Investigations of new methods for studying molecular dynamics with elastic neutron scattering

Dominik Zeller

Communauté
UNIVERSITÉ Grenoble Alpes

THÈSE

Pour obtenir le grade de

DOCTEUR DE LA

COMMUNAUTE UNIVERSITE GRENOBLE ALPES

Spécialité : Physique pour les Sciences du Vivant

Arrêté ministériel : 25 mai 2016

Présentée par

Dominik ZELLER

Thèse dirigée par **Judith PETERS,** Professeur, UGA, LIPhy
et codirigée par **Victoria GARCIA SAKAI,** Docteur, ISIS (STFC)

préparée au sein des **Laboratoires :**
Laboratoire Interdisciplinaire de Physique,
Institut Laue Langevin,
ISIS (STFC),
dans **l'École Doctorale de Physique**

Investigations de nouvelles méthodes pour étudier la dynamique moléculaire par diffusion neutronique élastique

Investigations of new methods for studying molecular dynamics with elastic neutron scattering

Thèse soutenue publiquement le **13 mars 2019**,
devant le jury composé de :

Madame Irina MIHALCESCU – Présidente du jury
Professeur, UGA, Laboratoire Interdisciplinaire de Physique, Examinatrice
Monsieur Gerald KNELLER
Professeur, Université d'Orléans, Rapporteur
Monsieur Martin MÜLLER
Professeur, Universität Kiel, Rapporteur
Monsieur Marc JOYEUX
Directeur de Recherche, CNRS, Laboratoire Interdisciplinaire de Physique, Examinateur

Final version.

Abstract français :

Dans cette thèse, la diffusion élastique incohérente de neutrons (EINS) est étudiée en détail par une analyse systématique des données de la protéine Alpha-Lactalbumine (A-L). En général, l'approximation gaussienne (AG) est le modèle privilégié pour extraire le déplacement carré moyen (MSD) des protéines des données EINS. Compte tenu des améliorations récentes apportées à l'analyse des données EINS pour aller au-delà de l'AG, avec une description plus complexe de la dynamique, il est important de vérifier si les nouveaux modèles fournissent des informations supplémentaires plus précises. Pour étudier systématiquement l'influence de quatre modèles choisis sur les MSD, l'échantillon a été mesuré sous forme de poudre à trois niveaux d'hydratation différents et sur trois spectromètres à rétrodiffusion de neutrons, afin de pouvoir étudier la dynamique dans une large gamme de temps et d'espace. De plus, A-L a été mesuré sous deux formes différentes, avec et sans calcium, pour vérifier si de légers changements dans la dynamique peuvent être observés. L'évaluation des données expérimentales a permis de conclure que l'AG donnait des résultats qualitativement similaires aux modèles incluant une hétérogénéité dans les mouvements, si l'intersection de l'intensité élastique avec l'axe de transfert du moment neutronique, EI(0), est traitée de la même manière pour tous les modèles. Néanmoins, l'inclusion de l'hétérogénéité fournit une meilleure description des données EINS, et permet d'inclure davantage de points de données. Dans la plupart des cas, une hétérogénéité comprenant deux types de mouvements distincts (bimodale) s'avère suffisante. Comme technique complémentaire, des simulations de dynamique moléculaire (DM) ont été effectuées sur les deux formes de poudre de A-L avec deux niveaux d'hydratation et deux valeurs de température. Les trajectoires ont été évaluées avec trois résolutions instrumentales différentes, correspondant aux jeux de données expérimentaux. Les MSD résultant de l'AG et de deux modèles supplémentaires ont été comparés aux MSD calculés à partir des trajectoires et aux résultats expérimentaux. L'évaluation indique un accord qualitatif entre les modèles. Les résultats expérimentaux sont du même ordre de grandeur mais ne concordent pas parfaitement dans la plupart des cas, ce qui indique une différence non négligeable entre les simulations MD et les expériences. En conclusion, l'importance de EI(0) sur les MSD est bien documentée et il est proposé d'entreprendre des expériences supplémentaires pour évaluer l'intensité élastique au transfert de moment zéro de manière plus détaillée.

Abstract english:

In this thesis, elastic incoherent neutron scattering (EINS) is investigated in detail by performing a systematic analysis of data from the protein Alpha-Lactalbumin (A-L). Almost exclusively, the Gaussian approximation (GA) is the preferred model used to extract the mean square displacement (MSD) in proteins from EINS data. With a number of recent improvements in the analysis of EINS data to go beyond the GA, describing a more complex dynamical picture, it is important and relevant to assess whether new models give additional and more precise information. For the systematic study of the influence of four chosen models on the extracted MSD, the A-L protein was measured as a powder at three different hydration levels on three neutron backscattering spectrometers, to be able to access a wide temporal and spatial range of dynamics. Furthermore, A-L was measured in two different forms, with and without calcium to check if small changes in dynamics can be observed. The evaluation of the models on the experimental data leads to the conclusion that the GA gives qualitatively similar results to the models that include dynamical heterogeneity, if the intercept of the elastic intensity with the neutron momentum transfer axis EI(0) is treated in the same manner for all models. However, the inclusion of heterogeneity provides a better description of EINS data and allows the inclusion of more data points. In most cases, a heterogeneous description comprising two distinct kind of motions (bi-modal) appears to be sufficient. To complement the experimental results, data from molecular dynamics (MD) simulations were analysed on both powder forms of A-L with two hydration levels and two temperature values. The trajectories were evaluated with three different instrumental resolutions corresponding to the experimental data sets. The resulting MSDs of the GA and two additional models were compared to the direct MSD obtained from the trajectories and the experimental results. The evaluation indicates a qualitative agreement between the models. The experimental results are of the same order of magnitude but not in good agreement in most cases, indicating a non-negligible difference between MD simulations and experiments. In conclusion, the importance of EI(0) on the MSD is well documented and further experiments to gain access to the elastic intensity at zero momentum transfer are suggested.

