

Review

An Evaluation of the Potential of NMR Spectroscopy and Computational Modelling Methods to Inform Biopharmaceutical Formulations

Akash Pandya ¹, Mark J. Howard ², Mire Zloh ^{3,4} and Paul A. Dalby ^{1,*} 

¹ Department of Biochemical Engineering, University College London, Gordon Street, London WC1E 7JE, UK; akash.pandya.15@ucl.ac.uk

² School of Chemistry, University of Leeds, Leeds LS2 9JT, UK; M.J.Howard@leeds.ac.uk

³ Faculty of Pharmacy, University Business Academy, Trg mladenaca 5, 21000 Novi Sad, Serbia; zloh@live.co.uk

⁴ NanoPuzzle Medicines Design, Business & Technology Centre, Bessemer Drive, Stevenage SG1 2DX, UK

* Correspondence: p.dalby@ucl.ac.uk; Tel.: +44-207-6769-9566

Received: 11 July 2018; Accepted: 17 September 2018; Published: 21 September 2018



Abstract: Protein-based therapeutics are considered to be one of the most important classes of pharmaceuticals on the market. The growing need to prolong stability of high protein concentrations in liquid form has proven to be challenging. Therefore, significant effort is being made to design formulations which can enable the storage of these highly concentrated protein therapies for up to 2 years. Currently, the excipient selection approach involves empirical high-throughput screening, but does not reveal details on aggregation mechanisms or the molecular-level effects of the formulations under storage conditions. Computational modelling approaches have the potential to elucidate such mechanisms, and rapidly screen *in silico* prior to experimental testing. Nuclear Magnetic Resonance (NMR) spectroscopy can also provide complementary insights into excipient–protein interactions. This review will highlight the underpinning principles of molecular modelling and NMR spectroscopy. It will also discuss the advancements in the applications of computational and NMR approaches in investigating excipient–protein interactions.

Keywords: formulation; excipients; aggregation; NMR; molecular dynamics; molecular docking

1. Introduction

Protein-based therapeutics have become the leading class of pharmaceuticals on the market by sales. The most prominent are monoclonal antibodies (mAbs), which belong to the immunoglobulin family [1]. Currently, most of the monoclonal antibody therapies approved are formulated at protein concentrations of >30 mg/mL [2] and are administered parenterally as liquid injections (intravenous or sub-cutaneous routes) [3]. Due to instability in the gastrointestinal tract, capsular and tablet oral dosage forms have not yet been found to be a practical way to deliver such protein-based drugs [4,5]. Storage of biopharmaceuticals in the preferred liquid aqueous form must achieve a long shelf-life of up to 2 years at 4 °C to be approved [6]. However, the continuous threat of aggregate formation, precipitation, chemical degradation and other modifications has made designing the ideal formulation very challenging [7].

Aggregation is considered to be one of the most pressing challenges, and can occur at all stages of protein pharmaceutical development, manufacture, formulation and storage. The propensity of protein-based products to aggregate during various stages of manufacturing has also been reviewed thoroughly [8,9]. One strategy to address this problem is to understand the mechanisms that dictate aggregation pathways under relevant solution conditions. The currently known range of aggregation pathways and mechanisms have been reviewed extensively [10,11], and are not the focus of this review.

Understanding detailed mechanisms is time-consuming and so the industry is typically under pressure to take an empirical approach to formulation, typically through combinatorial screening, to obtain products that are stable to aggregation, along with other degradation pathways. Such approaches often use convenient biophysical parameter determinations as indicators of aggregation propensity, such as the thermal transition midpoint (T_m) for global conformational unfolding, or the temperature (T_{agg}) at which aggregation is first detected. The capabilities and limitations of these rapid T_m and T_{agg} measurements have been explored extensively, and have been found to correlate well with rates of aggregation at elevated temperatures [12,13]. However, these correlations disappear for aggregation kinetics under low-temperature storage conditions, because the aggregation mechanisms are no longer primarily driven by global protein unfolding [14–17].

Protein-based pharmaceuticals often require the addition of formulation excipients to ensure their stability. The major groups of excipients include: amino acids; buffering agents; sugars; osmolytes and surfactants. There is significant overlap between certain excipient groups, and many excipients carry out more than one function in formulations, including control of aggregation, vial surface adsorption, viscosity, shear-thinning, chemical instabilities, and container instabilities. Although pharmaceutical excipients are readily available, the mechanisms of action for each vary. A number of theories such as preferential interaction/hydration, volume exclusion/crowding, cation- π interactions, and electrostatic interactions, dispersive and aromatic interactions have been hypothesised [18–20].

High-throughput (HTP) methods have been used in the pharmaceutical industry for decades, primarily for identifying lead compounds in drug discovery, and several in depth reviews have focused on their applications [21–23]. Outlined below (Table 1) are some of the prominent high-throughput biophysical methods used to assist the excipient selection process for protein-based formulations. An extensive study on the use of HTP methods to design a formulation for a highly concentrated IgG mAb is reported here [24].

Table 1. High throughput biophysical methods used for excipient screening.

Biophysical Method	Details	Limitations	Application References
Raman spectroscopy	Measures shifts in energy (wavelength) of photons re-emitted after interaction with molecular vibrational modes. Provides an empirical signature of protein structure, that can be used to monitor changes in intramolecular dynamics and intermolecular interactions.	Low sensitivity. Out of the millions of incoming photons interacting with molecules, there is only one scattered Raman photon.	[18,25]
Circular dichroism	Measures the difference in adsorption of circularly polarised light. Far-UV CD can determine the absolute and relative contributions of secondary structure types in proteins. Near UV CD can probe tertiary structure content. Can probe changes in protein structure in response to formulation.	A reference protein with known secondary structure is required to fit the experimental data. The quality of the fit also depends on the wavelengths used.	[18,26]
Isothermal titration calorimetry (ITC)	Measures the heat emitted or absorbed during the titration of a protein with a ligand. The amount of heat indicates the proportion of excipient that binds the protein and its associated enthalpy.	ITC can be used to determine the excipient mechanism directly and indirectly. However, no structural information of the protein is given.	[18,27]

Table 1. Cont.

Biophysical Method	Details	Limitations	Application References
Differential scanning calorimetry (DSC)	Routinely used in high-throughput screening of excipients for formulations. Determines the impact of excipients on the thermal stability of the protein, measured as the melting temperature and enthalpy of unfolding.	Useful for identifying excipients that preferentially interact with proteins, or that stabilise through crowding effects. Cannot be used to detect other mechanisms of action. Unable to characterise changes specific to the secondary or tertiary structure of proteins.	[16,28–30]
Differential scanning fluorimetry (DSF)	Uses a PCR thermocycler to scan the fluorescence of extrinsic dye-binding to proteins as a function of temperature in microtitre plates, and determine their melting temperatures.	The excitation source of the PCR equipment can potential limit the type extrinsic fluorescence dyes used. Unable to characterise excipient mechanisms of action and can only detect tertiary structure changes.	[16,31–33]

Given that high-throughput screening approaches are empirical, and hence time-consuming, and often only provide indirect surrogate evaluations of whether a protein is likely to be stable at low temperature over 2 years, there is a need to create a more fundamental understanding of aggregation mechanisms and the impact of different modes of formulation under these storage conditions. There is also the potential to use computational modelling approaches to both elucidate such mechanisms, and also to refine them to create better modelling methods.

The combination of such approaches with more detailed biophysical analyses, such as NMR thus has significant potential to advance the understanding and predictability of formulation for biologics. A schematic shown in Figure 1 highlights the different capabilities and limitations of molecular modelling techniques and NMR. NMR spectroscopy provides structural and functional insights into molecules under interrogation through observations including the chemical shift and spectral line shape analysis that inform both on molecular structure in addition to intermolecular interactions. Such information has the potential to be utilized to validate and refine molecular modelling approaches designed to predict protein-protein and protein-excipient interactions, as well as the propensity of proteins to aggregate.

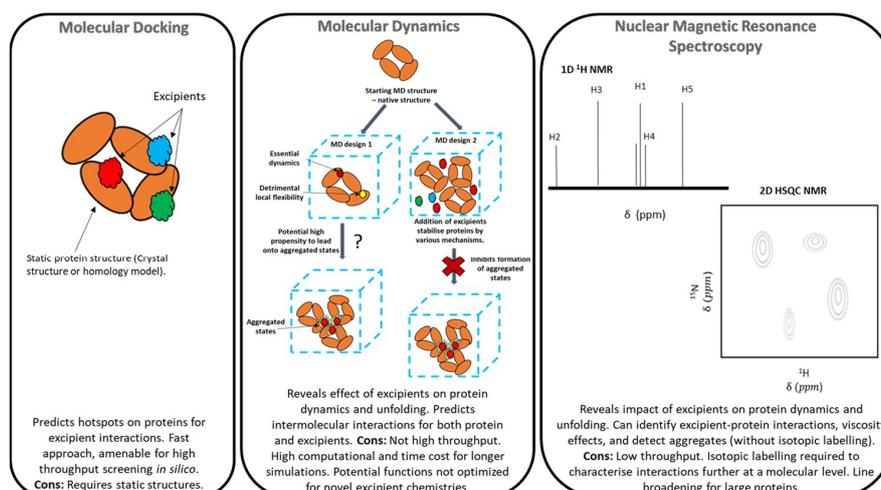


Figure 1. A schematic representation of the potential of *in silico* molecular docking, molecular dynamics simulation, and NMR spectroscopy to elucidate protein-excipient interactions to inform the rational design of protein-based formulations.

We start below with an informative summary of the underpinning principles of computational modelling and NMR spectroscopy. Based on these principles, this review reports on the complementary insights that can be obtained from Nuclear Magnetic Resonance (NMR) spectroscopy and computational modelling methods, to inform the rational design of protein-based formulations. We have a particular focus on the aggregation mechanisms observed under low-temperature storage conditions. These include those mechanisms driven by hot-spot interactions between proteins, and by local fluctuations within the native protein structure ensemble that can reveal aggregation-prone regions (APRs) [17,34–38]. Thus, it would also be useful to elucidate protein-excipient interactions with the potential to modify the ability of proteins to self-associate via hydrophobic hotspots or APRs on the surface, or by local fluctuations that reveal buried APRs.

2. Overview of Molecular Modelling, Methodologies and Limitations

2.1. Molecular Docking

Molecular docking is a computational chemistry tool that is extensively used in drug discovery processes [39], with more recent applications found in the drug development process. This method enables evaluation of intermolecular interactions between two molecules by predicting possible binding modes. The target molecule is a biopolymer, usually a protein, while the binding to a target can be predicted for either a small molecule (ligand-protein docking) as shown below in Figure 2 or another protein (protein-protein docking). The docking typically requires the known structure of the target molecule and seeks for an intermolecular complex with the most favourable binding pose based on the fit between target and its bound molecule, as well as the stabilization resulting from intermolecular interactions. Molecular docking generally involves generating a number of viable protein-ligand conformations by using search algorithms. Initially, a conformation of a complex (binding pose) is predicted in the form of the orientation of two molecules relative to each other. This is followed by estimation of the binding energies for each pose [40]. As absolute binding affinities are difficult to predict, scoring functions have been developed to establish a ranking of the predicted docked poses. This ranking of the generated conformations is achieved by using three main categories of scoring function. Force-field scoring functions use molecular mechanics and are established from calculations of atomic interactions including bond stretching or bending, torsional forces, van der Waals and electrostatic interactions [41]. Empirical scoring functions predict binding energies for docked conformations by considering van der Waals, hydrogen bonding, electrostatics and desolvation terms, whose relative weightings have been optimized through continual validation and refinement [42]. Lastly, knowledge-based scoring functions originate from experimental structural information [43]. The entropic effects contribution, an important component of protein-ligand binding energy, is usually taken into account through re-scoring or employing other methods [44]. A combined use of scoring and re-scoring allows selection of the most probable binding mode for a single ligand, or ranking of a series of ligands according to their predicted affinity for a selected target.

These scoring functions are usually implemented in software packages depending on their intended use, along with the appropriate conformational sampling methods that can also include different levels of molecular flexibility when generating binding poses. Rigid body docking allows for no conformational changes in the molecules of interest, and is thus the fastest. It is suitable for gaining insights into relative orientations of two docked proteins, or for scanning a protein surface with rigid small molecule ligands to identify binding hotspots. Notable docking programs that facilitate rigid protein-protein docking are Hex [45], and GRAMM-X [46]. These software packages sample protein-protein docked conformations based on a Fast Fourier Transform (FFT) algorithm. RosettaDock [47] is also used to predict the lowest energy docked protein-protein complexes by employing Monte Carlo approaches to position molecules in respect to each other, however it takes into account some flexibility by allowing the side-chains to move [47]. Protein-protein docking could provide possible insights into aggregation by identifying putative protein-protein interaction interfaces.

In contrast, flexible docking results in a conformational shift in the molecule of interest. This makes it favorable when studying ligand-protein complexes, and has the potential to predict and improve our understanding of specific interactions at the protein-ligand interface. The wealth of ligand-protein docking software packages available have increased over the decades. Some prominent flexible docking programs include AutoDock 4 [48], GOLD [49,50], and GEMDOCK [51]. The major limitation for molecular docking is the incorporation of protein flexibility. It is a well-known fact that upon ligand binding, a protein undergoes conformational changes. The flexible docking software packages mentioned above overlook ligand-induced binding effects and thus treat the protein as rigid [52]. However, there are software packages dedicated to flexible ligand-flexible protein docking that enable side-chain flexibility, including AutoDock Vina [53] and FlexX [54]. The protein-ligand docking may provide information not only on mode of binding within a specified active site, but it can provide information on possible interaction sites on the whole protein surface [55], which may allow evaluation of excipient binding on the protein surface (Figure 2).

