is not necessarily the consequence of a conformational change of the protein backbone. It was suggested that in cases like the above-mentioned example, the far less “condition-sensitive” but still very “conformation-sensitive” ^{1}H-^{13}C chemical shifts of the methyl signals observed in ^{1}H-^{13}C HSQC spectra should be compared (see also below) instead of a

comparison of the amide ^{1}H and ^{15}N resonances. In other cases, when the buffer conditions are the same for the compared samples and when no changes in the resonances are observed, it can be concluded with certainty that there are no differences in the protein conformation. Another important conclusion of this study was that while the CD method could not distinguish between single- or double-residue mutated protein variants, the 2D NMR fingerprint spectra of all of the investigated mutants were highly distinctive in that regard.

Ghasriani and coworkers [19] performed an interlaboratory round robin test on filgrastim samples in which they showed that the deviations in the fingerprint spectra can be partly attributed to instrumental differences.

A possible solution to avoid buffer-related spectral alterations is the use of less pH-sensitive nuclei as reporters. *Arbogast* [17] introduced a special ^{1}H-^{13}C 2D correlation experiment which is less sensitive to the solution conditions than ^{1}H-^{15}N HSQC. In this experiment, the investigated moieties are the sidechain methyl groups of aliphatic amino acids such as isoleucine, leucine, alanine and valine residues which are in a distinct region of the spectra, and are usually well-separated from the signals of the excipients. The chemical shifts of the methyl groups are sensitive enough to local structural changes, but they are not involved in chemical exchange processes which make them less condition-sensitive reporters.

2.2 Approaches for comparative evaluation

The goal of biosimilar development is to achieve the highest possible level of similarity between the reference and the biosimilar product. Using a complex and conscientiously planned analytical approach in order to determine the degree of (dis)similarity and to uncover the roots of the differences is of crucial importance. This requires not only the need to generate meaningful high-quality analytical data that can serve as a scientifically sound basis for the comparison, but also calls for the development of proper means by which the data-comparison itself is made. Comparing complex analytical data and assessing their degree of similarity, possibly by even somehow quantifying that similarity, is a challenging problem, especially when one is looking at this not simply as a formal chemometric problem, but as a task where biologically and structurally relevant differences in the analytical data should be separated from differences that are in those respects irrelevant.

The simplest means of comparing NMR fingerprints is by human visual inspection. This form of evaluation can carry some subjective elements and yields a qualitative statement declaring the similarity or dissimilarity of the two structures. In spite of these two drawbacks, visual comparison has some distinct advantages: a) the expert eye can, under favorable conditions, interpret whether a given difference in the fingerprint pattern (**Figure 4**) is relevant regarding the compared protein structures; b) visual evaluation can easily encompass a large number of spectra; c) it is independent of the resolution of the spectra. Overall, the visual comparison of NMR fingerprints may still be still regarded as a reliable way of assessing structural similarity. [8,9].

In contrast, chemometric tools give a quantitative measure of similarity and provide more objectivity, since the acceptance criteria can be defined prior to the analysis. The applicable mathematical approaches depend on the quality of the spectra. Less crowded, well-resolved spectra allow a peak-to peak comparison of chemical shifts and signal intensities as applied, e.g., in the case of the ^{15}N-^{1}H HSQC spectrum of filgrastim. [23]. However, congested spectra often require the use of other forms of data sampling such as binning or bucketing [11]. The method of choice also depends on the number of spectra that should be compared.

The number of compared spectra can also define the applied method of data evaluation. For pairwise comparisons, linear regression, sequential nearest-neighbor pattern recognition, or picture analysis can give a measure of similarity [11,12,24], while principal component analysis (PCA) [12] can be applied if a large of number of spectra is compared.

2.3 Investigation of post-translational modifications

Post-translational modifications (PTMs) are mostly enzymatically catalyzed covalent changes in the polypeptide chain, which occur after protein biosynthesis within the protein-producing cell. These modifications include, for instance, the formation of disulfide bridges between cysteine residues, glycosylation, and deamidation. Sometimes subsequent chemical modifications, such as the PEGylation of a protein (i.e., the covalent attachment of polyethylene glycol polymer chains to the protein so as to increase its biological activity, efficacy and circulatory time, and to reduce its immunogenicity and antigenicity), are also called PTMs. PTMs play an essential role in the formation and maintenance of the protein tertiary structure, and thus in the efficacy and stability of the drug product. For example, the

glycosylation pattern is a key factor determining the immunogenic properties of the product. Because PTMs must be identified and characterized even when present as minor components in the drug substance, their investigation is typically a mass spectrometry task. However, in certain cases NMR can provide invaluable complementary information on a PTM's structure, especially with regard to configurational or conformational problems that are not accessible by MS. Based on the recent literature (see below), NMR can play an important role in the determination of disulfide bond configurations, the characterization of PEGylation, and glycan pattern analysis.

Disulfide bonds between cysteine sidechains play a fundamental role in the formation of the tertiary structure of proteins, therefore their determination is of utmost importance. Although NMR applications for disulfide bond characterization were published in the field of biomolecular NMR [27], for biopharmaceutical proteins these methods are not feasible for a number of reasons. First, distance measurements between the β ^{1}H nuclei of the covalently connected cysteine pairs by NOESY methods can be extremely difficult even for isotope-enriched and well-characterized proteins, because the relevant (conformation-dependent) ^{1}H-^{1}H distance is usually quite large and spin diffusion involving the pertinent cystein CH_2 moieties can significantly dilute the measured inter-cystein NOEs. In principle, one known way to get round this problem is to substitute the non-NMR-active sulfur atoms with NMR-active selenium atoms, which would allow the measurement of the direct inter-selenocystein Se-Se through-bond connectivities. However, besides this being a biotechnologically rather difficult task, it is, similarly to isotopic enrichment, not an acceptable procedure in the framework of the biopharmaceutical industry. Fortunately, because 1D and 2D NMR fingerprints inherently encode all information about the HOS, including the disulfide pattern, a high degree of NMR fingerprint-level similarity is a very strong indication that the disulfide linkages are identical in the compared proteins.