Acknowledgements

First of all, I would like to thank my supervisor Prof. Dr. Judith Peters, for her support, encouragement and guidance without which this work would not have been possible. I also want to thank her for the possibility to participate in the ICNS 2017 in Daejeon, South-Korea where I had the opportunity to present my research. I would like to equally thank my co-supervisor Dr. Victoria Garcia Sakai for her support and the organisation of my stays in the UK.

I am grateful to the members of the jury, Prof. Dr. Gerald Kneller and Prof. Dr. Martin Müller, who accepted to be the referees of this work, and Prof. Dr. Irina Mihalcescu and Dr. Marc Joyeux for being part of the jury.

I am thankful for the Ph.D. scholarship for this thesis co-funded by the Université Grenoble Alpes, the Institut Laue Langevin (ILL) and ISIS Neutron and Muon Source (STFC Rutherford Appleton Laboratory) and for the support of my home laboratory, LIPhy. A special thanks to Marc Johnson who helped to acquire this grant.

I would like to thank the Institut Laue Langevin (ILL), ISIS Neutron and Muon Source and Heinz Maier-Leibnitz Zentrum (MLZ) for the allocation of neutron beamtime.

In particular, I am grateful towards the ILL for the opportunity to work in an international and welcoming environment and for the financial support it provided me for a three months prolongation of my PhD contract.

I want to thank the co-responsible of IN13, Francesca Natali for her help and useful discussions. During my first stay a ISIS, I have received invaluable support from Mark Telling and Ian Silverwood. In terms of MD simulations and the MDANSE software, Eric Pellegrini, Miguel Gonzalez and Remi Perenon were of a great help.

A big thank you goes to all fellow PhD students for making my stay in Grenoble a wonderful experience. I also want to thank them for electing me as PhD representative of the ILL in 2017/18 together with Lindsay McGregor. During this time, Henry Fischer, as the head of the ILL Graduate school, offered us great support and Ana Țuțueanu helped in organising numerous social events.

To all my office mates, Maksym, Loreto, Marta, Bastien, Aline and Irina, thank you for useful, but also funny discussions and the constant friendly atmosphere. A special thanks to Maksym who explained me a lot about QENS, the football sessions and his nice welcome in Estonia.

I want to thank Matthias, Hendrik, Marco, Ana and Marc for proofreading parts of this thesis and Aline for the help with the translation in French.

Finally, I want to express my deepest gratitude to Tese and my family for their continuous support and especially my grandfather who sadly passed away shortly before the conclusion of this thesis.

Contents

1. Introduction

1.1. English version

In 1969, Schoenborn [1] performed one of the first diffraction experiments with neutrons on biological samples. He investigated the structure of myoglobin and showed with his study that the analysis of proteins by means of neutron scattering was feasible. In 1989, Doster et al. [2] discovered the so called dynamical transition in proteins whose origins are still debated today [3, 4]. They used neutron spectroscopy to investigate the local dynamics of myoglobin in the temperature range of 20 K to 310 K and found that above 220 K, the internal motions increased significantly. Building on this work, an entire field of research in neutron scattering has developed. Over 300 articles in refereed journals have been published using this or similar neutron techniques over the last three decades [5], trying to either find the origin of the transition or, more generally, trying to understand the link between atomic scale dynamics and protein functionality.

In general, in a neutron scattering experiment a sample is illuminated by an incoming neutron beam, which upon interaction with the sample can either be scattered or absorbed. The measured scattered signal is composed of two different contributions, a coherent part and an incoherent one. The coherent part gives information about the structure and collective motions in the sample, whereas the incoherent part probes the average motions of individual atoms within the sample. The latter can be used to study molecular dynamics [6]. A particularity of biomolecules is their high content of hydrogen atoms (H), around 50%, which have the largest incoherent cross-section, 80 times larger than any other natural element present in them [6]. Furthermore, for neutron scattering experiments on biomolecules the isotope deuterium (D) can be used as a contrast agent, either to replace H atoms without significantly affecting the chemical nature of the biomolecule, or by using heavy water (D_2O) as the solvent instead of H_2O. This is possible since in neutron scattering it is not the element which defines the cross section (in contrast to, for example, X-rays) but the nucleus composition. The incoherent scattering cross section of H is 40 times larger than for D.

In the scope of this thesis the focus is on the incoherent part of the neutron scattering signal. During the interaction with the sample the neutron can not only change its direction but also its energy. The resulting energy gain or loss of the neutron can be classified into three different types, elastic incoherent neutron scattering (EINS, no energy exchange), quasi-elastic neutron scattering (QENS, a small energy exchange manifested as a broadening around EINS) and inelastic neutron scattering (INS, an energy exchange taking place at discrete values). Analysis of the contribution of EINS gives information about localised and confined motions, whereas QENS can give additional insights about the type of motions, like for example, translational or rotational diffusion. INS corresponds

to specific modes or excitations within the sample. The theoretical background and mathematical description of neutron scattering is given in Chapter 2.

In this thesis, the EINS extracted from biomolecules is investigated in detail. The analysis of such data is often used to understand their dynamical behaviour at a molecular level and to study the effect of a number of parameters of biological relevance such as temperature, pressure, hydration level and pH. Specifically in the case of proteins, the role of dynamics on ligand binding is important. For that, the mean square displacement (MSD) is obtained, which is characteristic for the local dynamics of the H atoms in the sample. Almost exclusively, one model is used to extract the MSD from the EINS signal, the Gaussian approximation [7, 8]. It is a harmonic approximation of all possible dynamical contributions, as there is to date no complete model that is able to account for all of them. In the past not many efforts were made to improve the description and only lately several new methods have been published [9, 10]. They mainly take into account the heterogeneity in the dynamics of different hydrogen atoms, by assuming different kinds of continuous distributions, for example, the Weibull [11] or Gamma [10] distributions. Other kinds of corrections are also discussed, but the main focus in the field remains the inclusion of heterogeneity, because it is considered the main deviation from the Gaussian approximation [12, 13]. In the thesis various models found in the literature and their assumptions are discussed in the second part of chapter 2, notably in section 2.2 and some of them are then chosen for the comparison with the Gaussian approximation.