The accuracy of docking predictions that do not account for protein flexibility and presence of other components the solutions (water, ions, buffers) have always been debated. A common solution is to set up molecular dynamics (MD) simulations that can potentially sample various protein conformations upon ligand binding in explicit presence of water molecules and buffers, where a most favourable docking pose can be used as a starting conformation for the simulation experiments.

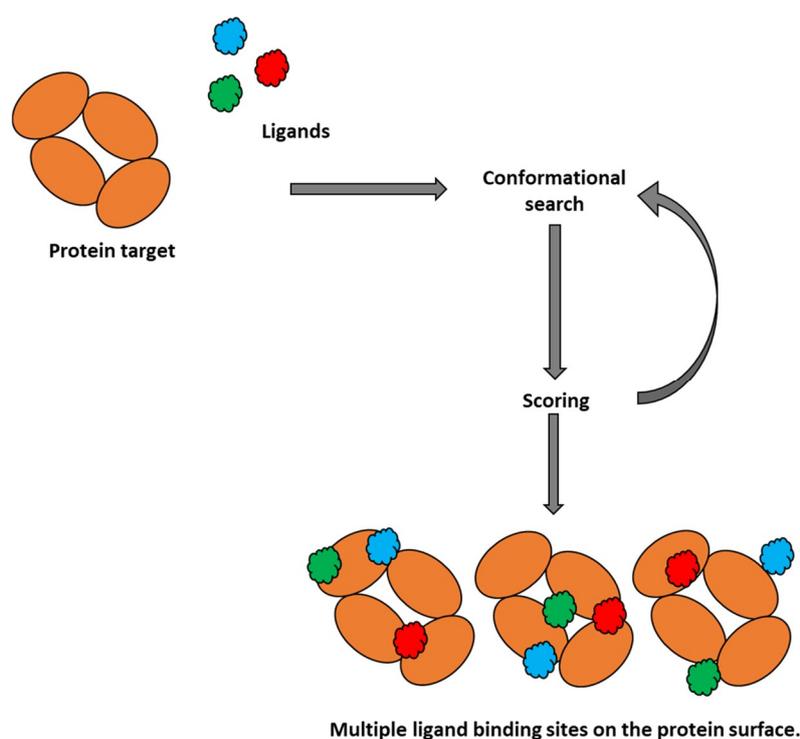


Figure 2. A schematic of the molecular docking process on a single multimeric protein.

2.2. Molecular Dynamics (MD)

Molecular dynamics (MD) is a powerful computational modelling tool, which enables the following of subtle atomic motions of a system of interest as a function of time [56]. All-atom (classical) MD simulations sample configurations by integrating Newton's law of motion to all the atoms in the system simultaneously over a femtosecond time step. A trajectory is recorded with precise atomic positions and velocities giving the user an indication of how the system evolves with time [57]. These positions, defined by Cartesian coordinates of all atoms in a system, allow calculation of the potential energy of the system and forces that act on each atom. Molecular dynamics commonly employs molecular mechanics approximations and use of the force fields, sets of equations and

associated constants to reproduce geometries of molecular systems [58]. These are calculated based on the some commonly used force fields including CHARMM22 [59], CHARMM27 [60], AMBER [61] and GROMOS [62]. There is a wide array of software packages available, including but not limited to Gromacs [63], Amber [64], NAMD [65] and CHARMM [66], that have proven invaluable for the advancement of molecular dynamics. A general summary of the molecular dynamics process is provided below in Figure 3. A list of typically employed MD simulation times for various protein dynamics events are listed below in Table 2. The main limitation of MD is the small time-step, relative to a typical requirement for much longer simulation times. Longer simulations have been carried out on the millisecond time scale [67,68]. This may potentially place a strain on computational resources when simulating larger systems, however the benefits outweigh limitations due to the wealth of information that can be obtained about the system on the atomistic level [69].

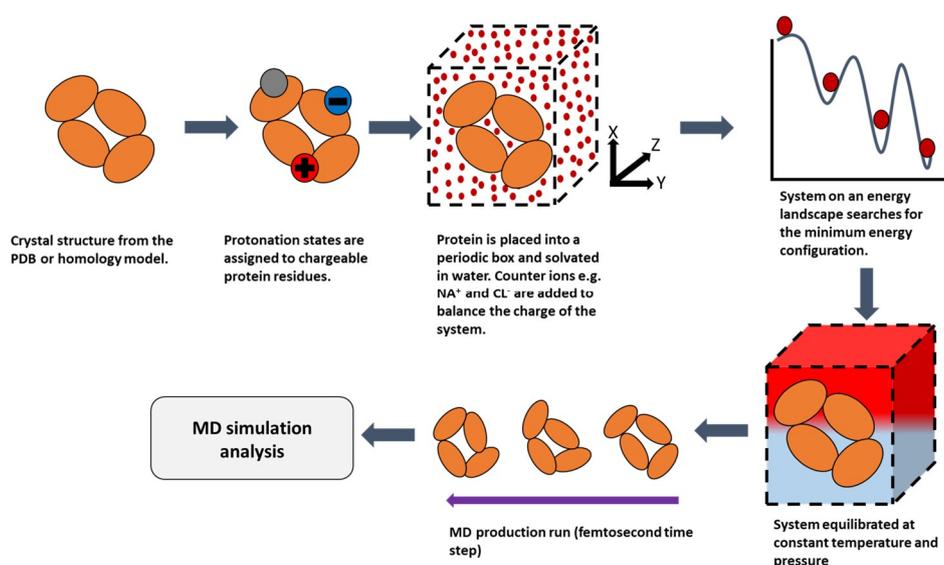


Figure 3. A schematic of the molecular dynamics process.

Table 2. The typical molecular dynamics (MD) simulations timescales that can observe various protein dynamics events.

Protein Dynamics Event	MD Simulation Time Range
Vibrational motions	Femtoseconds (10^{-15}) to picoseconds (10^{-12})
Rotational motions	Picoseconds (10^{-12}) to nanoseconds (10^{-9})
Loop dynamics	Picoseconds (10^{-12}) to milliseconds (10^{-3})
Ligand binding/unbinding	Nanoseconds (10^{-9}) to seconds
Protein folding/unfolding	Microseconds (10^{-6}) to seconds
Aggregation	Seconds and beyond

3. Overview of Nuclear Magnetic Resonance (NMR) Spectroscopy

Nuclear Magnetic Resonance (NMR) spectroscopy detects nuclei of isotopes with spin angular momentum (e.g., ^1H , ^{15}N , ^{13}C and ^{19}F) in a magnetic field with the goal to provide analytical information regarding the structural and/or physical nature of any molecule under investigation. The resultant NMR spectra contain highly defined resonances (often referred to as peaks) that reflect the chemical nature of specific atoms in the molecule of interest. NMR has been applied in a diverse range of pharmaceutical applications that include, characterisation of proteins, drug discovery and design [70]. Despite the power and utility of NMR spectroscopy, linewidths of resonance peaks become larger as molecular weight increases. This due to a reduction in nuclear T_2 relaxation times as the global tumbling of the molecule becomes slower. However, all is not lost as advances in NMR have allowed protein observations to be made over a wide-range of molecular timescales e.g., domain shifts

(microseconds to milliseconds), side chain motions (picoseconds to nanoseconds) and changes in loop regions (nanoseconds to microseconds) [71,72].

For the purpose of this review, we are considering solution-state NMR methods where there are several observable NMR parameters including chemical shifts (δ) and relaxation times (T_1 and T_2). A chemical shift of a nucleus in a molecule arises due to a nuclear shielding or de-shielding effect of the NMR static applied magnetic field and is the result of electrons surrounding the nucleus. Chemical shifts (δ) are measured in parts per million (ppm) [73,74]. The power of chemical shifts provides structural information by distinguishing chemically-equivalent nuclei but where they are in different molecular environments. For example, the proton (^1H) NMR of acetone (CH_3COCH_3) will only display one peak from two equivalent methyl groups but ethanol ($\text{CH}_3\text{CH}_2\text{OH}$) provides three ^1H NMR peaks representing those in the methyl (CH_3), methylene (CH_2) and hydroxyl (OH) moieties respectively. These ethanol peaks are additionally subject to fine-structural splitting from spin-spin J-coupling, a concept taught to many undergraduates and found in NMR and organic chemistry textbooks. This example is a simplistic view of the power within chemical shifts and it cannot be stressed enough that structural and chemical environment influence NMR peak observations. For example, two alanine residues within a protein will both have methyl groups, but their observed chemical shifts will be dependent on their individual structural environments and so display diverse chemical shifts. Chemical shifts are ultimately derived from the Larmor frequency of resonance of each NMR peak, where the distance (in Hertz) between any two peaks is dependent also on the spectrometer magnetic field strength, because the Larmor frequency is proportional to magnetic field strength. When defining peak separation in terms of chemical shift, a 1 ppm separation for a proton (^1H) spectrum on a 600 MHz (14.1 Tesla) spectrometer equates to a frequency separation of 600 Hz whereas for an 800 MHz (18.7 Tesla) spectrometer it would be 800 Hz. This field dependence of peak separation can be extremely useful when monitoring chemical exchange processes, as will be described further below.

NMR is a spectroscopic process that involves excitation and energy transfer between nuclear spin states which subsequently need to relax and return to equilibrium. This NMR relaxation is of fundamental interest to NMR spectroscopists as the time constants (T_1 and T_2) and rates (R_1 and R_2) of the associated mechanisms provide a wealth of information about the molecules being studied. Indeed, nuclear relaxation, and in particular relaxation dispersion, has found use as a tool to study protein folding and enzyme kinetics where chemical exchange events of interest, between species or molecular environments, can be quantified through their influence on NMR relaxation processes, as described in [75–77]. Each NMR-active nucleus in a molecule will possess a unique set of relaxation times from many fundamental molecular and bulk solution properties which is beyond this review. However, it is useful to note that molecular size and shape, proximity and the isotope of the nearest NMR-active nucleus, in addition to temperature, viscosity and the NMR magnetic field strength, all influence nuclear relaxation in quantifiable ways. The spin-lattice mechanism involves a nuclear spin exchanging energy with its surroundings (i.e., lattice), and thereafter returning to its ground state with time-constant T_1 and relaxation rate R_1 . The second spin-spin relaxation mechanism involves the loss of phase coherence and subsequent loss of bulk magnetization that underpins the NMR signal, with time-constant T_2 and relaxation rate R_2 . For both processes, $R_x = 1/T_x$. Although the processes giving R_1 and R_2 are independent events, the R_1 and R_2 relaxation rates converge for small molecules e.g., acetone and ethanol, whereas for larger species such as proteins in solution R_1 and R_2 are found to be significantly different. Furthermore, protein relaxation is extremely powerful and can differentiate the shape and internal motions of a molecule and reviews are available on this subject [78,79].

The linewidth of the individual NMR signals is proportional to R_2 and therefore to the correlation (time of the molecule). The crucial point is that R_2 has a proportionality to the correlation time (τ_c); the time for a molecule to rotate through 1 radian (i.e., molecular motion) which is subsequently related to molecular size and shape. Assuming a sphere, correlation time can be estimated from $[4\pi\eta r^3/3k_b T]$ where η —viscosity of the solvent, k_b —Boltzmann constant, T —temperature and r the radius of the

sphere. As a molecule becomes larger, r increases and so τ_c increases, which then makes R_2 increase and is finally observed in NMR as an increase in linewidth. This is the fundamental reason why protein NMR spectra are broader than small molecule spectra. Increasing the temperature and/or lowering the viscosity would drop linewidth but many proteins are limited in their operational solvent conditions and thermal limits. Taking this concept of molecular size influencing NMR a step further, a ligand that binds to a protein would experience a significantly different correlation time, that is dictated by the size of the complex, compared to when the same molecule is free solution. In many cases, ligand binding is not permanent, but dynamic and is defined by an equilibrium with an affinity that can further influence the NMR observation [80].

When a molecule is in equilibrium between two or more states, the rate of exchange determines whether the chemical shifts of each of each state is visible or the chemical shift of a single time-averaged state is observed. The observation is thus dictated by the equilibria involved in exchange as well as the difference in nuclear spin relaxation rates R_1 and R_2 . These two limits are known as slow and fast exchange in the chemical shift timescale but there is an intermediate condition where all resonances become broad and difficult to detect. At this condition, known as coalescence, the exchange rate constant is comparable to the difference (in Hz) between the Larmor frequencies of the two states. This ties in with the earlier concept of chemical shifts between peaks being frequency dependent and as the relationship between chemical shift and the equilibrium is not based on parts per million (ppm) but the fundamental Hertz distance between NMR resonances. Therefore, changing NMR field strengths can help move the observation between the slow, intermediate and fast exchange regimes.

3.1. Limitations

The potential for NMR spectra to become very challenging to interpret is very real when increasing the molecular weight of the protein. The number of proton resonances scales proportionally with molecular weight, but the spectral width remains constant, leading to increased crowding of peaks. In addition, line width increases with the molecular weight, leading to a further decrease in resolution and in signal intensity with the consequence that, resonance overlap increases rapidly with molecular size, until individual lines become too difficult to resolve. A commonly utilized workaround is to expand the NMR spectra from one to more dimensions (2D, 3D, 4D, etc.) to create more special representation of resonance peaks. Alternatively, data can be obtained at higher magnetic fields (e.g., at 14.1, 18.8, 21.1 and 23.5 Tesla that equates to 600, 800, 900 and 1000 MHz ^1H resonant frequencies) to improve resolution. Other approaches include the addition of alternative NMR-active nuclei than ^1H , such as ^{15}N and/or ^{13}C in addition to reducing spin-spin (R_2) based relaxation by deuteration; the process of replacing ^1H with ^2H . As discussed, R_2 is responsible for line broadening and reducing this process will create narrower lines. However, as discussed further below, TROSY NMR can be used to select slower NMR relaxation pathways to and therefore observe narrower lines, which then extends the range of molecular weight for which peaks can be resolved. Such a method is at its most powerful when combined with deuteration.