For PEGylated proteins, the identification of the PEGylation site is a critical issue in biosimilar characterization [1]. This task is typically performed by LC-MS methods, with which enzymatic protein digests are investigated at the residue level. Although this approach is usually successfully applicable, NMR can also be useful as a complementary method. Wang and coworkers [28] demonstrated on PEGylated interferon alpha-2b that by the combination of 1D and 2D NMR measurements the PEGylation site can be identified unambiguously for enzymatically digested and HPLC-separated PEGylated peptide samples [28]. The key step of the analysis was based on the fact that the enzymatic digestion leaves the PEG chain unaffected, which can be later separated by molecular weight.

According to the FDA guidance [1], in the case of glycoproteins such as mAbs the detailed characterization of the glycosylation pattern and the demonstration of its similarity to that of the reference is critical. This is a rather challenging task because, on the one hand, glycans themselves have very complex structures. On the other hand, mAbs exhibit much heterogeneity in terms of the attached glycans, and some glycans are often present in such minor quantities that they can only be investigated by MS. When the desired level of structural identification is the constitution of the individual sugar units, MS is therefore the method of choice. However, when one needs to distinguish between isobaric sugar units, and/or more detailed structural information is required, NMR comes into play. For example, NMR is one of the few tools capable of providing direct information about the anomeric configuration of glycosidic linkages between the sugar units. A nice example of this application was given in a study [29] in which LC-MS/MS and NMR methods were integrated for the characterization of N-glycans of cetuximab. This paper showed that although NMR is less sensitive than MS, the proper combination of these techniques allows the unambiguous identification of isobaric glycan structures even at a concentration range of approximately 15 pmol. The authors suggested the use of their method on a routine basis

when, as in the case of cetuximab, immunogenic α-1,3-Gal and the N-glycolylneuraminic acid epitopes cannot be distinguished from their isobaric forms by the commonly applied methods.

3. Quality control of raw materials and impurities

In the development and production of biologics, the role of NMR extends way beyond the structural characterization of the protein drug substance. In fact, as discussed below, NMR can be usefully employed to address a number of other analytical problems that are an inherent part of the whole biotechnology process, such as the characterization of raw materials and impurities, and the quantification of the drug substance in the drug product. As such, NMR can be envisaged as playing a certain role in some process analytical technologies (PAT). Below we give a few examples of such less well-known uses of NMR in biotechnology.

3.1 Extractables and leachables, and other small-molecule process-related impurities

Regulatory authorities emphasize the requirements for the purity of the biological drug substances and drug products and for the identification of their product- and process-related impurities [30]. Process-related impurities can originate from the manufacturing process, cell substrates (host cell proteins or DNA), fermentation (medium components, antibiotics), or downstream processing. The most common sources of small-molecule impurities are technology-related and/or extractable and leachable (L&E) compounds originating from the accessories of fermentation and raw material storage. Some common small-molecule process-related impurities and L&E compounds are isopropyl alcohol, acetic acid, ethyl acetate, acetone, silicone polymers, glycerol, plasticizers and antioxidants like sterically hindered phenolic antioxidants, etc. These impurities should be characterized analytically to the extent possible, and where feasible, their biological activities should be evaluated. The identification of such impurities besides the protein drug substance can be a rather difficult problem, often requiring a complex analytical approach using a variety of chromatographic and spectroscopic techniques such as thin layer chromatography (TLC), HPLC, HPLC-MS, GC, etc. Within the analytical repertoire employed for that purpose, NMR can find a significant and unique place due to its potential to detect and identify small-molecule contaminations in complex solutions without the physical separation of the components. For example, Skidmore et al. [16] introduced for that purpose an NMR-based approach which exploits the differences between the relaxational properties of large proteins and small molecules. Based on this difference, they used the so-called Carr-Purcell-Meiboom-Gill (CPMG) pulse sequence to decrease the intensities of the product without removing the protein from the solution (Figure 5).

3.2 Protein quantification

The determination of the absolute concentration of a protein drug substance in a drug product can be an analytically problematic task. This is because in some cases the commonly applied methods (such as UV-absorbance and colorimetric measurements) give contradictory results for a variety of reasons, and so alternative techniques should be evoked. [31] Under properly chosen experimental conditions, in principle the area under any given signal of a 1D ^{1}H NMR spectrum is directly proportional to the number of ^{1}H nuclei contributing to that signal. This seems to make the method uniquely reliable and convenient for the determination of absolute concentrations by using a suitable reference compound whose concentration is known. However, in practice the relation between the intensity of the NMR signals and the

protein concentration is not that straightforward, and therefore the data acquisition and evaluation of the spectra need great care. The use of an internal reference compound is generally not recommended for NMR-based protein quantification, because one cannot exclude the possibility that the protein will interact with the reference molecule, potentially distorting the results. If separate and assignable resonances of the protein are observable in the 1D ^{1}H NMR spectrum, or at least the number of ^{1}H nuclei belonging to a given signal or specific spectral region can be determined, an external reference material can be used for protein quantification by taking into account the conditions of the solution and the parameters of the NMR measurement. Wider and coworkers discussed in detail [32] the critical parameters that should be considered in that regard. After adjusting identically as many physicochemical and NMR measurement parameters as possible in the reference and the analyte samples, one must be careful to take into account the possibility that an excitation pulse applied for exactly the same length of time may affect the spins in the two samples to a slightly different degree, yielding slight unwanted differences in signal strength. The authors introduced a formula and an ideal workflow (called PULCON) involving pulse calibration and other experimental parameters (such as the temperature and additional correction factors) so as to compensate for this effect [33]. Later, the PULCON method was improved further with non-^{1}H-nucleus filtering [34] in order to make the method applicable for heavily overlapping cases, but this development requires isotopic labelling, consequently it has less relevance in the biopharmaceutical industry.