Another important point in the analysis of neutron scattering data for the extraction of information on the dynamical properties of a sample, is the instrumental time and space resolution. Each neutron spectrometer has its own specifications which can highly influence the results. Commonly in neutron spectroscopy two quantities are measured, the energy transfer of the neutron due to the interaction with the sample, and the scattering angle, which together define the momentum transfer. The instrumental energy resolution and energy transfer window define the timescale of motions which can be observed. For the spectrometers used in this thesis, the dynamical motions explored are in the pico- to nano-second timescale. The magnitude of the momentum transfer defines the observed spatial window and thus the range of amplitudes of motions which can be registered. Since different instruments cover different temporal and spatial scales, it is important to compare the model fitting of data from different instruments and check for consistency of results. The instrumental details and properties of the used spectrometers are introduced in chapter 3.

The focus of this thesis is to perform a systematic comparison of models and their impact on resulting dynamical information (e.g. MSD) using different but typically used neutron spectrometers on a common sample. To this end the protein Alpha-Lactalbumin was chosen and measured on four different instruments under three different hydration levels, including the dry powder. The particular interest of this study was the analysis of neutron scattering data, mainly the evaluation of EINS data and its influence on the obtained results, the MSD. Alpha-Lactalbumin was chosen as a good model system since it is commercially available in large quantities (grams), there is no published EINS data, it is a small biomolecule, which enables complementary molecular dynamics simulations to be performed in a reasonable time, and finally, it can be studied in one of two forms, calcium

rich or depleted. The sample properties and preparation, together with the experimental details are summarised in chapter 4.

Chapter 5 contains the comparison of four distinct models for the description of EINS data from three neutron spectrometers on the aforementioned protein. The advantages and disadvantages of using each model are discussed and the resulting MSDs are shown in detail. In addition, a comparison of the MSD for Alpha-Lactalbumin with and without calcium is shown, illustrating the importance of the described methods if only small changes are visible between samples.

In chapter 6 the experimental results are compared to molecular dynamics (MD) simulations. MD simulations are a useful tool in neutron spectroscopy since they cover the same spatial and temporal ranges. The two methods are highly complementary as neutron experiments can be used to validate MD simulations and in return, they can help explain, interpret and visualise experimental results. The simulations were performed under the same conditions as for the experiments, i.e. one as dry powder and one as hydrated powder. The results of two different temperatures, 280 K and 300 K, are compared using three different models and direct calculations through the MD trajectories.

Chapter 7 summarises the main finding and provides an outlook of further developments in the field and possible improvements.

During the course of the thesis four refereed articles have been published with my contributions. The main results of this thesis are contained in the first publication below, and some passages in this thesis have been quoted verbatim from there. Publications 2 and 3 are directly related to the research in this thesis by analysing the respective EINS data. The last publication was not related to neutron scattering and my contribution was smaller with some help during the experiments and a part of the data treatment.

1. **D. Zeller**, M. T. F. Telling, M. Zamponi, V. Garcia Sakai and J. Peters,
   Analysis of incoherent neutron scattering data beyond the Gaussian approximation,
   The Journal of Chemical Physics, **2018**, 149 (23), 234908.
2. C. D. Andersson, N. Martinez, **D. Zeller**, S. H. Rondahl, M. M. Koza, B. Frick, F. Ekström, J. Peters and A. Linusson,
   Changes in dynamics of α-chymotrypsin due to covalent inhibitors investigated by elastic incoherent neutron scattering,
   Physical Chemistry Chemical Physics, **2017**, 19, 25369-25379.
3. C. D. Andersson, N. Martinez, **D. Zeller**, A. Allgardsson, M. M. Koza, B. Frick, F. Ekström, J. Peters, and A. Linusson,
   Influence of Enantiomeric Inhibitors on the Dynamics of Acetylcholinesterase Measured by Elastic Incoherent Neutron Scattering,
   The Journal of Physical Chemistry B, **2018**, 122 (36), 8516-8525.
4. L. Kangur, K. Timpmann, **D. Zeller**, P. Masson, J. Peters, and A. Freiberg,
   Structural stability of human butyrylcholinesterase under high hydrostatic pressure,
   Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics, **2019**, 1867 (2), 107-113.

All publications can be found in Appendix B.

1.2. Version française

En 1969, Schoenborn [1] a réalisé l'une des premières expériences de diffraction neutronique sur des échantillons biologiques. Il a analysé la structure de la myoglobine et montré ainsi que l'étude de protéines par diffusion de neutrons était possible. En 1989, Doster et al. [2] ont découvert la transition dynamique des protéines dont les origines sont encore discutées aujourd'hui [3, 4]. Ils ont utilisé la diffusion de neutrons (in)élastique sur la myoglobine pour étudier sa dynamique locale dans le domaine de températures de 20K à 310K et ont découvert qu'au-dessus de 220K, les mouvements internes augmentent considérablement. Cette étude a ouvert un tout nouveau champ de recherche en diffusion neutronique. Au cours des trois dernières décennies [5], plus de 300 articles dans des revues référées ont été publiés, utilisant cette technique ou des techniques neutroniques similaires, dans le but de trouver l'origine de cette transition ou, plus généralement, de comprendre le lien entre la dynamique à l'échelle atomique et la fonctionnalité des protéines.