3.2. Protein-Observe Methods

As mentioned, the chemical shift is sensitive to the molecular environment around the nucleus and when a ligand or excipient interacts with a protein, physical characteristics for both are altered. The binding even creates a change in electron density and so influences the most prominent observable NMR parameter, chemical shift. As a result, chemical shift mapping (CSM) or chemical shift perturbation (CSP) methods are potential modes of investigation in protein-observe NMR [81]. CSM/CSP methods compare two protein NMR spectra, such as with and without the addition of a ligand, and track any changes in chemical shift and/or disappearance of peaks. These changes will identify any areas of the protein that are influenced by the binding event, with the largest changes being typically observed in the region around the binding site. In addition, careful experimental

design can utilize a suite of NMR spectra, obtained over a range of ligand concentrations, to provide isotherms from which dissociation constants for the binding process can be determined.

High molecular weight protein species such as whole mAb, Fab fragment and Fc region can be expressed with isotopic labelled nuclei, which enable greater resolution of peaks through the collection of multidimensional heteronuclear correlation spectra. There are several labelling protocols that exist for NMR studies. Uniform ^{15}N isotopic enrichment is the simplest form of labelling a protein. The protein is expressed in *E. coli* (BL21) [82] grown on minimal growth media and supplemented by $^{15}\text{NH}_4\text{Cl}$ and unlabelled glycerol/glucose that is purified using standard methods to provide a +90% enriched protein product as shown in the schematic below in Figure 4. Heteronuclear single quantum coherence (HSQC) NMR spectra are 2D NMR experiments recorded to show all nitrogen-hydrogen correlations, which typically are dominated by backbone amide groups from the protein [83]. In the case of studying excipient–protein interactions, a series of HSQC NMR spectra can be recorded of the protein in presence and in the absence of the ligand. Binding effects of these interactions can be investigated by overlaying the series of ^{15}N HSQC spectrums. If there is an interaction between the excipient and protein, the peaks on the spectra will be found to track from their initial position when no ligand was present [84].

More complex labelling approaches exist, such as $^2\text{H}/^{15}\text{N}/^{13}\text{C}$ triple labelling which uses bacteria, yeast or insect cells as an expression system supplemented by $^{15}\text{NH}_4\text{Cl}$, ^{13}C -glucose and deuterated water (D_2O). This form of labelling enables detailed mapping of whole structural changes in the protein upon excipient binding as it can offer significant coverage of the protein backbone and side chains. Deuteration replaces the aliphatic and aromatic protons for ^2H across the protein to reduce the significant nuclear relaxation effect of ^1H nuclei that facilitate line broadening. The proportion of ^1H to ^2H depends greatly on the carbon source used and whether H_2O or D_2O is present [85]. However, using specifically labelled methyl group probes ($^{13}\text{CH}_3$) in the study of high molecular weight proteins has proven useful in expanding the molecular weight limit reached with perdeuterated proteins [86]. Methyl groups tend to occur in the hydrophobic cores of proteins and provide an excellent probe for conformational modifications [87,88]. Several protocols for methyl group labelling are available [89–92].

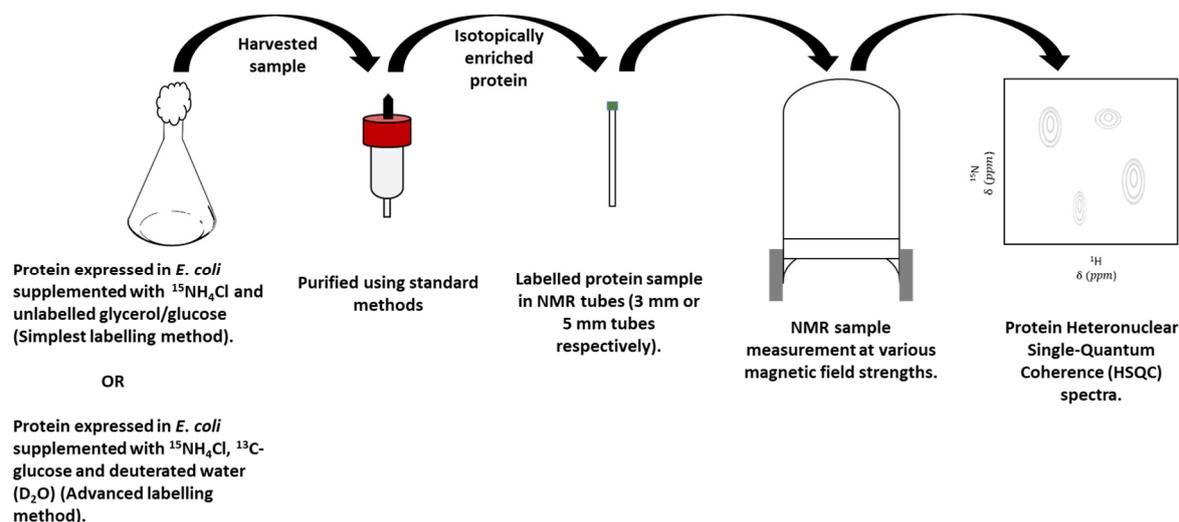


Figure 4. A schematic depicting the bacterial production of recombinant isotopically labelled protein, and recording of an NMR spectrum.

The increased relaxation already mentioned for proteins increases with increasing magnetic field strength, which creates additional issues. Consequently, several protein-NMR methodological advances have been utilized to address this challenge, the most popular being Transverse relaxation-optimised spectroscopy (TROSY) [93,94]. TROSY which uses spectroscopic means to reduce the observed T_2 relaxation by selecting the slower relaxation pathway for observation. Remember, T_2 is

related to line width and accessing the slowest relaxing pathway will produce narrower resonance lines. Therefore, TROSY accompanied by various isotopic labelling techniques especially deuteration allows the study of biomolecules above 25–30 kDa [95]. The effect of TROSY on the relaxation rate of an excipients and proteins is demonstrated in Figure 5.

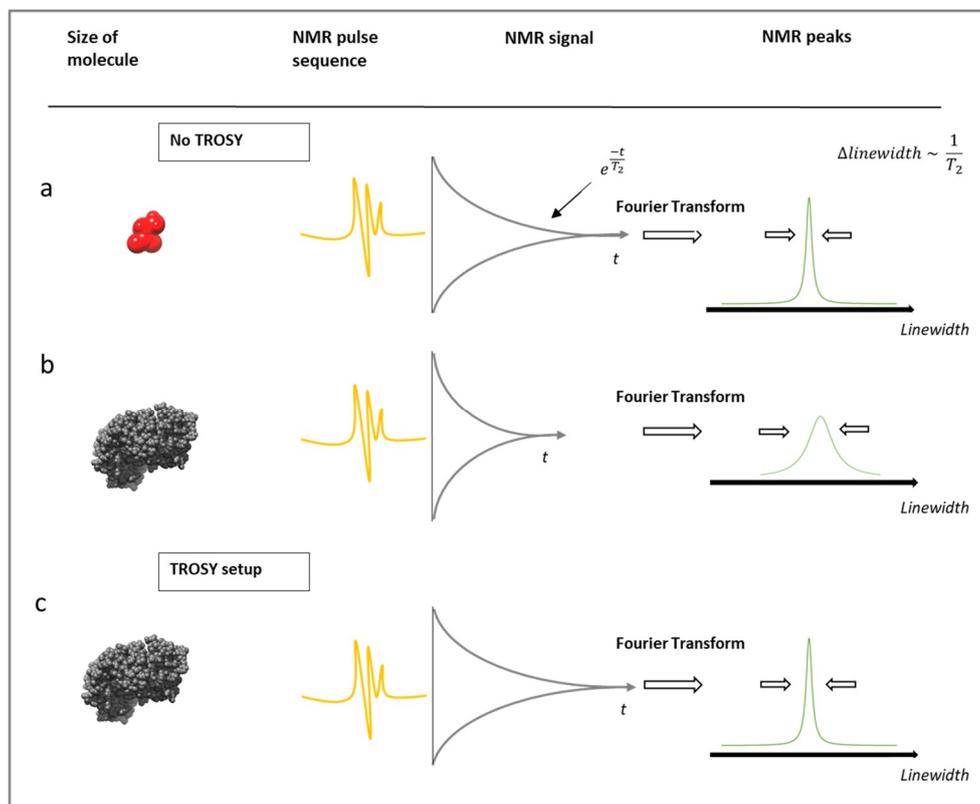


Figure 5. (a) The NMR spectrum for a small molecule ligand will depict a narrow linewidth due to a longer transverse relaxation time. (b) In contrast, a protein has a shorter transverse relaxation time and thus a broad linewidth is shown. (c) TROSY prolongs the transverse relaxation times and thereby improves the protein signal in the spectrum.

TROSY, when combined with higher magnetic fields (e.g., equal to or greater than 18.8 T–800 MHz ^1H) and the introduction of cryogenic probes has enabled NMR fingerprinting using $^1\text{H}/^2\text{H}$, ^{15}N ^{13}C isotopes at natural abundance of full mAb, Fab, Fc and other therapeutic proteins possible [96–99].

3.3. Ligand-Observe Methods

Ligand observe methods exploit the dependence of relaxation rates on molecular size described earlier in this review, specifically via the difference in molecular weight between a small ligand and a significantly larger protein. Nuclear Overhauser Effects (NOEs) are defined as the change in intensity of NMR resonances caused by dipole-dipole coupling. The sign and the magnitude of the NOE is dictated by the hydrodynamic radius (r^6) and the correlation time (τ_c). This makes NOEs available to detect intramolecular and intermolecular interactions. Large molecules tend to tumble slower and so it is anticipated that a negative NOE will be observed. In contrast ligands tumble fast resulting in a positive NOE [100]. Two prominent NOE based NMR methods include Saturation Transfer Difference (STD) and WaterLOGSY (Water Ligand-Observed via Gradient spectroscopy).

STD NMR involves recording a STD_{on} spectrum, whereby only the ^1H protons of a protein and not the excipient (ligand) are selectively saturated by a narrow selective ^1H NMR pulse typically placed between -1 and -3 ppm. The saturation is transferred throughout the protein via spin diffusion and onto any bound ligand, resulting in a reduction in the intensity of the ligand resonances as shown

in Figure 6a. In order to detect this transfer, the ligand must dissociate from the protein. Therefore, the dissociation constant (K_D), which describes the affinity of an excipient to the protein, has to be favourable for this process. A favourable K_D also allows multiple excipient molecules to bind and dissociate, each receiving saturation from the protein and increasing the observable signal in the STD_{on} experiment, as compared to when a small number of molecules bind during saturation. STD requires a difference spectrum which is created by subtracting a second control experiment spectrum (STD_{off}) acquired when no saturation of the protein takes place and is usually created by saturating outside the protein chemical shift envelope. In reality these two experiments (STD_{on} and STD_{off}) are acquired in a single interleaved experiment and processed simultaneously using a spectrometer macro to provide both 1D datasets STD_{off} and STD_{diff} (a spectrum automatically created as the subtraction of STD_{off} and STD_{on}). The nature of the transfer from protein to ligand uses a negative NOE which manifests for any peaks from nuclei involved in saturation transfer having lower magnitude signals in the a STD_{on} spectrum when compared to the STD_{off} spectrum. Any nuclei not involved in saturation transfer will have resonances of identical magnitude in both STD_{off} and STD_{on} . The resulting difference spectrum between STD_{off} and STD_{on} will display only resonances where saturation transfer has occurred [101]. Additionally, and with careful experimental set-up, the magnitude of the difference spectrum also provides an indication of the orientation of the molecule upon binding to the protein.

WaterLOGSY [102,103] operates by observing a NOE between the ligand and water molecules. The water molecules are either in the bulk solution or in the vicinity of a protein's surface. In the case of the latter, the water molecules take up the tumbling characteristics of the protein, resulting in a negative NOE upon ligand binding. Conversely, the ligand molecules interacting with the bulk water will result in positive NOE, by inheriting the tumbling characteristics of the water molecules. Binders and non-binders can be identified via a spectrum displaying negative and/or positive peaks. Control experiments with and without the protein are required to ascertain the sign of the positive NOE (Figure 6c).