Because a protein's hydration shell is an inherent constituent of its 3D structure, protein NMR studies are practically always conducted in aqueous solutions, whereby a large signal due to water appears in the ^{1}H NMR spectra, often covering important spectral regions and interfering severely with data evaluation. To alleviate this problem, several special NMR techniques that are commonly referred to as "water suppression" should be applied. In order to better appreciate the necessity of using these methods, it is worth pointing out that the concentrations of the water and protein ^{1}H nuclei in the sample are on the order of 100M and <1mM, respectively; note that this is a difference of more than 6 orders of magnitude! This means that, without suppression, the water signal can completely dominate the spectrum and conceal the protein resonances. The fact that the protein signals can be made visible, allowing the collection of atomic-level structural information, shows the technical ingenuity of these water-suppression techniques. However, water suppression methods can introduce various inherent undesired artifacts in the spectrum, whereby their application represents one of the main technical difficulties in these types of NMR experiment. Performing the measurements in pure deuterated buffer would seem to be a trivial option to solve this problem, but in everyday practice it is truly difficult to implement this option. Moreover, buffer exchange steps deprive NMR from one of its main advantages over other techniques, namely its potential to examine the intact sample as it is. Wider and his team [32] pointed out that those water suppression techniques which saturate the protein signals, such as "presaturation", must be avoided, since they have a detrimental effect on protein signal intensities as well [32]. Instead, those NMR experiments should be preferred which incorporate the so-called WATERGATE (water suppression by gradient tailored excitation) [35] building block, and the signal loss during the water suppression module of the pulse sequence has to be determined and considered as a correction factor when the concentration is calculated by the given formula. In cases where well-resolved peaks are not present in the ^{1}H spectrum, estimation of the number of protons ^{1}H nuclei and integration within specific spectral regions can be performed. This approach however must be treated with reservations and should be applied only for the comparison of samples rather than for the measurement of absolute protein concentration.

In order to reliably carry out the ^{1}H NMR integration and to be able to calculate the protein concentration from the integral value, one must find a signal (or a cluster of signals) that is a) purely of protein origin; b) sufficiently well-resolved from other resonances; c) for which we know exactly the number of contributing protein ^{1}H nuclei. For large proteins such as therapeutic mAbs that have a natural isotopic abundance and are accompanied by small molecules (components of the formulation buffer), it is in practice often almost impossible to find such protein signals, which can render the task of quantification rather arduous. This problem can be overcome by using an appropriate quantitative heteronuclear 2D method in which, for example, the areas under the separate methyl cross-peaks are integrated. As it was demonstrated recently [36,37], this approach carries a number of potential experimental biases when the common 2D NMR measurement methods are applied, even for small molecules. The reliability and accuracy of 2D correlation spectroscopy for concentration measurement depends on multiple factors, including the relaxation properties of the investigated molecule, the coupling constants between the correlated nuclei, and even the finer details of the pulse sequence that generates the 2D spectrum. All this makes the approach quite problematic and requires the validation of the method for each individual case. Hopefully, validated and universally applicable quantitative 2D NMR methods will be developed in the near future.

3.3 Analysis of cell culture medium composition

During the development of the upstream processes in which biomolecules are manufactured through fermentation, having the analytical capability to monitor the composition of the spent culture medium can be of much value. [38,39] When multiple small-molecule nutrients of the culture medium are being identified and quantified simultaneously, the procedure becomes methodically identical with the well-known field of metabonomics, which essentially deals with the quantitative analytical measurement of the dynamic metabolic profile of living systems and the statistical evaluation of the data. When these techniques were applied to the analysis of the spent fermentation broth of mammalian cell cultures, the literature sometimes referred to these studies as "fermentanomics" [39–41]. Traditionally, metabonomic studies are performed by LC-MS, NMR or a combination of the two techniques, depending on the sample number, and the desired information. Nowadays, in an industrial environment high-throughput techniques are widely used instead of adjusting the parameters one by one from one fermentation procedure to another in order to facilitate the optimization of the parameters of protein production. While large-scale bioreactors are often equipped with online detectors to follow the concentration-changes of specific nutrients and metabolites, the current solutions for the online monitoring of small-volume high-throughput-screening fermentation devices are limited in terms of the number of observable parameters. At this point, well-established NMR-based metabonomics tools come into play, since they have several advantages over LC-MS [42,43]. In particular, these NMR applications require minimal sample manipulation, they have high reproducibility and allow high-throughput data acquisition, and the accompanying statistical data analysis techniques are well consolidated.

In a significant study, Bradley et al. applied NMR fermentanomics for the optimization of the feed medium of mammalian cell cultures. [39]. In that work the authors used a series of so-called first-increment 1D ^{1}H-NOESY experiments with water suppression (these are essentially 1D ^{1}H NMR spectra by appearance that are more suitable for quantification than a conventional 1D ^{1}H NMR spectrum) for the time-course nutrient-concentration profiling of fermentations. After a 4-fold dilution of the centrifuged medium supernatant by D_2O-based phosphate buffer, a 20 minute per sample total experimental time was achieved, enabling high-throughput data recording and resulting in an appropriate signal-

to-noise ratio, proper pH control, and excellent water suppression. The nutrient resonances were assigned by the concerted use of public chemical shift databases and spiking with known purchased compounds, and data evaluation was aided by software designed for the spectral analysis of liquid mixtures. The authors demonstrated that the time-consuming and labor-intensive process of culture medium optimization can be very effectively facilitated by the fermentanomics approach: a correlation was observed between the depletion of certain NMR-observable nutrients and mAb productivity, and in that particular case histidine was identified as one of the productivity-limiting nutrients. This study also gave a surprising result: it revealed the accumulation of such volatile acetate and formate metabolites in the samples which had not been previously reported in mammalian cell cultures when investigated by LC-MS, confirming that NMR is the primary method for fermentanomics. The authors suggested a protocol for the monitoring of intracellular metabolites as well by using a combination of cell lysis and washing steps, followed by NMR measurements. In the case of these protein-rich samples CPMG-based NMR methods were preferred to filter out the broad lines of macromolecules.