En général, dans une expérience de diffusion neutronique, un échantillon est éclairé par un faisceau de neutrons, qui peut alors être diffusé ou absorbé après l'interaction. Les neutrons diffusés sont composés de deux contributions différentes, une partie cohérente et une partie incohérente. La partie cohérente fournit des informations sur la structure et les mouvements collectifs dans l'échantillon, tandis que la partie incohérente comprend les mouvements moyennés de chaque atome dans l'échantillon. Cette dernière peut être utilisée pour étudier la dynamique moléculaire [6]. La partie incohérente du signal diffusé est particulièrement importante pour les biomolécules, car l'atome d'hydrogène (H) possède l'une des plus grandes sections efficaces incohérentes [6]. Elle est en effet 80 fois plus grande que la section efficace incohérente de tout autre élément naturel dans les biomolécules. En outre, l'isotope deutérium (D) peut être utilisé comme agent de contraste, soit pour remplacer les atomes de H dans l'échantillon, sans changer significativement la nature chimique de la biomolécule, soit en utilisant de l'eau lourde (D2O) comme solvant au lieu de H2O. Cette méthode de contraste est réalisable car en diffusion neutronique, ce n'est pas l'élément chimique qui définit la section efficace (contrairement aux rayons X, par exemple), mais la composition du noyau. La section efficace de diffusion incohérente de l'hydrogène H est ainsi 40 fois plus grande que celle du deutérium D.

Dans le cadre de cette thèse, nous nous concentrons sur la partie incohérente de la diffusion des neutrons. Lors de l'interaction avec l'échantillon, le neutron peut non seulement changer de direction, mais également d'énergie. Le gain ou la perte en énergie des neutrons qui en résulte peut être divisé en trois parties différentes : la diffusion de neutrons élastique incohérente (EINS, pas d'échange d'énergie), la diffusion de neutrons quasi-élastique (QENS, petit échange d'énergie autour de l'EINS) et la diffusion de neutrons inélastique (INS, un échange d'énergie se produisant pour des valeurs discrètes). L'analyse de la contribution élastique, EINS, fournit des informations sur les mouvements localisés et confinés, tandis que le QENS peut fournir des informations sur les types de mouvement, comme par exemple, la diffusion translationnelle ou rotationnelle. INS correspond à des modes spécifiques ou à des excitations dans l'échantillon. Le chapitre 2 expose plus en détail le contexte théorique et la description mathématique de la diffusion des neutrons.

Dans cette thèse, la contribution EINS extraite pour les biomolécules est étudiée en détail.

La technique est souvent utilisée pour obtenir les propriétés dynamiques des protéines à l'échelle moléculaire et pour les étudier dans diverses conditions, par exemple dans différents environnements (température, pression, niveau d'hydratation, pH, etc.) ou, dans le cas des protéines, pour rechercher l'existence de ligands. Pour cela, le déplacement carré moyen (MSD) localisé de l'échantillon est calculé, qui est caractéristique de la dynamique locale des atomes d'H dans l'échantillon. De manière générale, un modèle est privilégié pour extraire les MSD du signal EINS : l'approximation gaussienne [7, 8]. Il s'agit d'une approximation harmonique de toutes les contributions dynamiques possibles, car il n'existe à ce jour aucun modèle complet les prenant toutes en compte. Peu d'efforts ont été déployés pour améliorer leur description, ce n'est seulement récemment que plusieurs nouvelles méthodes ont été publiées [9, 10]. Elles prennent principalement en compte l'hétérogénéité des différents mouvements des atomes d'hydrogène, ce qui implique différents types de distributions continues, comme par exemple la distribution de Weibull [11] ou Gamma [10]. D'autres types de corrections ont également été proposés, mais l'objectif principal dans ce domaine reste l'intégration de l'hétérogénéité, car elle est considérée comme la principale déviation par rapport à l'approximation gaussienne [12, 13]. Dans la thèse, différents modèles trouvés dans la littérature ainsi que leurs hypothèses sont discutés dans la deuxième partie du chapitre 2, notamment dans la section 2.2, et certains de ces modèles sont ensuite choisis pour les comparer à l'approximation gaussienne.

Un autre point important à considérer pour l'analyse des données de diffusion neutronique, et l'extraction des propriétés dynamiques des échantillons, est la résolution temporelle et spatiale des instruments. Chaque spectromètre à neutrons a ses propres caractéristiques qui peuvent grandement influer sur les résultats. En spectroscopie neutronique, nous mesurons généralement deux quantités : le transfert d'énergie du neutron dû à l'interaction avec l'échantillon, et l'angle de diffusion, qui définissent ensemble le transfert d'impulsion. La résolution instrumentale en énergie définit la fenêtre temporelle pouvant être observée. Pour les neutrons, la gamme temporelle qui peut être étudiée est dans le domaine des pico- aux nanosecondes. D'autre part, la valeur absolue du moment transféré définit la fenêtre spatiale observée, et donc les amplitudes maximales des mouvements qui peuvent être enregistrées. Puisque différents instruments couvrent des échelles de temps et d'espace différentes, il est important de comparer l'effet des modèles pour différents instruments et de vérifier la cohérence des résultats. Les détails instrumentaux et les propriétés des spectromètres utilisés sont présentés au chapitre 3.

L'objectif de cette thèse est d'effectuer une comparaison systématique de différents modèles et leur impact sur les propriétés dynamiques résultantes (par exemple les MSD), en utilisant des spectromètres différents, mais typiquement employés, sur un même échantillon. Pour cela, nous avons choisi la protéine Alpha-Lactalbumine et nous l'avons mesurée sur quatre instruments différents à trois niveaux d'hydratation différents : des poudres sèches et deux poudres hydratées distinctes. L'intérêt portait particulièrement sur l'analyse des données de diffusion neutronique, principalement l'évaluation des données EINS et leur influence sur les résultats obtenus, les MSD. L'alpha-lactalbumine a été choisie comme un bon système modèle, car elle est disponible dans le commerce en grandes quantités (grammes), aucune mesure élastique en température n'a été effectuée auparavant et il s'agit d'une petite biomolécule. Ce dernier point était important pour pouvoir effectuer des

simulations de dynamique moléculaire dans des temps raisonnables. Enfin, cette protéine peut être étudiée sous deux formes : avec ou sans calcium. Les propriétés de l'échantillon et la préparation d'Alpha-Lactalbumine, ainsi que les détails expérimentaux, sont résumés au chapitre 4.