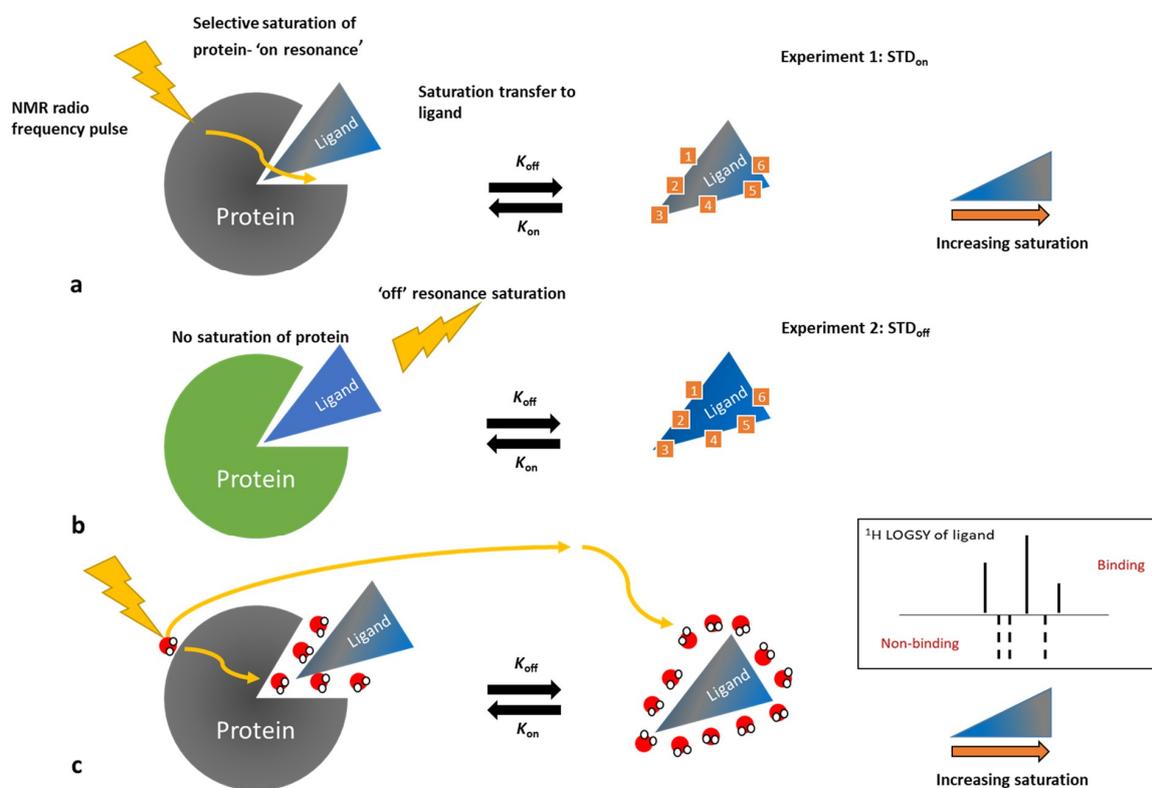


Figure 6. Schematic representation of saturation transfer difference (STD) NMR (a,b) adapted from [101] and WaterLOGSY (c) NMR. Increasing saturation of the ligand's resonances is indicated by a colour gradient from blue (no saturation) to grey (high saturation).

Diffusion based experiments such as pulsed-field gradient (PFG)-based pulse sequence are another group of ligand-observe experiments. The PFG experiments enable the study the general molecular displacements occurring in either complex or simple mixtures. Diffusion measurements can be obtained by experiments using either spin echo (SE) or stimulated-echo (STE). When combined with the acronym PFG, the full abbreviation becomes PFGSE and PFGSTE, respectively. Diffusion data can be demonstrated in many ways as diffusion curves or 2D maps. A pictorial representation of a DOSY spectrum as a 2D map is common, with one dimension constituting to chemical shifts and the other dimension representing diffusion coefficients. In addition to identifying the various components of a mixture, DOSY can also offer an insight into the hydrodynamics of the molecular system of interest by observing the self-diffusion of molecules in solution [104]. DOSY works by utilizing pulsed-field gradient NMR spectroscopy to measure translational diffusion of molecules where the pulsed-field gradient can spatially label resonances in a molecule of interest.

If 2D DOSY maps are used, they should be accompanied with diffusion curves that display the NMR peak intensity as the NMR gradients are perturbed. This curve can be analysed to obtain a diffusion constant. It is also important to define the process of DOSY calibration using known diffusion standards such as dioxane. PFG experiments can provide insights into mixtures containing excipients and these will be highlighted in the next section.

4. Nuclear Magnetic Resonance (NMR) Spectroscopy Applications in Aggregation and Formulation

Here we discuss the applications of NMR already implemented in the study of aggregation and formulation design. We also discuss the significance of the examples provided. A key component of many protein aggregation mechanisms is the formation of oligomeric protein complexes that are thermodynamically unfavourable, and therefore exist only transiently and at very low populations. Heteronuclear spin relaxation rates have been measured previously to determine weak association constants for the transient formation of oligomers of bovine low molecular weight protein tyrosine phosphatase (BPTP) under equilibrium conditions [105]. The approach combined hydrodynamic calculations with the conventional measurement of R_1 (longitudinal) and R_2 (transverse) at different BPTP concentrations, to reveal the formation of tetramers, and also that the tetramerisation interface was formed by a cluster of residues on the surface of the dimer. However, in irreversible protein aggregation mechanisms the soluble oligomers form under pseudo-equilibrium as transient and rarely populated intermediates, and so their detection by NMR is far more challenging. A significant amount of work using NMR has been focused on amyloid fibril aggregates which play a huge role several prominent diseases such as Parkinson's and Alzheimer's. Fibril formation of α -synuclein (α Syn), responsible for Parkinson's disease has been studied using Paramagnetic Relaxation Enhancement (PRE) NMR to depict the various contacts between heterogeneous disordered monomers. PRE NMR involves a nitroxide spin-label being attached to a particular protein region and while in its oxidized state (paramagnetic) improves the relaxation during the ^1H - ^{15}N HSQC experiment [106]. PRE-based contacts exceed the conventional NOE distances ($\leq 5 \text{ \AA}$) by at least 4-fold and provide probes that interrogate molecular structure over long-ranges [107].

The exchange dynamics between amyloid- β ($\text{A}\beta$) monomers and polydisperse, NMR-invisible ('dark') protofibrils was investigated by a novel technique called Dark-state Exchange Saturation Transfer (DEST). DEST follows similar principles to that of Saturation Transfer Difference (STD) NMR. This structural and kinetic study of the protofibril formation was highly significant as the build-up of toxic, soluble aggregate forms of $\text{A}\beta$, then forms larger assemblies which contribute to the development of Alzheimer's disease. The main findings showed ^{15}N - R_2 values being significantly larger for $\text{A}\beta_{42}$ than the closely related variant $\text{A}\beta_{40}$. This supported the known observation that $\text{A}\beta_{42}$ demonstrated a higher propensity for rapid aggregation and fibril formation than $\text{A}\beta_{40}$ [108].

The human immunoglobulin kIV light chain variable domain (LEN) has the potential to be converted into amyloid under stress conditions. CPMG [109,110] relaxation NMR experiments were used to identify residues undergoing slow millisecond motions. Multidimensional solution NMR

experiments were implemented at physiological and acidic pH. The main findings revealed that certain flexible residues at the dimer interface drive the formation of partially misfolded conformers. By identifying the specific residues and regions which contribute to the early stages of unfolding, such work may pave the way for the rational design of stable tertiary and even quaternary structures, that prevent aggregation [111].

NMR spectroscopy considers the structural properties of therapeutic proteins, however overlook the presence of a solvent, typically water. The solvent plays a vital role in influencing protein dynamics. A new method called water proton NMR was proposed to use the transverse relaxation (T_2) time of water protons to quantify protein aggregation. BSA (66 kDa) and γ -globulin (150 kDa) were subjected to temperature-induced aggregation. The T_2 of water protons increased linearly with the percentage of aggregate formation. The correlation was consistent at high and low magnetic fields [112]. Furthermore, a pH-induced aggregation procedure was also implemented on human insulin. A non-linear trend between T_2 of water protons and aggregates was observed in this case [113]. The water NMR method was extended to investigate mAb aggregation under various stress conditions, in which the transverse relaxation of water protons detected aggregate formation [114]. The NMR technique itself provides a rapid and non-invasive method in characterising the extent of aggregation in finished products.

A combined approach using DOSY-NMR and DLS was used to measure diffusion coefficients and particle size distributions for five commercially available insulin drug products. The authors revealed that DLS was more effective in detecting larger aggregates than DOSY-NMR due to the higher sensitivity to high molecular weight species. In contrast DOSY-NMR was found to be more suitable in detailing excipient behaviour in the formulation [115]. The findings do not diminish the potential of NMR in the study of aggregates, but rather emphasize the importance in selecting the appropriate NMR method based on prior information about the system being investigated.

There has been a significant contribution towards characterising solid dosage form formulations with techniques such as ^{15}N Dynamic Nuclear Polarisation [116] and ^{13}C Magic angle spinning NMR [117–119]. However, the applications of NMR on solid dosage formulations is beyond the scope of this review.

Antibody-based therapeutics are frequently formulated with small concentrations of sugars, amino acids, buffer salts and polysorbates. The presence of such excipients may potentially induce protein structural changes which solution NMR can pin-point. Several notable NMR studies have been carried out to assess the quality of new-age bio-therapeutics. Panjawani et al. have implemented 2D NMR fingerprinting assays to detect the effects of excipients and pH conditions on the conformations of two interferon (IFN) proteins (α -2a and α -2b). NMR spectra were recorded for both proteins and compared to reference spectra already recorded by regulatory agencies. The first stage of the formulation process was to add excipients used in products such as Roferon-A[®] and Intron-A[®]. A deformation process using Cation exchange chromatography was implemented to analyse the various components of the formulation. A series of 2D HSQC NMR spectra were recorded following the formulation and deformation process under various pH conditions. The study revealed that there was no alteration of the IFN structure during the deformation process. Below pH 3, the protein was found to unfold and at pH 4.5 the NMR spectra showed a tendency of the protein to oligomerise even though the tertiary structure was intact [120]. Similar NMR methods have been implemented on formulations relating to recombinant methionyl human GCSF (Neupogen[®] by Amgen) in preparation for subsequent biosimilars entry into the market [121] and also on antibodies in presence of common excipients, Tween[®] [122]. The three studies highlighted above all demonstrate the potential of NMR to guide the formulation process by providing structural insights into changes in the active ingredient under various formulation conditions.

The NMR methods for formulation design so far discussed all involve a form of isotopic labelling. However, Golovanov and co-workers have attempted to address the challenge of working directly with industrial formulation samples, by using 1D ^1H NMR methods. One dimensional NMR spectra were recorded for mAbs in presence of an excipient mixture arginine glutamate. The ^1H NMR experiments

helped to identify conditions whereby protein-protein interactions were restricted. They also revealed that translational diffusion measurements were less useful than transverse relaxation data in finding the most suited formulation [123]. The other ligand-based approach mentioned earlier was diffusion NMR, this technique was used to differentiate NMR signals of the excipient that may potentially mask mAb and insulin NMR signals [124–126]. Although, 1D ^1H NMR spectra are unsuitable for in-depth structural understanding of proteins, they can still provide an initial insight into whether the protein of interest is in its folded or unfolded state within formulations.

5. Molecular Modelling Applications in Aggregation and Formulation Design

Ligand-protein and protein-protein interaction sites are functionally important for modulating protein stability, including potentially for the control of aggregation in therapeutic protein formulations. Steering the design of such sites can be challenging as often there are large surface areas involved, structural dynamics and the presence of solvent molecules [127,128]. Strategies to engineer these sites include mutating residues that are involved in the ligand-protein and protein-protein interactions. Although there are a significant number of available protein and peptide structures obtained by NMR or X-crystallography (PDB), there is a lack of experimental information on protein-protein interfaces, particularly those involved in protein aggregation. MD has been found to be a useful tool in predicting the influences of various factors on protein structures, including guiding site-directed mutagenesis [129].

Protein unfolding may lead to aggregated states, which itself presents as an immense challenge in maintaining conformational stability. Daggett and Levitt have simulated unfolding pathways using MD to understand the mechanisms behind protein folding [130]. MD has also been found to be crucial in providing molecular insights into the amyloidogenesis process. Simulations at neutral and lower pH conditions were carried out to characterise the conformational changes of a prion protein from its native cellular isoform (PrP^{C}) to an infectious form (PrP^{Sc}). The latter is known to lead to neurodegenerative diseases. The protein structure was found to be intact at a neutral pH, whereas at lower pH there was more flexibility in the structure. Furthermore, the total sheet-like structure increased via the native β -sheet and an additional portion formed in the N terminal of PrP [131,132].

Daggett and co-workers hypothesized that rarely formed α -sheet structures are shared by amyloidogenic proteins which are linked to toxicity. Molecular Dynamics was used to generate structures for the rational design of novel anti- α -sheet peptides which were also tested experimentally. These α -sheet structures provide a suitable target for neutralising the toxicity and preventing fibril formation [133]. The novel α sheet design has the potential to prevent aggregation in several amyloid proteins $\text{A}\beta$ Alzheimer's and amylin (type 2 diabetes) as they bind to the toxic species [134].

An extensive review of the wealth of computational tools to determine aggregation-prone hotspots has already been carried out. The review highlights specific regions in either the structure or in the sequence of a protein that induces aggregation. It proposes a list of sequence based computational methods whereby Aggregation Prone Regions (APRs) can be identified [135]. Some of the more prominent tools include; AGGRESCAN [136], TANGO [137] and PASTA [138]. A structural tool to identify APRs has been used on therapeutically relevant mAbs. The Spatial Aggregation Propensity (SAP) method employs molecular dynamics to map fluctuations and identify the number of aggregation prone hydrophobic regions exposed on the antibody surface. These hydrophobic patches can be exposed natively, via fluctuations or conformational shifts and can be observed by molecular simulations. A high SAP value meant that there was a presence of APRs and these were used to inform of potential mutations that inherently improved antibody stability [139–141].

Computational simulations can potentially provide molecular insights into drug formulations. Although small molecule formulations are beyond the scope of this review, there are two notable modelling studies that employ docking and MD to study their respective formulations. A combined approach using both methods was used to investigate the interactions between hydrophilic excipients; lactose, hydroxypropyl methyl cellulose (HPMC), mannitol and the poorly soluble, Bicalutamide (BIC).