In a later study [40] a similar NMR approach was used, and further culture medium optimization was achieved, in the case of an immunoglobulin G (IgG) antibody producing cell culture, preventing the depletion of several nutrients from the fermentation which resulted in a higher yield of the desired antibody without any undesirable impact on the protein structure and *N*-linked glycan profile. A further improvement on the fermentanomic approach was subsequently published [41] in which multivariate data analysis (MVDA) was introduced for the statistical evaluation of the NMR spectra of the spent culture media. By this alternative approach product quality attributes (such as protein glycosylation) could be correlated with process variables. It was demonstrated that fermentanomics, together with MVDA, is efficiently applicable not only for the rapid finding of the optimal composition of the culture media, but also for process optimization and for controlling the consistency of batch-to-batch variations of the product quality attribute profile.

The need for the compositional analysis of the raw materials used in mammalian cell fermentations also crops up from the point of view of the internal quality control of the purchased culture media and feeds. Such quality control is essential to ensure the sustainable batch-to-batch quality of the produced protein drug substances. The characterization of these complex mixtures, which contain essential nutrients and vitamins in a well-defined ratio, is challenging because vendors usually regard their compositions as proprietary knowledge. 1D ^{1}H NMR spectra provide, as described above, fingerprint-like quantitative and qualitative information (see Fig. 6) on almost all of the organic components and can demonstrate the batch-to-batch identity of the purchased media in a robust, quick and convenient way, requiring much simpler sample preparation than with other methods such as HPLC-MS.

4. Conclusions

Based partly on our own experiences and partly on the recent literature, in this review we collected the main areas in which modern NMR spectroscopy can usefully contribute to the development and the quality control of biotechnological drug substances in a pharmaceutical industrial environment, mainly focusing in that regard on protein biosimilars. We discussed the unique capability of NMR to provide highly detailed fingerprint-like information about a protein's primary and high-order structure at an atomic resolution. Several 1D and 2D NMR experiments can be used to generate various types of such fingerprints, and these offer powerful tools for comparing the structural similarity of the follow-on and the reference proteins. In addition, we attempted to draw attention to the utility

of NMR in related analytical tasks such as the identification of post-translational modifications and small-molecule impurities, the quantification of the protein drug substance in the drug product, and the analysis of cell culture media. The presented methods demonstrate the great benefits that NMR can provide, as well as the limitations inherent to this technique. Most of the papers that we found to be relevant to this topic had been published in the last three to five years, indicating that this field is young and is rapidly evolving. With the expected future advancements in the sensitivity and resolution of NMR, its role should be further expanding conspicuously in the field of biopharmaceuticals. This optimistic future prospect however will not only depend on the increasing technical capabilities of NMR, but also on whether the general appreciation of what NMR can do in the biopharmaceutical industry will also increase in the minds of the non-NMR experts. Biotechnology R&D programs are multidisciplinary projects whose success depends on the close collaboration of its expert participants and a mutual understanding of the scopes and limitations of each analytical technique being used. Thus, such general understanding of the role of NMR, at least at a conceptual level, is critical, and we modestly hope that this review may have contributed to that goal.

References

[1] CDER/CBER, FDA, Scientific Considerations in Demonstrating Biosimilarity to a Reference Product Guidance for Industry Scientific Considerations in Demonstrating Biosimilarity to a Reference Product, Guid. Ind. (2015).

[2] D.S. Wishart, Characterization of biopharmaceuticals by NMR spectroscopy, TrAC - Trends Anal. Chem. 48 (2013) 96–111. doi:10.1016/j.trac.2013.03.009.

[3] European Medicines Agency, Guideline on Similar Biological Medicinal Products Containing Biotechnology-Derived Proteins As Active Substance: Quality Issues, London, 2006. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003953.pdf.

[4] S. Berkowitz, J. Engen, J. Mazzeo, G. Jones, Analytical tools for characterizing biopharmaceuticals and the implications for biosimilars, Nat. Rev. Drug Discov. 11 (2013) 527–540. doi:10.1038/nrd3746.Analytical.

[5] H. Wei, J. Mo, L. Tao, R.J. Russell, A.A. Tymiak, G. Chen, R.E. Iacob, J.R. Engen, Hydrogen/deuterium exchange mass spectrometry for probing higher order structure of protein therapeutics: Methodology and applications, Drug Discov. Today. 19 (2014) 95–102. doi:10.1016/j.drudis.2013.07.019.

[6] J. Fang, Y.Q. Yu, A. Beck, W. Chen, A. Chakraborty, HDX-MS for Assessing Comparability Between Innovator and Biosimilar Biotherapeutics, Poster. (2016) 2016.

[7] J. Pan, S. Zhang, C.H. Borchers, Comparative higher-order structure analysis of antibody biosimilars using combined bottom-up and top-down hydrogen-deuterium exchange mass spectrometry, Biochim. Biophys. Acta - Proteins Proteomics. 1864 (2016) 1801–1808. doi:10.1016/j.bbapap.2016.08.013.

[8] F. Sorgel, A. Schwebig, J. Holzmann, S. Prasch, P. Singh, M. Kinzig, Comparability of biosimilar filgrastim with originator filgrastim: Protein characterization, pharmacodynamics, and pharmacokinetics, BioDrugs. 29 (2015) 123–131. doi:10.1007/s40259-015-0124-7.

[9] J. Visser, I. Feuerstein, T. Stangler, T. Schmiederer, C. Fritsch, M. Schiestl, Physicochemical and functional comparability between the proposed biosimilar rituximab GP2013 and originator rituximab, BioDrugs. 27 (2013) 495–507. doi:10.1007/s40259-013-0036-3.

[10] D.I. Freedberg, Using nuclear magnetic resonance spectroscopy to characterize biologicals., Dev. Biol.

(Basel). 122 (2005) 77–83.

[11] C.A. Amezcua, C.M. Szabo, Assessment of Higher Order Structure Comparability in Therapeutic Proteins Using Nuclear Magnetic Resonance Spectroscopy, J Pharm Sci. 102 (2013) 1724–1733. doi:10.1002/jps.23531.

[12] B. Japelj, G. Ilc, J. Marušič, J. Senčar, D. Kuzman, J. Plavec, Biosimilar structural comparability assessment by NMR: from small proteins to monoclonal antibodies, Sci. Rep. 6 (2016) 32201. doi:10.1038/srep32201.