Le chapitre 5 expose la comparaison de quatre modèles distincts d'EINS, pour trois spectromètres neutroniques, appliqués à la protéine Alpha-Lactalbumine. Les avantages et les inconvénients de chaque modèle sont discutés et les MSD résultants sont présentés en détail. De plus, une comparaison entre Alpha-Lactalbumine avec et sans calcium est effectuée via les MSD, démontrant l'importance des méthodes décrites lorsque seules de petites modifications sont présentes dans les échantillons.

Dans le chapitre 6, les résultats expérimentaux sont comparés aux simulations de dynamique moléculaire (MD). Les simulations MD constituent des outils précieux en spectroscopie neutronique, car elles couvrent les mêmes échelles de temps et d'espace. Les deux méthodes sont ainsi complémentaires : si les expériences neutroniques peuvent être utilisées pour valider des simulations MD, en retour, elles peuvent aider à expliquer, interpréter et visualiser des résultats expérimentaux. Les simulations ont été réalisées dans les mêmes conditions que les expériences : avec une poudre sèche et une poudre hydratée. Les résultats pour deux températures différentes, 280K et 300K, sont comparés grâce à trois modèles différents et grâce à des calculs directs utilisant les trajectoires MD.

Le chapitre 7 résume les principaux résultats et apporte un aperçu des développements futurs dans ce domaine, ainsi que des améliorations possibles.

Au cours de la thèse, quatre publications ont été publiées avec ma contribution. Les principaux résultats de cette thèse peuvent être trouvées dans la première publication, citée ci-dessous, et certains passages de cette thèse sont directement repris de ce manuscrit. Les publications 2 et 3 sont directement liés au sujet de recherche de cette thèse, avec l'analyse des données EINS correspondantes. La dernière publication ne concerne pas la diffusion neutronique et ma contribution se résume à avoir aidé lors des expériences et pour une partie du traitement des données.

1. **D. Zeller**, M. T. F. Telling, M. Zamponi, V. Garcia Sakai and J. Peters,
   Analysis of incoherent neutron scattering data beyond the Gaussian approximation,
   The Journal of Chemical Physics, **2018**, 149 (23), 234908.
2. C. D. Andersson, N. Martinez, **D. Zeller**, S. H. Rondahl, M. M. Koza, B. Frick, F. Ekström, J. Peters and A. Linusson,
   Changes in dynamics of α-chymotrypsin due to covalent inhibitors investigated by elastic incoherent neutron scattering,
   Physical Chemistry Chemical Physics, **2017**, 19, 25369-25379.
3. C. D. Andersson, N. Martinez, **D. Zeller**, A. Allgardsson, M. M. Koza, B. Frick, F. Ekström, J. Peters, and A. Linusson,
   Influence of Enantiomeric Inhibitors on the Dynamics of Acetylcholinesterase Measured by Elastic Incoherent Neutron Scattering,
   The Journal of Physical Chemistry B, **2018**, 122 (36), 8516-8525.
4. L. Kangur, K. Timpmann, **D. Zeller**, P. Masson, J. Peters, and A. Freiberg,
   Structural stability of human butyrylcholinesterase under high hydrostatic pressure,
   Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics, **2019**, 1867 (2), 107-113.

Toutes les publications se trouvent dans l'annexe B.

2. Theoretical background

In the first part of this chapter, the basics of neutron scattering will be presented including a mathematical description relevant to the neutron scattering experiments performed for this thesis. In the second part, the description will focus on the elastic incoherent neutron scattering part in detail.

2.1. Basics of neutron scattering

This section is a brief introduction to neutron scattering and is mainly based on the books of G.L. Squires [8] and M. Bée[6]. For more in depth treatment the reader is invited to read these books or the article of H. Schober [14].

2.1.1. Properties of the neutron

The neutron is a subatomic particle with a mass of m =1.675 $\times$ 10^{-27} kg and a spin of 1/2. It is uncharged and has a magnetic dipole moment μ_{n}. The main characteristics are summarised in Table 2.1. In general, the neutron is found in the atomic nucleus together with protons, where it is stable. To use it as a probe for neutron scattering experiments it has to be extracted from the nucleus. As a free neutron has a lifetime of only around 15 min, it has to be produced steadily to perform experiments. To date there are two main methods used, reactor based sources or spallation sources, and they will be described in more detail in section 3.1).

In the non-relativistic case (which will be assumed in all further treatment here) the kinetic energy of a neutron with a velocity v is

$$E = \frac{1}{2}mv^2 \tag{2.1}$$

Due to the wave-particle duality in quantum mechanics, it is possible to describe the neutron also as a wave, with a wavelength λ defined by de Broglie [8]:

$$\lambda = \frac{h}{mv} \tag{2.2}$$

where h is Planck's constant. The wavevector $\mathbf{k}$ is defined to have the magnitude

$$|\mathbf{k}| = k = \frac{2\pi}{\lambda}, \tag{2.3}$$

and its direction is that of $\mathbf{v}$. The momentum $\mathbf{p}$ of the neutron is [8]

$$\mathbf{p} = \hbar\mathbf{k}, \text{ with } \hbar = \frac{h}{2\pi}. \tag{2.4}$$