BIC is a non-steroidal antiandrogen drug used to treat prostate cancer. A Lamarckian genetic algorithm within the AutoDock 4 software package [48] was used to seek out the lowest binding energy of BIC-excipient during docking. The best conformation of each BIC-excipient complex was then used as the starting MD structure. MD simulations of the docked complexes revealed Lennard-Jones interactions between BIC-HMPC/mannitol and coulomb interactions between BIC-lactose. Lactose formed the most hydrogen bonds with water and provided the best dissolution performance in experimental studies [142]. Jha and Larson conducted a detailed molecular dynamics study to assess the effects of polymeric excipients on a Phenytoin, which exhibits poor solubility in water. This work demonstrates the use of MD utilities such as radial distribution functions between API-API molecules to characterise aggregation and also between excipient-API to flesh out potential interactions [143].

As mentioned earlier, excipients interact with proteins via many different mechanisms. Arginine is particular has several mechanisms of actions, which makes it fascinating to study by MD. A series of aqueous molecular simulations to investigate arginine's aggregation inhibition properties were set up. The simulations revealed arginine molecules had the tendency to form hydrogen bonded clusters when in solution. The hydrogen bonds were found to be stronger within clusters than those between arginine and water. A similar cluster formation was observed upon addition of proteins. The clusters effectively crowded out protein-protein interactions. By contrast, cation- π interactions between arginine and the protein were found to stabilise the unfolded intermediates [144].

In order for a protein to be conformationally stable, it would have to demonstrate good solubility in water. An excipient mixture of arginine (L-Arg) and glutamate (L-Glu) has been explored experimentally as a way forward to inhibit aggregation in proteins [145,146]. Preferential interaction coefficients were derived in simulations to investigate this behaviour further. Preferential interaction coefficients are a measure of the excess number of excipient molecules in the vicinity of a protein compared to that in the bulk solution. The simulation conditions included a combination of Arg-Glu mixtures at a range of concentrations, as well as single component L-arginine and L-glutamate system in the presence of the well-characterised *Drosophila* Su dx protein WW34. The main findings indicated that the equimolar mixture system enhanced the protein solubility more than systems consisting of single co-solvent components. Again, the increase in crowding and hydrogen-bond formation between co-solvent molecules led to a suppression of protein-protein interactions [147]. Another, MD study investigated Arginine's role as an eluent during Protein A chromatography. The main purpose of this complementary study was to witness the effects of adding an excipient in an attempt to limit aggregate formation at a crucial manufacturing step. Addition of arginine was successful in disrupting the affinity between Protein A and an antibody. Meanwhile, another excipient, citrate, was found to reverse the process [148].

Similarly, MD simulations have been carried out with other excipients such as polyethylene glycol (PEG) [149], polyvinyl alcohol [150], trehalose and its derivatives [151–153]. Mechanistic insights into excipient interactions has also been studied using molecular dynamics in freeze-dried systems [154,155]. A recent computational study has implemented molecular docking approaches to identify binding hotspots on a Fab A33 fragment surface, for a set of commercially available excipients as shown below in Figure 7. The excipients selected for the study were from various categories such as amino acids, sugars, surfactants and osmolytes. Each excipient pose was characterized in terms of their predicted *in silico* binding energy, and also the Fab residues with which they interacted. All eight excipients were found to bind on three particular hotspots. Protein-protein docking was implemented on two Fab A33 molecules using Hex [45] and Grammx [46]. This allowed an appraisal of the interacting Fab residues with excipients and whether they coincided with protein-protein interaction sites. The regions through which two Fab molecules are predicted by docking to interact during aggregation, can potentially be shielded by adding excipients that also bind to those regions. The presence of such an effect was validated with the experimental determination of thermal stability for each formulation [156].

There have been efforts to demonstrate the possibility of using non-conventional classes of excipients such as anti-inflammatory drugs to enhance therapeutic protein stability. For example, the anti-inflammatory drug Dexamethasone phosphate was docked onto an in silico-built model of the Bevacizumab[®]–Bevacizumab[®] interface to identify potential binding sites. A region on the Fc structure was found to interact with a Fab fragment on the second bevacizumab molecule. An interaction with Dexamethasone phosphate and Lys 445 on the protein was revealed to potentially suppress dimerization. Bevacizumab[®] is a humanised whole mAb that is used to treat various cancers and eye disease [157,158].

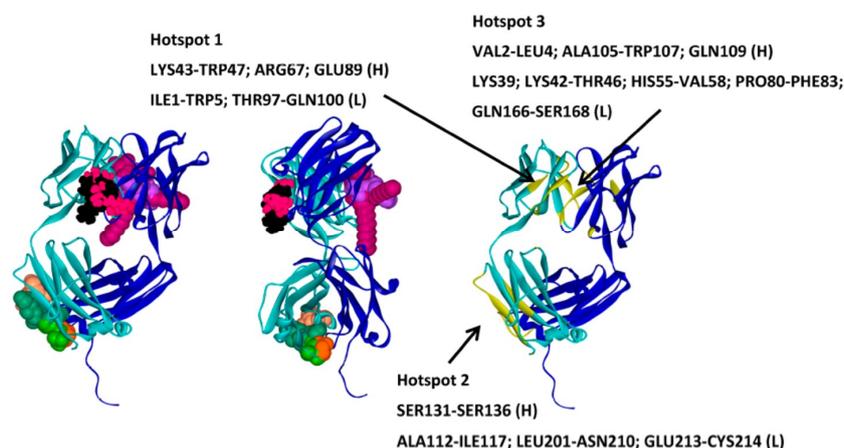


Figure 7. Binding hotspots for eight commercial excipients on the A33Fab surface. Reprinted from [156] with permission.

6. Future Perspectives

Molecular modelling and NMR spectroscopy are promising techniques for improving the understanding of the structure and dynamics of proteins and the impact of formulation excipients. However, the current wealth of literature only demonstrates individual use of these techniques in the study of aggregation and designing formulations. There are opportunities to fill this void by combining computational and NMR methods. Molecular Dynamics and NMR share a complementary relationship, whereby observable NMR parameters can be used to validate those derived by molecular simulations. This has been made possible by the introduction of methodologies such as CamShift [159,160], which can be used to evaluate molecular dynamics derived backbone chemical shifts of proteins. The significance of this going forward may help the characterisation of protein motions on an atomic level, which ultimately can be invaluable when studying aggregation and informing the design of formulations.

Molecular docking has proven to be effective in finding excipient-binding hotspots on a protein's surface. During drug discovery, 2D ¹H-¹⁵N HSQC chemical shift perturbations (CSPs) have been used to guide molecular docking by selection of binding sites [161]. This also provides an excellent way of refinement and/or validation of molecular docking of excipients where low affinity binding and possibly non-specific interactions may occur. Once binding site(s) and results from docking are experimentally validated for a range of excipients by NMR, molecular docking would then have the potential to become more predictive, and therefore adopted in formulation design. Excipients could be screened in silico against validated protein binding-sites for local structure stabilisation, or for the shielding of protein surfaces that otherwise self-interact. Potentially even the design of novel excipients may be undertaken. Eventually, as accuracy of docking improves, this could be converted into an in silico high-throughput process which provides detailed insights into the nature of excipient–protein interactions.

Such computer aided formulation design can be complemented with molecular dynamics simulations by taking into consideration the simultaneous presence of multiple components, by including not only water, salts and buffers, but also including a combination of excipients. This would then enable a more detailed understanding of the complex interactions found between

the effects of excipients, for example through competing affinity for either the same binding site or multiple binding sites. Ligand observe NMR methods may be used to confirm these in silico results. Judicious analyses of trajectories may also indicate a possible rationale for the effects of excipient combinations on protein structure and/or stability.

The computer modelling approaches may have higher significance during the development stages, particularly in industry where the availability of protein (native or isotope enriched) is low, and not yet fully characterized by NMR. As the protein production and purification processes are established, the issue of the availability for both NMR and experimental formulation, becomes less important. At those stages, an extensive set of experimental information may be acquired to inform the rational design of formulations, but will be able to build upon a wide pre-screen of available excipients in silico.

7. Conclusions

This review has highlighted the principles that govern molecular modelling tools and NMR spectroscopy. It also has given an overview of the applications of NMR and molecular simulations to provide atomic level insights into possible interactions between therapeutic proteins and excipients. A key aim going forward will be to harness the relationship between NMR and computational modelling, and the insights they provide, in the design of new formulations that will ensure product stability.

Author Contributions: Conceived and designed the manuscript: A.P., M.J.H., M.Z. and P.A.D.

Funding: This research was funded by the Engineering and Physical Sciences Research Council (EPSRC) Future Targeted Healthcare Manufacturing Hub (EP/P006485/1), the EPSRC Centre for Doctoral Training (CDT) in Emergent Macromolecular Therapies (EP/L015218/), and EPSRC Enabling rapid formulation grant (EP/N025105/1).

Conflicts of Interest: The authors declare no conflict of interest. NanoPuzzle Medicines Design will derive no financial gain from this work. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

References

1. Wang, W.; Singh, S.; Zeng, D.L.; King, K.; Nema, S. Antibody structure, instability, and formulation. *J. Pharm. Sci.* **2007**, *96*, 1–26. [[CrossRef](#)] [[PubMed](#)]
2. Garidel, P.; Kuhn, A.B.; Schäfer, L.V.; Karow-Zwick, A.R.; Blech, M. High-concentration protein formulations: How high is high? *Eur. J. Pharm. Biopharm.* **2017**, *119*, 353–360. [[CrossRef](#)] [[PubMed](#)]
3. Roberts, C.J. Protein aggregation and its impact on product quality. *Curr. Opin. Biotechnol.* **2014**, *30*, 211–217. [[CrossRef](#)] [[PubMed](#)]
4. Smart, A.L.; Gaisford, S.; Basit, A.W. Oral peptide and protein delivery: Intestinal obstacles and commercial prospects. *Expert Opin. Drug Deliv.* **2014**, *11*, 1323–1335. [[CrossRef](#)] [[PubMed](#)]
5. Li, X.; Yu, M.; Fan, W.; Gan, Y.; Hovgaard, L.; Yang, M. Orally active-targeted drug delivery systems for proteins and peptides. *Expert Opin. Drug Deliv.* **2014**, *11*, 1435–1447. [[CrossRef](#)] [[PubMed](#)]
6. Bye, J.W.; Platts, L.; Falconer, R.J. Biopharmaceutical liquid formulation: A review of the science of protein stability and solubility in aqueous environments. *Biotechnol. Lett.* **2014**, *36*, 869–875. [[CrossRef](#)] [[PubMed](#)]
7. Wang, W. Advanced protein formulations. *Protein Sci.* **2015**, *24*, 1031–1039. [[CrossRef](#)] [[PubMed](#)]
8. Cromwell, M.E.M.; Hilario, E.; Jacobson, F. Protein aggregation and bioprocessing. *AAPS J.* **2006**, *8*, E572–E579. [[CrossRef](#)] [[PubMed](#)]
9. Vázquez-Rey, M.; Lang, D.A. Aggregates in monoclonal antibody manufacturing processes. *Biotechnol. Bioeng.* **2011**, *108*, 1494–1508. [[CrossRef](#)] [[PubMed](#)]
10. Philo, J.S.; Arakawa, T. Mechanisms of protein aggregation. *Curr. Pharm. Biotechnol.* **2009**, *10*, 348–351. [[CrossRef](#)] [[PubMed](#)]
11. Wang, W.; Nema, S.; Teagarden, D. Protein aggregation-Pathways and influencing factors. *Int. J. Pharm.* **2010**, *390*, 89–99. [[CrossRef](#)] [[PubMed](#)]
12. Menzen, T.; Friess, W. Temperature-ramped studies on the aggregation, unfolding, and interaction of a therapeutic monoclonal antibody. *J. Pharm. Sci.* **2014**, *103*, 445–455. [[CrossRef](#)] [[PubMed](#)]