[13] L. Poppe, J.B. Jordan, K. Lawson, M. Jerums, I. Apostol, P.D. Schnier, Profiling formulated monoclonal antibodies by 1H NMR spectroscopy, Anal. Chem. 85 (2013) 9623–9629. doi:10.1021/ac401867f.

[14] X.U.M. Fang, J.D.F. Yan, Higher Order Structure Assessment of Biosimilars Based on the Correlation of NMR Spectral Fingerprints, Chinese J. Magn. Reson. 32 (2015) 342–353. doi:10.11938/cjmr20150216.

[15] J. Franks, J.N. Glushka, M.T. Jones, D.H. Live, Q. Zou, J.H. Prestegard, Spin Diffusion Editing for Structural Fingerprints of Therapeutic Antibodies, Anal. Chem. 88 (2016) 1320–1327. doi:10.1021/acs.analchem.5b03777.

[16] K. Skidmore, D. Hewitt, Y.H. Kao, Quantitation and characterization of process impurities and extractables in protein-containing solutions using proton NMR as a general tool, Biotechnol. Prog. 28 (2012) 1526–1533. doi:10.1002/btpr.1620.

[17] L.W. Arbogast, R.G. Brinson, J.P. Marino, Mapping Monoclonal Antibody Structure by 2D 13C NMR at Natural Abundance, Anal. Chem. 87 (2015) 3556–3561. doi:10.1021/ac504804m.

[18] M.R. Palmer, B.R. Wenrich, P. Stahlfeld, D. Rovnyak, Performance tuning non-uniform sampling for sensitivity enhancement of signal-limited biological NMR, in: J. Biomol. NMR, 2014: pp. 303–314. doi:10.1007/s10858-014-9823-5.

[19] H. Ghasriani, D.J. Hodgson, R.G. Brinson, I. McEwen, L.F. Buhse, S. Kozlowski, J.P. Marino, Y. Aubin, D.A. Keire, Precision and robustness of 2D-NMR for structure assessment of filgrastim biosimilars, Nat. Biotechnol. 34 (2016) 139–141. doi:10.1038/nbt.3474.

[20] D.J. Hodgson, Y. Aubin, Assessment of the structure of pegylated-recombinant protein therapeutics by the NMR fingerprint assay, J. Pharm. Biomed. Anal. 138 (2017) 351–356. doi:10.1016/j.jpba.2017.01.058.

[21] X. Jin, S. Kang, H. Kwon, S. Park, Heteronuclear NMR as a 4-in-1 analytical platform for detecting modification-specific signatures of therapeutic insulin formulations, Anal. Chem. 86 (2014) 2050–2056. doi:10.1021/ac403218f.

[22] N. Panjwani, D.J. Hodgson, S. Sauvé, Y. Aubin, Assessment of the effects of pH, formulation and deformulation on the conformation of interferon alpha-2 by NMR, J. Pharm. Sci. 99 (2010) 3334–3342. doi:10.1002/jps.22105.

[23] Y. Aubin, D.J. Hodgson, W.B. Thach, G. Gingras, S. Sauvé, Monitoring Effects of Excipients, Formulation Parameters and Mutations on the High Order Structure of Filgrastim by NMR, Pharm. Res. 32 (2015) 3365–3375. doi:10.1007/s11095-015-1713-3.

[24] S̆. Župerl, P. Pristovšek, V. Menart, V. Gaberc-Porekar, M. Novič, Chemometric approach in quantification of structural identity/similarity of proteins in biopharmaceuticals, J. Chem. Inf. Model. 47 (2007) 737–743. doi:10.1021/ci6005273.

[25] K. Chen, D.S. Long, S.C. Lute, M.J. Levy, K.A. Brorson, D.A. Keire, Simple NMR methods for evaluating higher order structures of monoclonal antibody therapeutics with quinary structure, J. Pharm. Biomed. Anal. 128 (2016) 398–407. doi:10.1016/j.jpba.2016.06.007.

[26] L.W. Arbogast, R.G. Brinson, T. Formolo, J.T. Hoopes, J.P. Marino, 2D 1HN, 15N Correlated NMR Methods at Natural Abundance for Obtaining Structural Maps and Statistical Comparability of Monoclonal Antibodies, Pharm. Res. 33 (2016) 462–475. doi:10.1007/s11095-015-1802-3.

[27] M. Mobli, G.F. King, NMR methods for determining disulfide-bond connectivities, Toxicon. 56 (2010) 849–854. doi:10.1016/j.toxicon.2010.06.018.

[28] Y.S. Wang, S. Youngster, J. Bausch, R. Zhang, C. McNemar, D.F. Wyss, Identification of the major positional isomer of pegylated interferon alpha-2b., Biochemistry. 39 (2000) 10634–10640.

[29] A. Wiegandt, B. Meyer, Unambiguous Characterization of N Glycans of Monoclonal Antibody Cetuximab by Integration of LC-MS/MS and 1H-NMR Spectroscopy, Anal. Chem. 86 (2014) 4807−4814. doi:dx.doi.org/10.1021/ac404043g.

[30] C. ICH, Q 6 B Specifications: Test Procedures and Acceptance Criteria for Biotechnological/Biological Products, Eur. Med. Agency. (1999) 1–17. papers2://publication/uuid/3FAC0933-7C71-4C87-BC90-81C0D1DE0716.

[31] A.C. Conibear, N.L. Daly, D.J. Craik, Quantification of small cyclic disulfide-rich peptides., Biopolymers. 98 (2012) 518–24. doi:10.1002/bip.22121.

[32] G. Wider, L. Dreier, Measuring protein concentrations by NMR spectroscopy, J. Am. Chem. Soc. 128 (2006) 2571–2576. doi:10.1021/ja055336t.

[33] D.I. Hoult, R.E. Richards, The signal-to-noise ratio of the nuclear magnetic resonance experiment, J. Magn. Reson. 24 (1976) 71–85. doi:10.1016/0022-2364(76)90233-X.

[34] L. Dreier, G. Wider, Concentration measurements by PULCON using X-filtered or 2D NMR spectra, Magn. Reson. Chem. 44 (2006). doi:10.1002/mrc.1838.