In order to change the wavelength of the incoming neutrons, the velocity has to be changed. This is often done with moderators resulting in a Maxwellian distribution of the neutron flux as a function of temperature. Therefore it is conventional to say that a neutron with energy E corresponds to a temperature T [8] according to

$$E = k_B T. \tag{2.5}$$

We can rewrite the energy of the neutron as:

$$E = k_B T = \frac{1}{2}mv^2 = \frac{h^2}{2m\lambda^2} = \frac{\hbar^2 k^2}{2m} \tag{2.6}$$

For thermal neutrons (see section 3.1) a value of $v = 2.20\,\mathrm{km/s}$ is conventionally taken as standard velocity. For this value following properties of the neutron can be calculated [8]:

$$\begin{aligned} v &= 2.20\,\mathrm{km\,s^{-1}}, & \frac{1}{v} &= 455\,\mathrm{\mu s\,m^{-1}} \\ E &= 25.3\,\mathrm{meV}, & T &= 293\,\mathrm{K} \\ \lambda &= 1.798\,\mathrm{\AA}, & k &= 3.49 \times 10^{10}\,\mathrm{m^{-1}} \end{aligned} \tag{2.7}$$

Table 2.1. - Basic properties of the neutron. μ_N is the nuclear magneton with $\mu_\mathrm{N} = 5.051 \times 10^{-27}\,\mathrm{JT^{-1}}$. All values are taken from [8], except neutron mean lifetime which is taken from [15].

mass m	$1.675 \times 10^{-27}\,\mathrm{kg}$
charge	0
spin	1/2
magnetic dipole moment μ_n	$1.913\,\mu_\mathrm{N}$
mean lifetime	$(880.2 \pm 1.0)\,\mathrm{s}$

2.1.2. Scattering theory

In a neutron experiment, it is the interaction of the neutron with the nuclei of a sample that is investigated. During such interaction the neutron either changes its momentum $\mathbf{k}$ or it is absorbed. In a scattering event with neutrons two quantities are measured: 1) The energy transfer $\hbar\omega$, which is the difference between the initial E_0 and the final neutron energy E_f after the scattering event; 2) The momentum transfer $\mathbf{Q}$, which is the difference between the incoming momentum $\mathbf{k}_0$ and the final momentum $\mathbf{k}_f$ [6], and can be written

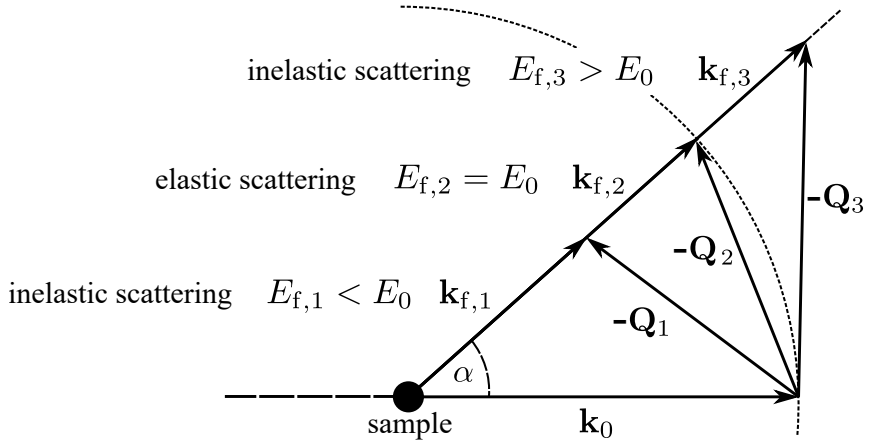


Figure 2.1. - Schematic of $\mathbf{Q} = \mathbf{k_0} - \mathbf{k}_f$ and its influence on the energy. Variation of the momentum transfer vector $\mathbf{Q}$ at constant scattering angle α. See also Eq. (3.8) for calculation of the elastic and inelastic momentum transfer for a spectrometer. Figure inspired by [6, p.11].

as:

$$\hbar\omega = E_0 - E_f = \frac{\hbar^2}{2m}(k_0^2 - k_f^2) \tag{2.8}$$

$$\mathbf{Q} = \mathbf{k}_0 - \mathbf{k}_f. \tag{2.9}$$

A change in the momentum transfer can happen with or without energy transfer. If the energy of the neutron changes, $\mathbf{Q}$ cannot be kept constant if the scattering angle between $\mathbf{k}_0$ and $\mathbf{k}_f$ stays the same [6]. A schematic is shown in Figure 2.1.

The results of a scattering event can be expressed in terms of a quantity known as a the cross section σ [8]. Since a scattering event can results in a change of momentum and energy of the neutron, it can be described by the double differential cross section

$$\frac{\partial^2\sigma_\mathrm{s}}{\partial\mathbf{\Omega}\partial E} = \frac{1}{\hbar}\frac{\partial^2\sigma_\mathrm{s}}{\partial\mathbf{\Omega}\partial\omega} \tag{2.10}$$

It represents the probability that a neutron is scattered into a small solid angle $\mathrm{d}\Omega$ in the direction $\mathbf{\Omega}$ with a final energy between $E = \hbar\omega$ and $E + \mathrm{d}E = \hbar(\omega + \mathrm{d}\omega)$. Assuming an incoming neutron flux of I_0 on the sample, the total scattered neutron flux I_s is then defined as

$$I_\mathrm{s} = I_0 \int \mathrm{d}E \int \mathrm{d}\mathbf{\Omega}\, \frac{\partial^2\sigma_\mathrm{s}}{\partial\mathbf{\Omega}\partial E} = I_0 \int \mathrm{d}\omega \int \mathrm{d}\mathbf{\Omega}\, \frac{1}{\hbar}\frac{\partial^2\sigma_\mathrm{s}}{\partial\mathbf{\Omega}\partial\omega} = I_0\sigma_\mathrm{s} \tag{2.11}$$