13. Maddux, N.R.; Iyer, V.; Cheng, W.; Youssef, A.M.K.; Joshi, S.B.; Volkin, D.B.; Ralston, J.P.; Winter, G.; Russell Middaugh, C. High throughput prediction of the long-term stability of pharmaceutical macromolecules from short-term multi-instrument spectroscopic data. *J. Pharm. Sci.* **2014**, *103*, 828–839. [[CrossRef](#)] [[PubMed](#)]
14. Brader, M.L.; Estey, T.; Bai, S.; Alston, R.W.; Lucas, K.K.; Lantz, S.; Landsman, P.; Maloney, K.M. Examination of thermal unfolding and aggregation profiles of a series of developable therapeutic monoclonal antibodies. *Mol. Pharm.* **2015**, *12*, 1005–1017. [[CrossRef](#)] [[PubMed](#)]
15. Chakroun, N.; Hilton, D.; Ahmad, S.S.; Platt, G.W.; Dalby, P.A. Mapping the Aggregation Kinetics of a Therapeutic Antibody Fragment. *Mol. Pharm.* **2016**, *13*, 307–319. [[CrossRef](#)] [[PubMed](#)]
16. Robinson, M.J.; Matejtschuk, P.; Bristow, A.F.; Dalby, P.A. Tm-Values and Unfolded Fraction Can Predict Aggregation Rates for Granulocyte Colony Stimulating Factor Variant Formulations but Not under Predominantly Native Conditions. *Mol. Pharm.* **2018**, *15*, 256–267. [[CrossRef](#)] [[PubMed](#)]
17. Zhang, C.; Samad, M.; Yu, H.; Chakroun, N.; Hilton, D.; Dalby, P.A. Computational-design to reduce conformational flexibility and aggregation rates of an antibody Fab fragment. *Mol. Pharm.* **2018**, *15*, 3079–3092. [[CrossRef](#)] [[PubMed](#)]
18. Kamerzell, T.J.; Esfandiary, R.; Joshi, S.B.; Middaugh, C.R.; Volkin, D.B. Protein-excipient interactions: Mechanisms and biophysical characterization applied to protein formulation development. *Adv. Drug Deliv. Rev.* **2011**, *63*, 1118–1159. [[CrossRef](#)] [[PubMed](#)]
19. Timasheff, S.N. Protein-solvent preferential interactions, protein hydration, and the modulation of biochemical reactions by solvent components. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 9721–9726. [[CrossRef](#)] [[PubMed](#)]
20. Shukla, D.; Schneider, C.P.; Trout, B.L. Molecular level insight into intra-solvent interaction effects on protein stability and aggregation. *Adv. Drug Deliv. Rev.* **2011**, *63*, 1074–1085. [[CrossRef](#)] [[PubMed](#)]
21. Fox, S.; Farr-Jones, S.; Sopchak, L.; Boggs, A.; Nicely, H.W.; Khoury, R.; Biros, M. High-throughput screening: Update on practices and success. *J. Biomol. Screen.* **2006**, *11*, 864–869. [[CrossRef](#)] [[PubMed](#)]
22. Macarron, R.; Banks, M.N.; Bojanic, D.; Burns, D.J.; Cirovic, D.A.; Garyantes, T.; Green, D.V.S.; Hertzberg, R.P.; Janzen, W.P.; Paslay, J.W.; et al. Impact of high-throughput screening. *Nature* **2011**, *10*, 188–195. [[CrossRef](#)]
23. Mayr, L.M.; Bojanic, D. Novel trends in high-throughput screening. *Curr. Opin. Pharmacol.* **2009**, *9*, 580–588. [[CrossRef](#)] [[PubMed](#)]
24. Bhambhani, A.; Kissmann, J.M.; Joshi, S.B.; Volkin, D.B.; Kashi, R.S.; Middaugh, C.R. Formulation design and high-throughput excipient selection based on structural integrity and conformational stability of dilute and highly concentrated IgG1 monoclonal antibody solutions. *J. Pharm. Sci.* **2012**, *101*, 1120–1135. [[CrossRef](#)] [[PubMed](#)]
25. Wen, Z.Q. Raman spectroscopy of protein pharmaceuticals. *J. Pharm. Sci.* **2007**, *96*, 2861–2878. [[CrossRef](#)] [[PubMed](#)]
26. Woody, R.W. On the analysis of membrane protein circular dichroism spectra. *Protein Sci.* **2004**, *13*, 100–112. [[CrossRef](#)]
27. Joshi, S.B.; Kamerzell, T.J.; McNown, C.; Middaugh, C.R. The interaction of heparin/polyanions with bovine, porcine, and human growth hormone. *J. Pharm. Sci.* **2008**, *97*, 1368–1385. [[CrossRef](#)] [[PubMed](#)]
28. He, F.; Hogan, S.; Latypov, R.F.; Narhi, L.O.; Razinkov, V.I. High throughput thermostability screening of monoclonal antibody formulations. *J. Pharm. Sci.* **2010**, *99*, 1707–1720. [[CrossRef](#)] [[PubMed](#)]
29. Kamerzell, T.J.; Ramsey, J.D.; Middaugh, C.R. Immunoglobulin dynamics, conformational fluctuations, and nonlinear elasticity and their effects on stability. *J. Phys. Chem. B* **2008**, *112*, 3240–3250. [[CrossRef](#)] [[PubMed](#)]
30. Harn, N.; Allan, C.; Oliver, C.; Middaugh, C.R. Highly concentrated monoclonal antibody solutions: Direct analysis of physical structure and thermal stability. *J. Pharm. Sci.* **2007**, *96*, 532–546. [[CrossRef](#)] [[PubMed](#)]
31. Goldberg, D.S.; Bishop, S.M.; Shah, A.U.; Sathish, H.A. Formulation development of therapeutic monoclonal antibodies using high-throughput fluorescence and static light scattering techniques: Role of conformational and colloidal stability. *J. Pharm. Sci.* **2011**, *100*, 1306–1315. [[CrossRef](#)] [[PubMed](#)]
32. He, F.; Phan, D.H.; Hogan, S.; Bailey, R.; Becker, G.W.; Narhi, L.O.; Razinkov, V.I. Detection of IgG Aggregation by a High Throughput Method Based on Extrinsic Fluorescence. *J. Pharm. Sci.* **2010**, *99*, 2598–2608. [[CrossRef](#)] [[PubMed](#)]
33. Li, Y.; Mach, H.; Blue, J.T. High Throughput Formulation Screening for Global Aggregation Behaviors of Three Monoclonal Antibodies. *J. Pharm. Sci.* **2011**, *100*, 2120–2135. [[CrossRef](#)] [[PubMed](#)]

34. Espargaró, A.; Castillo, V.; de Groot, N.S.; Ventura, S. The in Vivo and in Vitro Aggregation Properties of Globular Proteins Correlate with Their Conformational Stability: The SH3 Case. *J. Mol. Biol.* **2008**, *378*, 1116–1131. [[CrossRef](#)] [[PubMed](#)]
35. Chi, E.; Krishnan, S.; Kendrick, B. Roles of conformational stability and colloidal stability in the aggregation of recombinant human granulocyte colony-stimulating factor. *Protein Sci.* **2003**, *12*, 903–913. [[CrossRef](#)] [[PubMed](#)]
36. Chiti, F.; Dobson, C.M. Amyloid formation by globular proteins under native conditions. *Nat. Chem. Biol.* **2009**, *5*, 15–22. [[CrossRef](#)] [[PubMed](#)]
37. Costanzo, J.A.; O'Brien, C.J.; Tiller, K.; Tamargo, E.; Robinson, A.S.; Roberts, C.J.; Fernandez, E.J. Conformational stability as a design target to control protein aggregation. *Protein Eng. Des. Sel.* **2014**, *27*, 157–167. [[CrossRef](#)] [[PubMed](#)]
38. Neudecker, P.; Robustelli, P.; Cavalli, A.; Walsh, P.; Lundström, P.; Zarrine-Afsar, A.; Sharpe, S.; Vendruscolo, M.; Kay, L.E. Structure of an Intermediate State in Protein Folding and Aggregation. *Science* **2012**, *336*, 362. [[CrossRef](#)] [[PubMed](#)]
39. Kontoyianni, M. Docking and Virtual Screening in Drug Discovery. In *Proteomics for Drug Discovery: Methods and Protocols*; Lazar, I.M., Kontoyianni, M., Lazar, A.C., Eds.; Springer: New York, NY, USA, 2017; pp. 255–266. ISBN 978-1-4939-7201-2.
40. Novikov, F.N.; Chilov, G.G. Molecular docking: Theoretical background, practical applications and perspectives. *Mendeleev Commun.* **2009**, *19*, 237–242. [[CrossRef](#)]
41. Huang, N.; Kalyanaraman, C.; Irwin, J.J.; Jacobson, M.P. Physics-based scoring of protein–Ligand complexes: Enrichment of known inhibitors in large-scale virtual screening. *J. Chem. Inf. Model.* **2006**, *46*, 243–253. [[CrossRef](#)] [[PubMed](#)]
42. Wang, R.; Lu, Y.; Wang, S. Comparative evaluation of 11 scoring functions for molecular docking. *J. Med. Chem.* **2003**, *46*, 2287–2303. [[CrossRef](#)] [[PubMed](#)]
43. Huang, S.-Y.; Grinter, S.Z.; Zou, X. Scoring functions and their evaluation methods for protein–ligand docking: Recent advances and future directions. *Phys. Chem. Chem. Phys.* **2010**, *12*, 12899. [[CrossRef](#)] [[PubMed](#)]
44. Sousa, S.F.; Ribeiro, A.J.M.; Coimbra, J.T.S.; Neves, R.P.P.; Martins, S.A.; Moorthy, N.S.H.N.; Fernandes, P.A.; Ramos, M.J. Protein-Ligand Docking in the New Millennium—A Retrospective of 10 Years in the Field. *Curr. Med. Chem.* **2013**, *20*, 2296–2314. [[CrossRef](#)] [[PubMed](#)]
45. Macindoe, G.; Mavridis, L.; Venkatraman, V.; Devignes, M.D.; Ritchie, D.W. HexServer: An FFT-based protein docking server powered by graphics processors. *Nucleic Acids Res.* **2010**, *38*, 445–449. [[CrossRef](#)] [[PubMed](#)]
46. Tovchigrechko, A.; Vakser, I.A. GRAMM-X public web server for protein-protein docking. *Nucleic Acids Res.* **2006**, *34*, 310–314. [[CrossRef](#)] [[PubMed](#)]
47. Lyskov, S.; Gray, J.J. The RosettaDock server for local protein-protein docking. *Nucleic Acids Res.* **2008**, *36*, 233–238. [[CrossRef](#)] [[PubMed](#)]
48. Morris, G.; Huey, R. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **2009**, *30*, 2785–2791. [[CrossRef](#)] [[PubMed](#)]
49. Jones, G.; Willett, P.; Glen, R.C. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *J. Mol. Biol.* **1995**, *245*, 43–53. [[CrossRef](#)]
50. Jones, G.; Willett, P.; Glen, R.C.; Leach, A.R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727–748. [[CrossRef](#)] [[PubMed](#)]
51. Yang, J.M.; Chen, C.C. GEMDOCK: A Generic Evolutionary Method for Molecular Docking. *Proteins Struct. Funct. Genet.* **2004**, *55*, 288–304. [[CrossRef](#)] [[PubMed](#)]
52. B-Rao, C.; Subramanian, J.; Sharma, S.D. Managing protein flexibility in docking and its applications. *Drug Discov. Today* **2009**, *14*, 394–400. [[CrossRef](#)] [[PubMed](#)]
53. Trott, O.; Olson, A. NIH Public Access. *J. Comput. Chem.* **2010**, *31*, 455–461. [[CrossRef](#)] [[PubMed](#)]
54. Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A Fast Flexible Docking Method using an Incremental Construction Algorithm. *J. Mol. Biol.* **1996**, *261*, 470–489. [[CrossRef](#)] [[PubMed](#)]
55. Sánchez-Linares, I.; Pérez-Sánchez, H.; Cecilia, J.M.; García, J.M. High-Throughput parallel blind Virtual Screening using BINDSURF. *BMC Bioinform.* **2012**, *13*. [[CrossRef](#)] [[PubMed](#)]

56. Karplus, M.; McCammon, J.A. Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* **2002**, *9*, 646–652. [[CrossRef](#)] [[PubMed](#)]
57. De Vivo, M.; Masetti, M.; Bottegoni, G.; Cavalli, A. Role of Molecular Dynamics and Related Methods in Drug Discovery. *J. Med. Chem.* **2016**, *59*, 4035–4061. [[CrossRef](#)] [[PubMed](#)]
58. González, M.A. Force fields and molecular dynamics simulations. *Collect. SFN* **2011**, *12*, 169–200. [[CrossRef](#)]
59. MacKerell, A.D.; Bashford, D.; Bellott, M.; Dunbrack, R.L.; Evanseck, J.D.; Field, M.J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; et al. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616. [[CrossRef](#)] [[PubMed](#)]
60. MacKerell, A.D.; Banavali, N.; Foloppe, N. Development and current status of the CHARMM force field for nucleic acids. *Biopolymers* **2000**, *56*, 257–265. [[CrossRef](#)]
61. Weiner, P.K.; Kollman, P.A. AMBER: Assisted model building with energy refinement. A general program for modeling molecules and their interactions. *J. Comput. Chem.* **1981**, *2*, 287–303. [[CrossRef](#)]
62. Scott, W.R.P.; Hünenberger, P.H.; Tironi, I.G.; Mark, A.E.; Billeter, S.R.; Fennen, J.; Torda, A.E.; Huber, T.; Krüger, P.; Van Gunsteren, W.F. The GROMOS biomolecular simulation program package. *J. Phys. Chem. A* **1999**, *103*, 3596–3607. [[CrossRef](#)]
63. Lindahl, E.; Hess, B. GROMACS 3.0: A package for molecular simulation and trajectory analysis. *Mol. Model. Annu.* **2001**, 306–317. [[CrossRef](#)]
64. Case, D.A.; Cheatham, T.E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K.M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R.J. The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005**, *26*, 1668–1688. [[CrossRef](#)] [[PubMed](#)]
65. Phillips, J.C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R.D.; Kalé, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781–1802. [[CrossRef](#)] [[PubMed](#)]
66. Brooks, B.; Brooks, C.; MacKerell, A.; Nilsson, L.; Petrella, R.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; et al. CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **2009**, *30*, 1545–1614. [[CrossRef](#)] [[PubMed](#)]
67. Klepeis, J.L.; Lindorff-Larsen, K.; Dror, R.O.; Shaw, D.E. Long-timescale molecular dynamics simulations of protein structure and function. *Curr. Opin. Struct. Biol.* **2009**, *19*, 120–127. [[CrossRef](#)] [[PubMed](#)]
68. Lane, T.J.; Shukla, D.; Beauchamp, K.A.; Pande, V.S. To milliseconds and beyond: Challenges in the simulation of protein folding. *Curr. Opin. Struct. Biol.* **2013**, *23*, 58–65. [[CrossRef](#)] [[PubMed](#)]
69. Dror, R.O.; Jensen, M.Ø.; Borhani, D.W.; Shaw, D.E. Exploring atomic resolution physiology on a femtosecond to millisecond timescale using molecular dynamics simulations. *J. Gen. Physiol.* **2010**, *135*, 555–562. [[CrossRef](#)] [[PubMed](#)]
70. Wishart, D.S. Characterization of biopharmaceuticals by NMR spectroscopy. *TrAC Trends Anal. Chem.* **2013**, *48*, 96–111. [[CrossRef](#)]
71. Osawa, M.; Takeuchi, K.; Ueda, T.; Nishida, N.; Shimada, I. Functional dynamics of proteins revealed by solution NMR. *Curr. Opin. Struct. Biol.* **2012**, *22*, 660–669. [[CrossRef](#)] [[PubMed](#)]
72. Li, C.; Tang, C.; Liu, M. Protein dynamics elucidated by NMR technique. *Protein Cell* **2013**, *4*, 726–730. [[CrossRef](#)] [[PubMed](#)]
73. Saitô, H.; Ando, I.; Ramamoorthy, A. Chemical shift tensor—The heart of NMR: Insights into biological aspects of proteins. *Prog. Nucl. Magn. Reson. Spectrosc.* **2010**, *57*, 181–228. [[CrossRef](#)] [[PubMed](#)]
74. Wishart, D.S. Interpreting protein chemical shift data. *Prog. Nucl. Magn. Reson. Spectrosc.* **2011**, *58*, 62–87. [[CrossRef](#)] [[PubMed](#)]
75. Neudecker, P.; Lundström, P.; Kay, L.E. Relaxation dispersion NMR spectroscopy as a tool for detailed studies of protein folding. *Biophys. J.* **2009**, *96*, 2045–2054. [[CrossRef](#)] [[PubMed](#)]
76. Farber, P.J.; Mittermaier, A. Relaxation dispersion NMR spectroscopy for the study of protein allostery. *Biophys. Rev.* **2015**, *7*, 191–200. [[CrossRef](#)] [[PubMed](#)]
77. Ban, D.; Smith, C.A.; de Groot, B.L.; Griesinger, C.; Lee, D. Recent advances in measuring the kinetics of biomolecules by NMR relaxation dispersion spectroscopy. *Arch. Biochem. Biophys.* **2017**, *628*, 81–91. [[CrossRef](#)] [[PubMed](#)]
78. Sapienza, P.; Lee, A. Using NMR to study fast dynamics in proteins: Methods and applications. *Curr. Opin. Pharmacol.* **2010**, *10*, 723–730. [[CrossRef](#)] [[PubMed](#)]