[35] M. Piotto, V. Saudek, V. Sklenar, Gradient-Tailored Excitation for Single-Quantum Nmr-Spectroscopy of Aqueous-Solutions, J. Biomol. Nmr. 2 (1992) 661–665.

[36] F. Fardus-Reid, J. Warren, A. Le Gresley, Validating heteronuclear 2D quantitative NMR, Anal. Methods. 8 (2016) 2013–2019. doi:10.1039/C6AY00111D.

[37] P. Giraudeau, Quantitative 2D liquid-state NMR, Magn. Reson. Chem. 52 (2014) 259–272. doi:10.1002/mrc.4068.

[38] P. Gronemeyer, R. Ditz, J. Strube, Trends in Upstream and Downstream Process Development for Antibody Manufacturing, (2014) 188–212. doi:10.3390/bioengineering1040188.

[39] S.A. Bradley, A. Ouyang, J. Purdie, T.A. Smitka, T. Wang, A. Kaerner, Fermentanomics: Monitoring mammalian cell cultures with NMR spectroscopy, J. Am. Chem. Soc. 132 (2010) 9531–9533. doi:10.1021/ja101962c.

[40] E.K. Read, S.A. Bradley, T.A. Smitka, C.D. Agarabi, S.C. Lute, K.A. Brorson, Fermentanomics informed amino acid supplementation of an antibody producing mammalian cell culture, Biotechnol. Prog. 29 (2013) 745–753. doi:10.1002/btpr.1728.

[41] A.S. Rathore, S. Kumar Singh, M. Pathak, E.K. Read, K.A. Brorson, C.D. Agarabi, M. Khan, Fermentanomics: Relating quality attributes of a monoclonal antibody to cell culture process variables and raw materials using multivariate data analysis, Biotechnol. Prog. 31 (2015) 1586–1599. doi:10.1002/btpr.2155.

[42] M.R. Viant, E.S. Rosenblum, R.S. Tieerdema, NMR-based metabolomics: a powerful approach for

characterizing the effects of environmental stressors on organism health., Environ. Sci. Technol. 37 (2003) 4982–9. http://www.ncbi.nlm.nih.gov/pubmed/14620827.

[43] Abdul-Hamid M. Emwas, The Strengths and Weaknesses of NMR Spectroscopy and Mass Spectrometry with Particular Focus on Metabolomics Research, in: J.B. Bjerrum (Ed.), Metabonomics Methods Protoc., 2015: pp. 161–193. doi:10.1007/978-1-4939-2377-9_13.

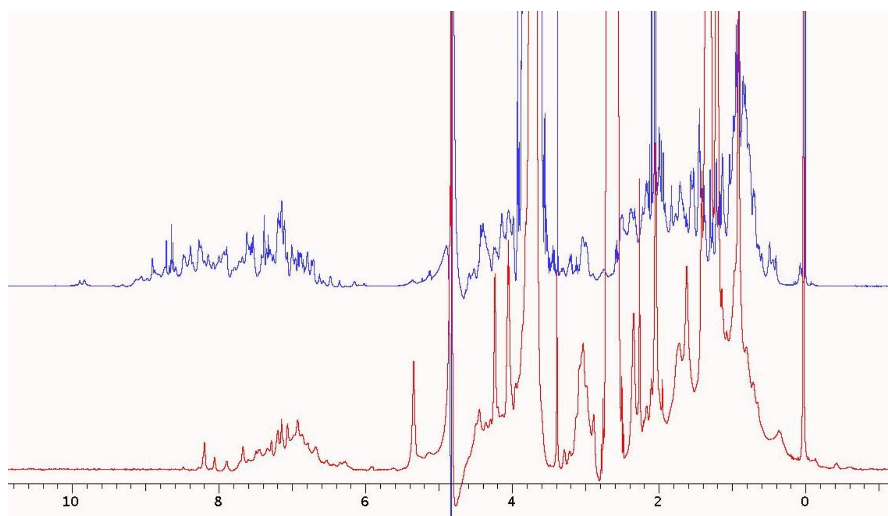

Figure 1 1D ^{1}H NMR spectrum of a middle-sized therapeutic protein (17kDa) (upper trace) and that of a 150 kDa monoclonal antibody drug product, which contains approximately seven times more ^{1}H nuclei (lower trace). Because of the increased spectral congestion and the line-broadening effects associated with the increased molecular weight, in the latter case far fewer resolved signals can be detected in the spectrum. The intensive signals in the aliphatic region correspond to the excipients.