where σ_s is the scattering cross section. Similarly an absorption cross section σ_a can be defined for neutrons absorbed, i.e. trapped by a nucleus in the sample. Together they evaluate to the total cross section of the sample:

$$\sigma_\mathrm{tot} = \sigma_\mathrm{s} + \sigma_\mathrm{a} \tag{2.12}$$

Neutrons interact with a nucleus via nuclear and magnetic forces. The magnetic interactions due to the spin of the neutron (via dipole-dipole interaction) will not be considered here since for diamagnetic systems it is negligible compared to nuclear interactions [6]. Nuclear forces are in the range of about 10^{-14} to 10^{-15} m whereas the wavelength of thermal neutrons is in the order of 10^{-10} m (see Eq. (2.7)). Thus the neutron-nucleus scattering can be entirely described by s-waves, which corresponds to spherically symmetric, i.e. isotropic, scattering [6, 8]. This leads to a simple description of the scattering process via one single parameter, the scattering length b. It is independent of the neutron energy and as a complex number, the real part can be negative or positive depending of the attractive or repulsive nature of the interaction (by definition a positive b implies a repulsive portential). The imaginary part represents absorption [6].

Using this description, the interaction between the neutron and a nucleus i can be expressed by the Fermi pseudopotential

$$V(\mathbf{r}) = \frac{2\pi\hbar^2}{m} b_i \delta(\mathbf{r} - \mathbf{R}_i), \tag{2.13}$$

where $\mathbf{r}$ is the location of the neutron, $\mathbf{R}_i$ the position of the nucleus i and δ the Dirac delta function. Each atomic species has its respective scattering length b_i which changes not only with the atom type but also with each isotope because the interaction does not only depend on the nature of the nucleus but also on the total spin state of the nucleus-neutron system [6]. The mean value $< b_i >$ of b_i over all the isotopes and spin states is called the coherent scattering length, whereas the root mean square deviation (RMSD) of b_i from $< b_i >$ is called incoherent scattering length. Mathematically this is written as:

$$b_i^{\mathrm{coh}} =< b_i > \tag{2.14}$$

$$b_i^{\mathrm{inc}} = \left[< b_i^2 > - < b_i >^2\right]^{1/2}. \tag{2.15}$$

The scattering length can then be related to the scattering cross section via

$$\sigma_s = \sigma_{\mathrm{coh}} + \sigma_{\mathrm{inc}} = 4\pi(b_{\mathrm{coh}}^2 + b_{\mathrm{inc}}^2). \tag{2.16}$$

The incoherent and coherent scattering cross sections, as well as the absorption cross sections of the most important atom types and isotopes used in this thesis are given in Table 2.2. There are two important things to mention. First, Fermi's pseudopotential does not correspond to the actual potential in the nucleus and a positive scattering length does not imply that the actual potential is repulsive. Secondly, the scattering length is defined for a fixed nucleus, also known as the bound scattering length. If the nucleus is free, the scattering must be treated in the center of mass system and the mass m of the neutron in Eq. (2.13) must be replaced by the reduced mass $\mu = mM/(m + M)$, where M is the mass of the nucleus [8].

If the perturbation experienced by the initial wave function (incident neutron) in the course of scattering is small, the Born approximation can be used [14]. Then the double

differential cross section can be written as [6, 8]

$$\frac{\partial^2\sigma_\mathrm{s}}{\partial\mathbf{\Omega}\partial\omega} = \frac{k_f}{k_0}\frac{1}{2\pi}\sum_i\sum_j b_i b_j \int_{-\infty}^{\infty} \langle \exp\{-i\mathbf{Q}\cdot\mathbf{R}_i(0)\}\exp\{i\mathbf{Q}\cdot\mathbf{R}_j(t)\}\rangle \times \exp(i\omega t)\,\mathrm{d}t, \tag{2.17}$$

where <> is the thermal average. It is important to remark that the equation is derived from quantum mechanics (QM) and $\mathbf{R}(\mathbf{t})$ is a Heisenberg operator and not a classical vector. This result can be rewritten in terms of the coherent and incoherent cross section, assuming no correlation between the scattering lengths of the different nuclei:

$$\frac{\partial^2\sigma_\mathrm{s}}{\partial\mathbf{\Omega}\partial\omega} = \left(\frac{\partial^2\sigma_\mathrm{s}}{\partial\mathbf{\Omega}\partial\omega}\right)_\mathrm{coh} + \left(\frac{\partial^2\sigma_\mathrm{s}}{\partial\mathbf{\Omega}\partial\omega}\right)_\mathrm{inc}. \tag{2.18}$$

As in this thesis mainly the scattering on one isotope, hydrogen H^1, is important, in the following equations only one isotope is assumed (for the inclusion of several atom types see [6]). For one isotope, Eq. (2.17) and (2.18) result in

$$\left(\frac{\partial^2\sigma_\mathrm{s}}{\partial\mathbf{\Omega}\partial\omega}\right)_\mathrm{coh} = \frac{\sigma_\mathrm{coh}}{4\pi}\frac{k_f}{k_0}\,NS_\mathrm{coh}(\mathbf{Q},\omega) \tag{2.19}$$

$$\left(\frac{\partial^2\sigma_\mathrm{s}}{\partial\mathbf{\Omega}\partial\omega}\right)_\mathrm{inc} = \frac{\sigma_\mathrm{inc}}{4\pi}\frac{k_f}{k_0}\,NS_\mathrm{inc}(\mathbf{Q},\omega), \tag{2.20}$$

where N is the total number of nuclei in the scattering system. $S_\mathrm{coh}(\mathbf{Q},\omega)$ and $S_\mathrm{inc}(\mathbf{Q},\omega)$ are the so-called coherent and incoherent dynamic scattering functions (DSF), defined as

$$S_\mathrm{coh}(\mathbf{Q},\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} I_\mathrm{coh}(\mathbf{Q},t)\times\exp(i\omega t)\,\mathrm{d}t \tag{2.21}$$

$$S_\mathrm{inc}(\mathbf{Q},\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} I_\mathrm{inc}(\mathbf{Q},t)\times\exp(i\omega t)\,\mathrm{d}t. \tag{2.22}$$