79. Charlier, C.; Cousin, S.F.; Ferrage, F. Protein dynamics from nuclear magnetic relaxation. *Chem. Soc. Rev.* **2016**, *45*, 2410–2422. [[CrossRef](#)] [[PubMed](#)]
80. Gossert, A.D.; Jahnke, W. NMR in drug discovery: A practical guide to identification and validation of ligands interacting with biological macromolecules. *Prog. Nucl. Magn. Reson. Spectrosc.* **2016**, *97*, 82–125. [[CrossRef](#)] [[PubMed](#)]
81. Becker, W.; Bhattachipolu, K.C.; Gubensäk, N.; Zangger, K. Investigating protein-ligand interactions by solution NMR spectroscopy. *ChemPhysChem* **2018**, 895–906. [[CrossRef](#)] [[PubMed](#)]
82. Lee, Y.J.; Lee, D.H.; Jeong, K.J. Enhanced production of human full-length immunoglobulin G1 in the periplasm of Escherichia coli. *Appl. Microbiol. Biotechnol.* **2014**, *98*, 1237–1246. [[CrossRef](#)] [[PubMed](#)]
83. McIntosh, L.P.; Dahlquist, F.W. Biosynthetic Incorporation of ¹⁵N and ¹³C for Assignment and Interpretation of Nuclear Magnetic Resonance Spectra of Proteins. *Q. Rev. Biophys.* **1990**, *23*, 1–38. [[CrossRef](#)] [[PubMed](#)]
84. Williamson, M.P. Using chemical shift perturbation to characterise ligand binding. *Prog. Nucl. Magn. Reson. Spectrosc.* **2013**, *73*, 1–16. [[CrossRef](#)] [[PubMed](#)]
85. Lemaster, D.M. Uniform and selective deuteration in two-dimensional NMR of proteins. *Annu. Rev. Biophys. Chem.* **1990**, *19*, 243–266. [[CrossRef](#)] [[PubMed](#)]
86. Rosenzweig, R.; Kay, L.E. Bringing Dynamic Molecular Machines into Focus by Methyl-TROSY NMR. *Annu. Rev. Biochem.* **2014**, *83*, 291–315. [[CrossRef](#)] [[PubMed](#)]
87. Tugarinov, V.; Kay, L.E. Methyl groups as probes of structure and dynamics in NMR studies of high-molecular-weight proteins. *ChemBioChem* **2005**, *6*, 1567–1577. [[CrossRef](#)] [[PubMed](#)]
88. Tugarinov, V.; Kanelis, V.; Kay, L.E. Isotope labeling strategies for the study of high-molecular-weight proteins by solution NMR spectroscopy. *Nat. Protoc.* **2006**, *1*, 749–754. [[CrossRef](#)] [[PubMed](#)]
89. Ayala, I.; Sounier, R.; Usé, N.; Gans, P.; Boisbouvier, J. An efficient protocol for the complete incorporation of methyl-protonated alanine in perdeuterated protein. *J. Biomol. NMR* **2009**, *43*, 111–119. [[CrossRef](#)] [[PubMed](#)]
90. Ruschak, A.M.; Velyvis, A.; Kay, L.E. A simple strategy for ¹³C,¹H labeling at the Ile- γ_2 methyl position in highly deuterated proteins. *J. Biomol. NMR* **2010**, *48*, 129–135. [[CrossRef](#)] [[PubMed](#)]
91. Ayala, I.; Hamelin, O.; Amero, C.; Pessey, O.; Plevin, M.J. An optimized isotopic labelling strategy of isoleucine- γ_2 methyl groups for solution NMR studies of high molecular weight proteins. *Chem. Commun.* **2012**, 1434–1436. [[CrossRef](#)] [[PubMed](#)]
92. Lichtenecker, R.J.; Coudeville, N.; Konrat, R.; Schmid, W. Selective Isotope Labelling of Leucine Residues by Using α -Ketoacid Precursor Compounds. *ChemBioChem* **2013**, *14*, 818–821. [[CrossRef](#)] [[PubMed](#)]
93. Pervushin, K.; Riek, R.; Wider, G.; Wüthrich, K. Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 12366–12371. [[CrossRef](#)] [[PubMed](#)]
94. Pervushin, K.V.; Wider, G.; Riek, R.; Wüthrich, K. The 3D NOESY-[(1)H,(15)N,(1)H]-ZQ-TROSY NMR experiment with diagonal peak suppression. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 9607–9612. [[CrossRef](#)] [[PubMed](#)]
95. Fernández, C.; Wider, G. TROSY in NMR studies of the structure and function of large biological macromolecules. *Curr. Opin. Struct. Biol.* **2003**, *13*, 570–580. [[CrossRef](#)] [[PubMed](#)]
96. Arbogast, L.W.; Brinson, R.G.; Marino, J.P. Mapping Monoclonal Antibody Structure by 2D ¹³C NMR at Natural Abundance. *Anal. Chem.* **2015**, *87*, 3556–3561. [[CrossRef](#)] [[PubMed](#)]
97. Amezcua, C.A.; Szabo, C.M. Assessment of higher order structure comparability in therapeutic proteins using nuclear magnetic resonance spectroscopy. *J. Pharm. Sci.* **2013**, *102*, 1724–1733. [[CrossRef](#)] [[PubMed](#)]
98. Fávero-Retto, M.P.; Palmieri, L.C.; Souza, T.A.C.B.; Almeida, F.C.L.; Lima, L.M.T.R. Structural meta-analysis of regular human insulin in pharmaceutical formulations. *Eur. J. Pharm. Biopharm.* **2013**, *85*, 1112–1121. [[CrossRef](#)] [[PubMed](#)]
99. Quinternet, M.; Starck, J.P.; Delsuc, M.A.; Kieffer, B. Heteronuclear NMR provides an accurate assessment of therapeutic insulin's quality. *J. Pharm. Biomed. Anal.* **2013**, *78–79*, 252–254. [[CrossRef](#)] [[PubMed](#)]
100. Pellecchia, M.; Sem, D.S.; Wüthrich, K. NMR in Drug Discovery. *Nat. Rev. Drug Discov.* **2002**, *1*, 211–219. [[CrossRef](#)] [[PubMed](#)]
101. Wagstaff, J.L.; Taylor, S.L.; Howard, M.J. Recent developments and applications of saturation transfer difference nuclear magnetic resonance (STD NMR) spectroscopy. *Mol. BioSyst.* **2013**, *9*, 571–577. [[CrossRef](#)] [[PubMed](#)]

102. Dalvit, C.; Pevarello, P.; Tato, M.; Veronesi, M.; Vulpetti, A.; Sundström, M. Identification of compounds with binding affinity to proteins via magnetization transfer from bulk water. *J. Biomol. NMR* **2000**, *18*, 65–68. [[CrossRef](#)] [[PubMed](#)]
103. Dalvit, C.; Fogliatto, G.P.; Stewart, A.; Veronesi, M.; Stockman, B. WaterLOGSY as a method for primary NMR screening: Practical aspects and range of applicability. *J. Biomol. NMR* **2001**, *21*, 349–359. [[CrossRef](#)] [[PubMed](#)]
104. Pagès, G.; Gilard, V.; Martino, R.; Malet-Martino, M. Pulsed-field gradient nuclear magnetic resonance measurements (PFG NMR) for diffusion ordered spectroscopy (DOSY) mapping. *Analyst* **2017**, *142*, 3771–3796. [[CrossRef](#)] [[PubMed](#)]
105. Bernadó, P.; Åkerud, T.; García de la Torre, J.; Akke, M.; Pons, M. Combined use of NMR relaxation measurements and hydrodynamic calculations to study protein association. Evidence for tetramers of low molecular weight protein tyrosine phosphatase in solution. *J. Am. Chem. Soc.* **2003**, *125*, 916–923. [[CrossRef](#)] [[PubMed](#)]
106. Wu, K.-P.; Baum, J. Detection of transient interchain interactions in the intrinsically disordered protein alpha-synuclein by NMR paramagnetic relaxation enhancement. *J. Am. Chem. Soc.* **2010**, *132*, 5546–5547. [[CrossRef](#)] [[PubMed](#)]
107. Dedmon, M.M.; Lindorff-Larsen, K.; Christodoulou, J.; Vendruscolo, M.; Dobson, C.M. Mapping long-range interactions in α -synuclein using spin-label NMR and ensemble molecular dynamics simulations. *J. Am. Chem. Soc.* **2005**, *127*, 476–477. [[CrossRef](#)] [[PubMed](#)]
108. Fawzi, N.L.; Ying, J.; Ghirlando, R.; Torchia, D.A.; Clore, G.M. Atomic-resolution dynamics on the surface of amyloid- β protofibrils probed by solution NMR. *Nature* **2011**, *480*, 268–272. [[CrossRef](#)] [[PubMed](#)]
109. Carr, H.Y.; Purcell, E.M. Effects of diffusion on free precession in nuclear magnetic resonance experiments. *Phys. Rev.* **1954**, *94*, 630–638. [[CrossRef](#)]
110. Meiboom, S.; Gill, D. Modified spin-echo method for measuring nuclear relaxation times. *Rev. Sci. Instrum.* **1958**, *29*, 688–691. [[CrossRef](#)]
111. Mukherjee, S.; Pondaven, S.P.; Jaroniec, C.P. Conformational flexibility of a human immunoglobulin light chain variable domain by relaxation dispersion nuclear magnetic resonance spectroscopy: Implications for protein misfolding and amyloid assembly. *Biochemistry* **2011**, *50*, 5845–5857. [[CrossRef](#)] [[PubMed](#)]
112. Feng, Y.; Taraban, M.B.; Yu, Y.B. Water proton NMR—A sensitive probe for solute association. *Chem. Commun.* **2015**, *51*, 6804–6807. [[CrossRef](#)] [[PubMed](#)]
113. Taraban, M.B.; Truong, H.C.; Feng, Y.; Jouravleva, E.V.; Anisimov, M.A.; Yu, Y.B. Water Proton NMR for in Situ Detection of Insulin Aggregates. *J. Pharm. Sci.* **2015**, *104*, 4132–4141. [[CrossRef](#)] [[PubMed](#)]
114. Taraban, M.B.; Depaz, R.A.; Lobo, B.; Yu, Y.B. Water Proton NMR: A Tool for Protein Aggregation Characterization. *Anal. Chem.* **2017**, *89*, 5494–5502. [[CrossRef](#)] [[PubMed](#)]
115. Patil, S.M.; Keire, D.A.; Chen, K. Comparison of NMR and Dynamic Light Scattering for Measuring Diffusion Coefficients of Formulated Insulin: Implications for Particle Size Distribution Measurements in Drug Products. *AAPS J.* **2017**, *19*, 1760–1766. [[CrossRef](#)] [[PubMed](#)]
116. Rossini, A.J.; Widdifield, C.M.; Zagdoun, A.; Lelli, M.; Schwarzwälder, M.; Copéret, C.; Lesage, A.; Emsley, L. Dynamic nuclear polarization enhanced NMR spectroscopy for pharmaceutical formulations. *J. Am. Chem. Soc.* **2014**, *136*, 2324–2334. [[CrossRef](#)] [[PubMed](#)]
117. Pisklak, D.M.; Zielińska-Pisklak, M.; Szeleszczuk, Ł.; Wawer, I. ^{13}C cross-polarization magic-angle spinning nuclear magnetic resonance analysis of the solid drug forms with low concentration of an active ingredient-propranolol case. *J. Pharm. Biomed. Anal.* **2014**, *93*, 68–72. [[CrossRef](#)] [[PubMed](#)]
118. Maciej, D.; Zieli, M.A.; Szeleszczuk, Ł. ^{13}C solid-state NMR analysis of the most common pharmaceutical excipients used in solid drug formulations, Part I: Chemical shifts assignment. *J. Pharm. Biomed. Anal.* **2016**, *122*, 81–89. [[CrossRef](#)]
119. Pisklak, D.M.; Zielińska-Pisklak, M.; Szeleszczuk, Ł.; Wawer, I. ^{13}C solid-state NMR analysis of the most common pharmaceutical excipients used in solid drug formulations Part II: CP kinetics and relaxation analysis. *J. Pharm. Biomed. Anal.* **2016**, *122*, 29–34. [[CrossRef](#)] [[PubMed](#)]
120. Panjwani, N.; Hodgson, D.J.; Sauvé, S.; Aubin, Y. Assessment of the effects of pH, formulation and deformulation on the conformation of interferon alpha-2 by NMR. *J. Pharm. Sci.* **2010**, *99*, 3334–3342. [[CrossRef](#)] [[PubMed](#)]