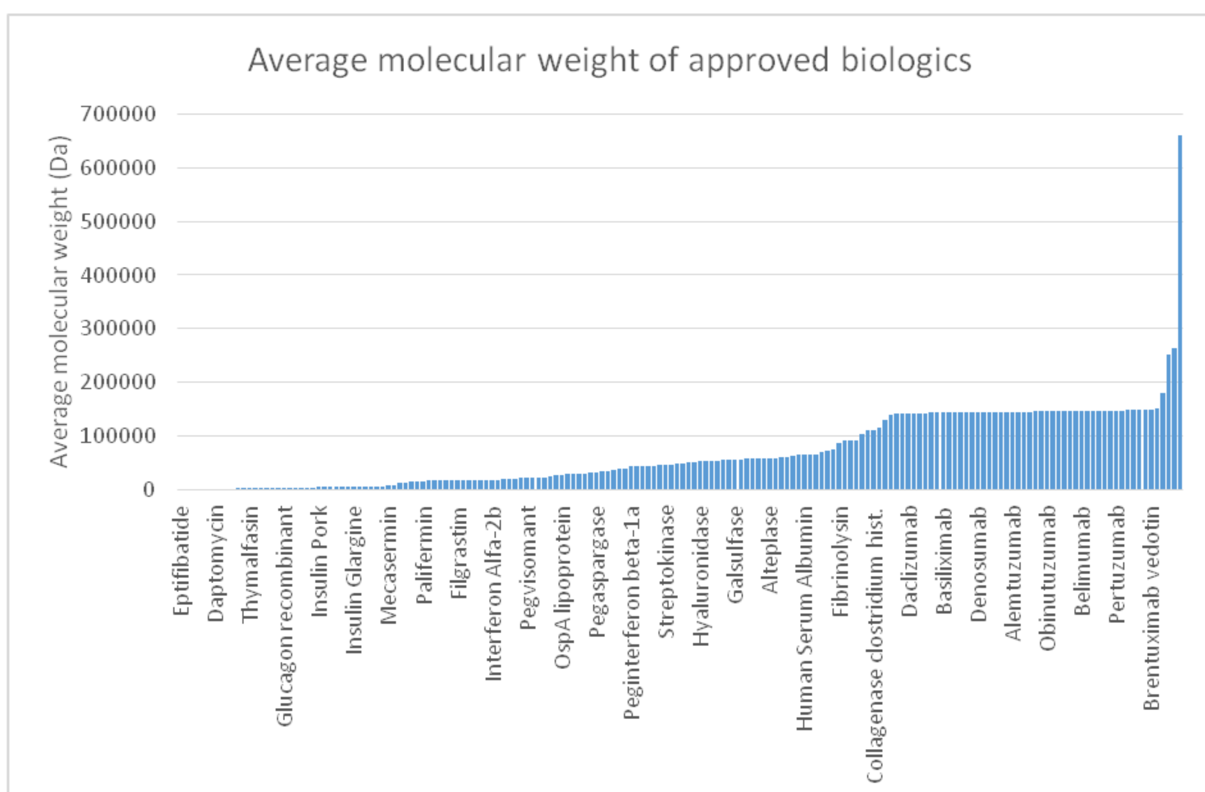

Figure 2 Average molecular weight of some randomly selected approved therapeutic proteins in the US and EU. The average molecular weight influences the appropriate NMR approach to investigate the structural properties.

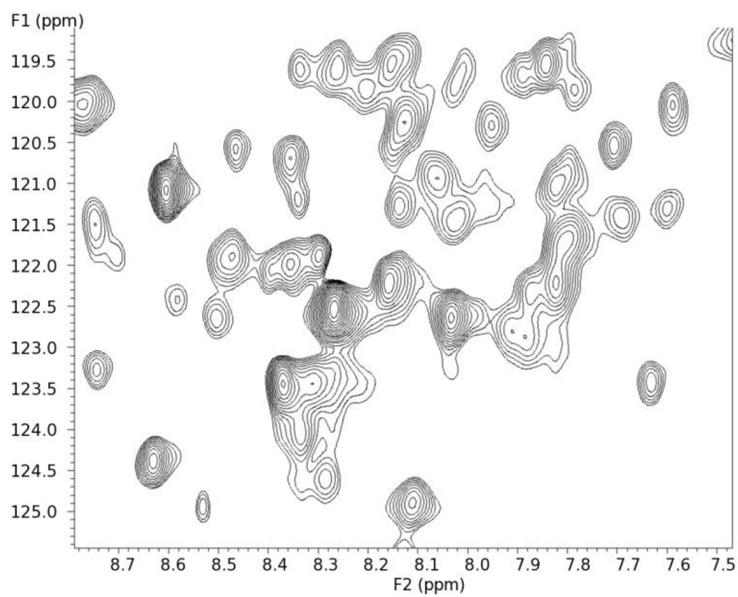

Figure 3 Part of a ^{1}H-^{15}N HSQC spectrum recorded in-house at natural isotopic abundance from of a ~ 20kDa protein at 800MHz ^{1}H frequency.

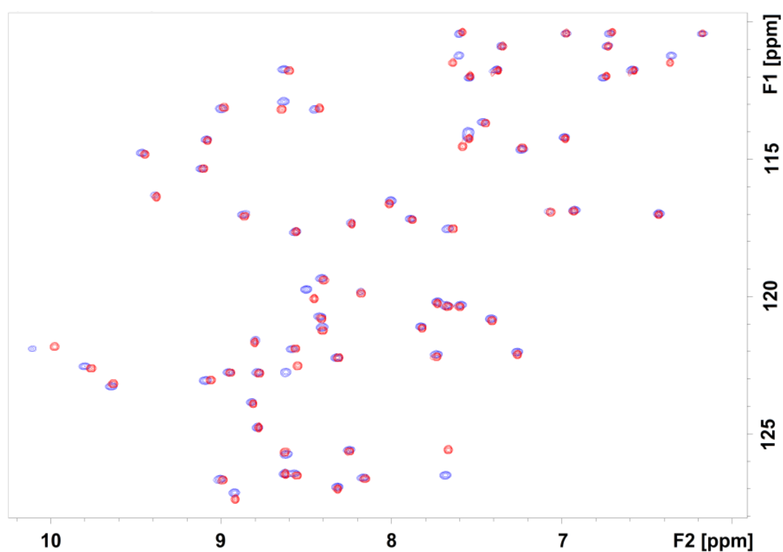

Figure 4 Overlaid view of two ^{1}H-^{15}N HSQC spectra of a small protein (~ 6 kDa). Clearly, the spectra are very similar but not completely identical. While the number of peaks and the majority of peak positions match, slight chemical shift differences can be observed in the case of some signal pairs because of a slight difference in the pH of the two samples. Being aware of this difference in the solution conditions, a human expert can make the judgment that the two corresponding protein structures are similar, while a chemometric approach would indicate dissimilarity in this specific case.

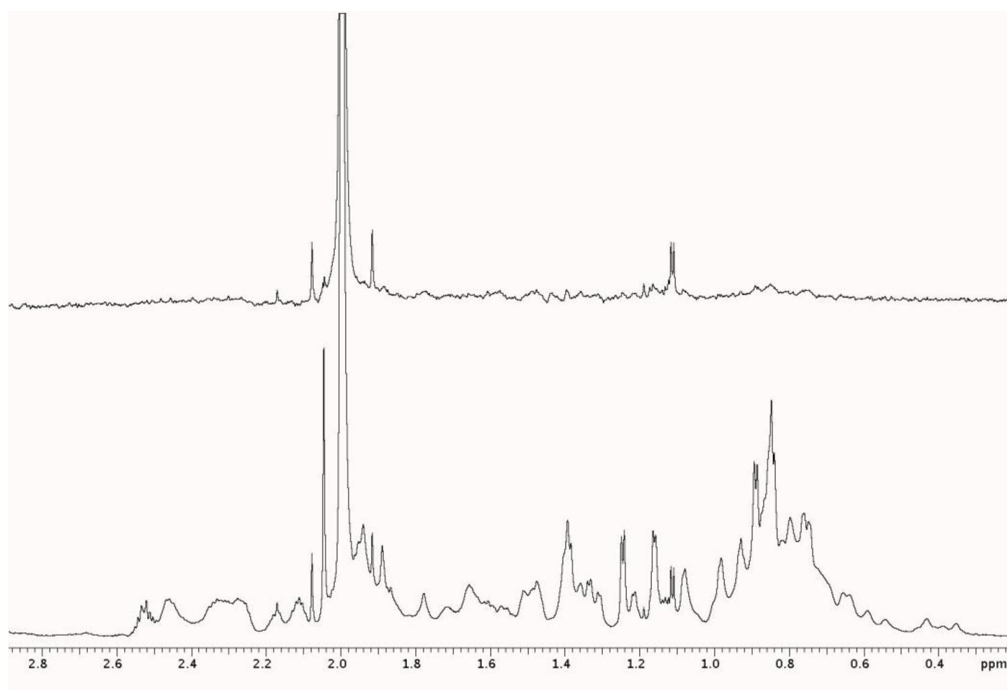

Figure 5. In NMR spectroscopy, the signals coming from the protein and small-molecule impurities can be distinguished without the physical separation of the components. The lower trace shows the normal 1D ^{1}H NMR spectrum (800MHz, in-house measurement) of a biological drug product. By the application of a CPMG-based signal-suppression scheme, the signals of the protein can be eliminated from the spectrum (upper trace). The doublet at 1.17 ppm corresponds to a small-molecule impurity, while the intensive signal at 2 ppm comes from a buffer component.

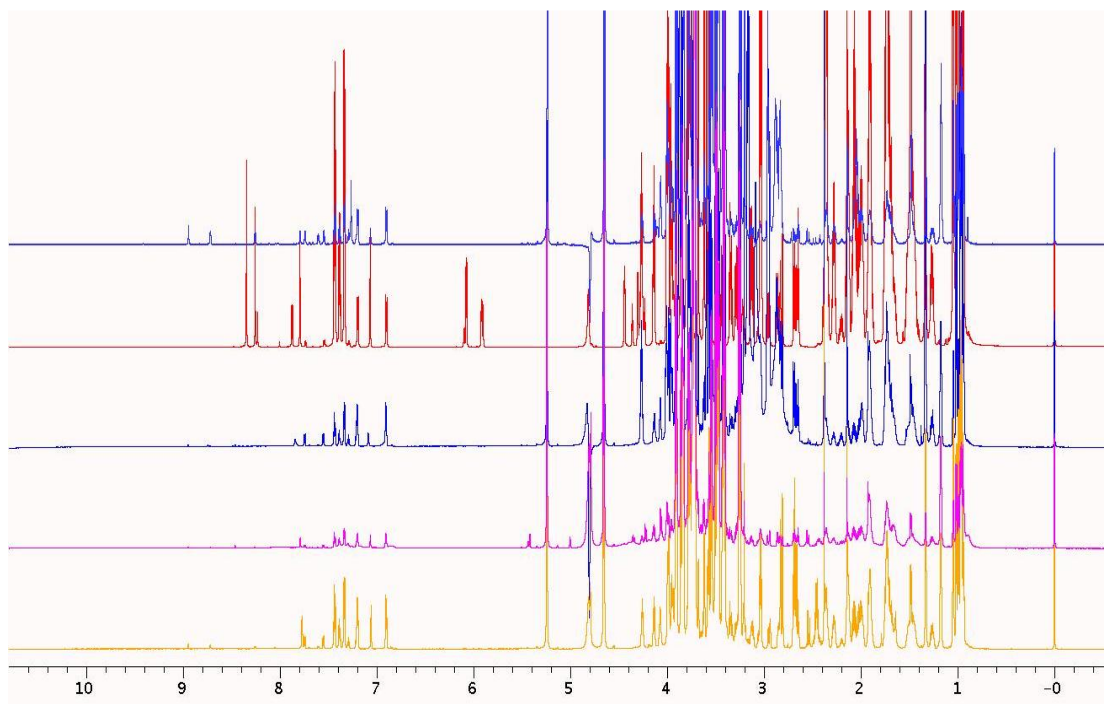

Figure 6. ^{1}H NMR spectra (800MHz, in-house measurement obtained with water suppression) of different mammalian cell culture media. From a series of this simple experiment, useful qualitative and quantitative information can be collected from the mixture, which facilitates medium development and can potentially serve as an in-process quality control tool.

Table 1. Investigated biological products referenced in the literature, and the applied NMR approaches.

INVESTIGATED PROTEIN	AVERAGE MOLECULAR WEIGHT	APPLIED NMR TECHNIQUES	REFERENCE
Cubicin (Daphtomycin)	1.6 kDa	^{15}N HSQC	[14]
Insulin	7.6 kDa	^{15}N HSQC	[21]
Interferon-α2	19.2 kDa	^{15}N HSQC	[22]
MET-G-CSF	18.8 Da	^{1}H 2D NOESY ^{15}N HSQC	[8] [24] [19]
PEG Interferon-α2B	>19.2 kDa	^{15}N HSQC	[20]
PEG-GCSF	~ 38.8 kDa	^{15}N HSQC	[20]
Rituximab	143.9 Da	1D-^{1}H 1D-^{1}H NOESY	[9] [25]
Infliximab	144.2 kDa	1D-^{1}H	[25]
Trastuzumab	145.5 kDa	1D-^{1}H	[14]
mAb (unidentified)	~ 150 kDa	1D-^{1}H enhanced by pulsed field gradient stimulated echo	[13]
mAb (unidentified)	~ 150 kDa	enzymatic digestion, ^{1}H, ^{15}N SOFAST-HMQC	[26]
mAb (unidentified)	~ 150 kDa	enzymatic digestion, ^{1}H, ^{15}N SOFAST-HMQC	[26]