121. Aubin, Y.; Gingras, G.; Sauvé, S. Assessment of the three-dimensional structure of recombinant protein therapeutics by NMR fingerprinting: Demonstration on recombinant human granulocyte macrophage-colony stimulation factor. *Anal. Chem.* **2008**, *80*, 2623–2627. [[CrossRef](#)] [[PubMed](#)]
122. Singh, S.M.; Bandi, S.; Jones, D.N.M.; Mallela, K.M.G. Effect of Polysorbate 20 and Polysorbate 80 on the Higher-Order Structure of a Monoclonal Antibody and Its Fab and Fc Fragments Probed Using 2D Nuclear Magnetic Resonance Spectroscopy. *J. Pharm. Sci.* **2017**, *106*, 3486–3498. [[CrossRef](#)] [[PubMed](#)]
123. Kheddo, P.; Cliff, M.J.; Uddin, S.; Van Der Walle, C.F.; Golovanov, A.P.; Kheddo, P.; Cliff, M.J.; Uddin, S.; Van Der, C.F.; Golovanov, A.P.; et al. Characterizing monoclonal antibody formulations in arginine glutamate solutions using ¹H NMR spectroscopy. *Mabs* **2016**, *8*, 1245–1258. [[CrossRef](#)] [[PubMed](#)]
124. Poppe, L.; Jordan, J.B.; Lawson, K.; Jerums, M.; Apostol, I.; Schnier, P.D. Profiling Formulated Monoclonal Antibodies by ¹H NMR Spectroscopy. *Anal. Chem.* **2013**, *85*, 9623–9629. [[CrossRef](#)] [[PubMed](#)]
125. Poppe, L.; Jordan, J.B.; Rogers, G.; Schnier, P.D. On the analytical superiority of ¹D NMR for fingerprinting the higher order structure of protein therapeutics compared to multidimensional NMR methods. *Anal. Chem.* **2015**, *87*, 5539–5545. [[CrossRef](#)] [[PubMed](#)]
126. Bednarek, E.; Sitkowski, J.; Bocian, W.; Borowicz, P.; Płucienniczak, G.; Stadnik, D.; Surmacz-Chwedoruk, W.; Jaworska, B.; Kozerski, L. Structure and pharmaceutical formulation development of a new long-acting recombinant human insulin analog studied by NMR and MS. *J. Pharm. Biomed. Anal.* **2017**, *135*, 126–132. [[CrossRef](#)] [[PubMed](#)]
127. Morin, A.; Meiler, J.; Mizoue, L.S. Computational design of protein-ligand interfaces: Potential in therapeutic development. *Trends Biotechnol.* **2011**, *29*, 159–166. [[CrossRef](#)] [[PubMed](#)]
128. Liang, S.; Li, L.; Hsu, W.-L.; Pilcher, M.N.; Uversky, V.; Zhou, Y.; Dunker, A.K.; Meroueh, S.O. Exploring the Molecular Design of Protein Interaction Sites with Molecular Dynamics Simulations and Free Energy Calculations. *Biochemistry* **2009**, *48*, 399–414. [[CrossRef](#)] [[PubMed](#)]
129. Childers, M.C.; Daggett, V. Insights from molecular dynamics simulations for computational protein design. *Mol. Syst. Des. Eng.* **2017**. [[CrossRef](#)] [[PubMed](#)]
130. Daggett, V.; Levitt, M. A model of the molten globule state from molecular dynamics simulations. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 5142–5146. [[CrossRef](#)] [[PubMed](#)]
131. Alonso, D.O.V.; DeArmond, S.J.; Cohen, F.E.; Daggett, V. Mapping the early steps in the pH-induced conformational conversion of the prion protein. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 2985–2989. [[CrossRef](#)] [[PubMed](#)]
132. Alonso, D.O.V.; An, C.; Daggett, V. Simulations of biomolecules: Characterization of the early steps in the pH-induced conformational conversion of the hamster, bovine and human forms of the prion protein. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2002**, *360*, 1165–1178. [[CrossRef](#)] [[PubMed](#)]
133. Daggett, V. α -sheet: The toxic conformer in amyloid diseases? *Acc. Chem. Res.* **2006**, *39*, 594–602. [[CrossRef](#)] [[PubMed](#)]
134. Kellock, J.; Hopping, G.; Caughey, B.; Daggett, V. Peptides Composed of Alternating L- and D-Amino Acids Inhibit Amyloidogenesis in Three Distinct Amyloid Systems Independent of Sequence. *J. Mol. Biol.* **2016**, *428*, 2317–2328. [[CrossRef](#)] [[PubMed](#)]
135. Agrawal, N.J.; Kumar, S.; Wang, X.; Helk, B.; Singh, S.K.; Trout, B.L. Aggregation in Protein-Based Biotherapeutics: Computational Studies and Tools to Identify Aggregation-Prone Regions. *J. Pharm. Sci.* **2011**, *100*, 5081–5095. [[CrossRef](#)] [[PubMed](#)]
136. Conchillo-Solé, O.; de Groot, N.S.; Avilés, F.X.; Vendrell, J.; Daura, X.; Ventura, S. AGGRESCAN: A server for the prediction and evaluation of “hot spots” of aggregation in polypeptides. *BMC Bioinform.* **2007**, *8*. [[CrossRef](#)] [[PubMed](#)]
137. Fernandez-Escamilla, A.M.; Rousseau, F.; Schymkowitz, J.; Serrano, L. Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat. Biotechnol.* **2004**, *22*, 1302–1306. [[CrossRef](#)] [[PubMed](#)]
138. Trovato, A.; Seno, F.; Tosatto, S.C.E. The PASTA server for protein aggregation prediction. *Protein Eng. Des. Sel.* **2007**, *20*, 521–523. [[CrossRef](#)] [[PubMed](#)]
139. Chennamsetty, N.; Voynov, V.; Kayser, V.; Helk, B.; Trout, B. Design of therapeutic proteins with enhanced stability. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 11937–11942. [[CrossRef](#)] [[PubMed](#)]
140. Chennamsetty, N.; Voynov, V.; Kayser, V.; Helk, B.; Trout, B.L. Prediction of aggregation prone regions of therapeutic proteins. *J. Phys. Chem. B* **2010**, *114*, 6614–6624. [[CrossRef](#)] [[PubMed](#)]

141. Chennamsetty, N.; Helk, B.; Voynov, V.; Kayser, V.; Trout, B. Aggregation-Prone Motifs in Human Immunoglobulin G. *J. Mol. Biol.* **2009**, *391*, 404–413. [[CrossRef](#)] [[PubMed](#)]
142. Li, C.; Wang, J.X.; Le, Y.; Chen, J.F. Studies of bicalutamide-excipients interaction by combination of molecular docking and molecular dynamics simulation. *Mol. Pharm.* **2013**, *10*, 2362–2369. [[CrossRef](#)] [[PubMed](#)]
143. Jha, P.K.; Larson, R.G. Assessing the efficiency of polymeric excipients by atomistic molecular dynamics simulations. *Mol. Pharm.* **2014**, *11*, 1676–1686. [[CrossRef](#)] [[PubMed](#)]
144. Shukla, D.; Trout, B. Interaction of Arginine with Proteins and the Mechanism by Which It Inhibits Aggregation. *J. Phys. Chem. B* **2010**, *114*, 13426–13438. [[CrossRef](#)] [[PubMed](#)]
145. Kheddo, P.; Tracka, M.; Armer, J.; Dearman, R.J.; Uddin, S.; Van Der Walle, C.F.; Golovanov, A.P. The effect of arginine glutamate on the stability of monoclonal antibodies in solution. *Int. J. Pharm.* **2014**, *473*, 126–133. [[CrossRef](#)] [[PubMed](#)]
146. Kheddo, P.; Golovanov, A.P.; Mellody, K.T.; Uddin, S.; Van Der Walle, C.F.; Dearman, R.J. The effects of arginine glutamate, a promising excipient for protein formulation, on cell viability: Comparisons with NaCl. *Toxicol. In Vitro* **2016**, *33*, 88–98. [[CrossRef](#)] [[PubMed](#)]
147. Shukla, D.; Trout, B. Understanding the Synergistic Effect of Arginine and Glutamic Acid Mixtures on Protein Solubility. *J. Phys. Chem. B* **2011**, *115*, 11831–11839. [[CrossRef](#)] [[PubMed](#)]
148. Shukla, D.; Zamolo, L.; Cavallotti, C.; Trout, B.L. Understanding the role of arginine as an eluent in affinity chromatography via molecular computations. *J. Phys. Chem. B* **2011**, *115*, 2645–2654. [[CrossRef](#)] [[PubMed](#)]
149. Kurinomaru, T.; Kuwada, K.; Tomita, S.; Kameda, T.; Shiraki, K. Noncovalent PEGylation through Protein-Polyelectrolyte Interaction: Kinetic Experiment and Molecular Dynamics Simulation. *J. Phys. Chem. B* **2017**, *121*, 6785–6791. [[CrossRef](#)] [[PubMed](#)]
150. Rawat, S.; Gupta, P.; Kumar, A.; Garg, P.; Suri, C.; Sahoo, D. Molecular Mechanism of Poly(vinyl alcohol) Mediated Prevention of Aggregation and Stabilization of Insulin in Nanoparticles. *Mol. Pharm.* **2015**, *12*, 1018–1030. [[CrossRef](#)] [[PubMed](#)]
151. Kale, S.S.; Akamanchi, K.G. Trehalose Monooleate: A Potential Antiaggregation Agent for Stabilization of Proteins. *Mol. Pharm.* **2016**, *13*, 4082–4093. [[CrossRef](#)] [[PubMed](#)]
152. Messina, M.S.; Ko, J.H.; Yang, Z.; Strouse, M.J.; Houk, K.N.; Maynard, H.D. Effect of trehalose polymer regioisomers on protein stabilization. *Polym. Chem.* **2017**, *8*, 4781–4788. [[CrossRef](#)]
153. Kale, S.; Akamanchi, K. Rational approach for design and evaluation of anti-aggregation agents for protein stabilization: A case study of trehalose phenylalaninate. *Int. J. Pharm.* **2017**, *524*, 215–225. [[CrossRef](#)] [[PubMed](#)]
154. Arsiccio, A.; Pisano, R. Stability of Proteins in Carbohydrates and Other Additives during Freezing: The Human Growth Hormone as a Case Study. *J. Phys. Chem. B* **2017**, *121*, 8652–8660. [[CrossRef](#)] [[PubMed](#)]
155. Arsiccio, A.; Pisano, R. Clarifying the role of cryo- and lyo-protectants in the biopreservation of proteins. *Phys. Chem. Chem. Phys.* **2018**, *20*, 8267–8277. [[CrossRef](#)] [[PubMed](#)]
156. Barata, T.; Zhang, C.; Dalby, P.; Brocchini, S.; Zloh, M. Identification of Protein–Excipient Interaction Hotspots Using Computational Approaches. *Int. J. Mol. Sci.* **2016**, *17*, 853. [[CrossRef](#)] [[PubMed](#)]
157. Westermaier, Y.; Veurink, M.; Riis-Johannessen, T.; Guinchard, S.; Gurny, R.; Scapozza, L. Identification of aggregation breakers for bevacizumab (Avastin[®]) self-association through similarity searching and interaction studies. *Eur. J. Pharm. Biopharm.* **2013**, *85*, 773–780. [[CrossRef](#)] [[PubMed](#)]
158. Veurink, M.; Westermaier, Y.; Gurny, R.; Scapozza, L. Breaking the aggregation of the monoclonal antibody bevacizumab (Avastin[®]) by dexamethasone phosphate: Insights from molecular modelling and asymmetrical flow field-flow fractionation. *Pharm. Res.* **2013**, *30*, 1176–1187. [[CrossRef](#)] [[PubMed](#)]
159. Robustelli, P.; Kohlhoff, K.; Cavalli, A.; Vendruscolo, M. Using NMR chemical shifts as structural restraints in molecular dynamics simulations of proteins. *Structure* **2010**, *18*, 923–933. [[CrossRef](#)] [[PubMed](#)]
160. Robustelli, P.; Stafford, K.A.; Palmer, A.G. Interpreting protein structural dynamics from NMR chemical shifts. *J. Am. Chem. Soc.* **2012**, *134*, 6365–6374. [[CrossRef](#)] [[PubMed](#)]
161. Stark, J.; Powers, R. Rapid protein-ligand costructures using chemical shift perturbations. *J. Am. Chem. Soc.* **2008**, *130*, 535–545. [[CrossRef](#)] [[PubMed](#)]