They are the time-Fourier transform of the so-called intermediate scattering functions (ISF), $I_\mathrm{coh}(\mathbf{Q},t)$ and $I_\mathrm{inc}(\mathbf{Q},t)$:

$$I_\mathrm{coh}(\mathbf{Q},t) = \frac{1}{N}\sum_i\sum_j \langle \exp\{-i\mathbf{Q}\cdot\mathbf{R}_i(0)\}\exp\{i\mathbf{Q}\cdot\mathbf{R}_j(t)\}\rangle \tag{2.23}$$

$$I_\mathrm{inc}(\mathbf{Q},t) = \frac{1}{N}\sum_i \langle \exp\{-i\mathbf{Q}\cdot\mathbf{R}_i(0)\}\exp\{i\mathbf{Q}\cdot\mathbf{R}_i(t)\}\rangle. \tag{2.24}$$

The space-Fourier transformation of $I_\mathrm{coh}(\mathbf{Q},t)$ and $I_\mathrm{inc}(\mathbf{Q},t)$ give rise to the van Hove pair-correlation function $G(\mathbf{r},t)$ in case of coherent scattering and to the self-correlation

Table 2.2. - Summary of scattering and absorption cross sections. The values are given in barn ($1\,\mathrm{b} = 10^{-24}\,\mathrm{cm}^2$) and are taken from [16]. If no isotope is mentioned, the average value for the natural abundance is used. Atom notation: $\mathrm{Atom}^{Z}_{A=Z+N}$

atom	σ_{coh} [b]	σ_{inc} [b]	σ_{a} [b]
H^{1}_{1}	1.758	80.27	0.3326
H^{1}_{2}/D	5.592	2.05	0.000 52
C^{6}	5.551	0.001	0.0035
N^{7}	11.01	0.5	1.9
O^{8}	4.232	0.0008	0.000 19
S^{16}	1.017	0.007	0.53
Al^{13}	1.495	0.0082	0.231
Ca^{20}	2.78	0.05	0.43
V^{23}	0.018	5.08	5.08
In^{49}	2.08	0.54	193.8

function $G_{\mathrm{s}}(\mathbf{r},t)$ in case of incoherent scattering [8]:

$$G(\mathbf{r},t) = \frac{1}{N}\sum_{i}\sum_{j}\int \langle \delta\left\{\mathbf{r}' - \mathbf{R}_i(0)\right\}\delta\left\{\mathbf{r}+\mathbf{r}' - \mathbf{R}_j(t)\right\}\rangle \,\mathrm{d}\mathbf{r}' \tag{2.25}$$

$$G_{\mathrm{s}}(\mathbf{r},t) = \frac{1}{N}\sum_{i}\int \langle \delta\left\{\mathbf{r}' - \mathbf{R}_i(0)\right\}\delta\left\{\mathbf{r}+\mathbf{r}' - \mathbf{R}_i(t)\right\}\rangle \,\mathrm{d}\mathbf{r}'. \tag{2.26}$$

Importantly, the operators $\mathbf{R}_i(0)$ and $\mathbf{R}_j(t)$ do not commute, except for $t = 0$, and therefore the order of the operators has to be preserved. In the classical limit it can be ignored that $\mathbf{R}_i(0)$ and $\mathbf{R}_j(t)$ do not commute and the integration of Eq. (2.25) and (2.26) can be solved [8]:

$$G^{\mathrm{cl}}(\mathbf{r},t) = \frac{1}{N}\sum_{i}\sum_{j}\langle \delta\left\{\mathbf{r} - \mathbf{R}_j(t) + \mathbf{R}_i(0)\right\}\rangle \tag{2.27}$$

$$G^{\mathrm{cl}}_{\mathrm{s}}(\mathbf{r},t) = \frac{1}{N}\sum_{i}\langle \delta\left\{\mathbf{r} - \mathbf{R}_i(t) + \mathbf{R}_i(0)\right\}\rangle\,. \tag{2.28}$$

In conclusion, the coherent scattering depends on the correlation between the positions of the same nucleus at different times and on the correlation between the positions of different nuclei at different times, leading to interference effects. In contrast, the incoherent scattering depends only on the correlation between the positions of the same nucleus at different times which do not give interference effects [8]. Thus, coherent scattering can give information on structure and collective motions (correlations between all scatterers), whereas incoherent scattering gives information about the space evolution in time (self-correlation), leading to a probe of the local dynamics of the sample.

2.1.3. The dynamical scattering function $S(\mathbf{Q}, \omega)$

In many neutron scattering experiments, the dynamical scattering function (DSF), $S(\mathbf{Q}, \omega)$, which is the sum of the coherent and incoherent DSF defined in Eq. (2.21) and (2.22), is directly measured. It is the probability density of neutrons which are scattered under the scattering vector $\mathbf{Q}$ and with an energy transfer $\hbar\omega$. Depending on the energy transfer the DSF can be split in three parts: 1) The elastic neutron scattering (ENS, S_{ENS}) for neutrons which do not exchange energy with the sample, i.e. $\hbar\omega \approx 0$; 2) the quasi-elastic neutron scattering (QENS, S_{QENS}) where only small amounts of energy are exchanged (typically $< 2\,\mathrm{meV}$) which manifests itself as a broadening of the zero-energy transfer centered elastic peak; and 3) the inelastic neutron scattering (INS, S_{INS}) which appears as satellite peaks ($\hbar\omega \neq 0$) well separated from the elastic peak and corresponds to specific modes or excitations within the sample. A schematic is shown in Figure 2.2.

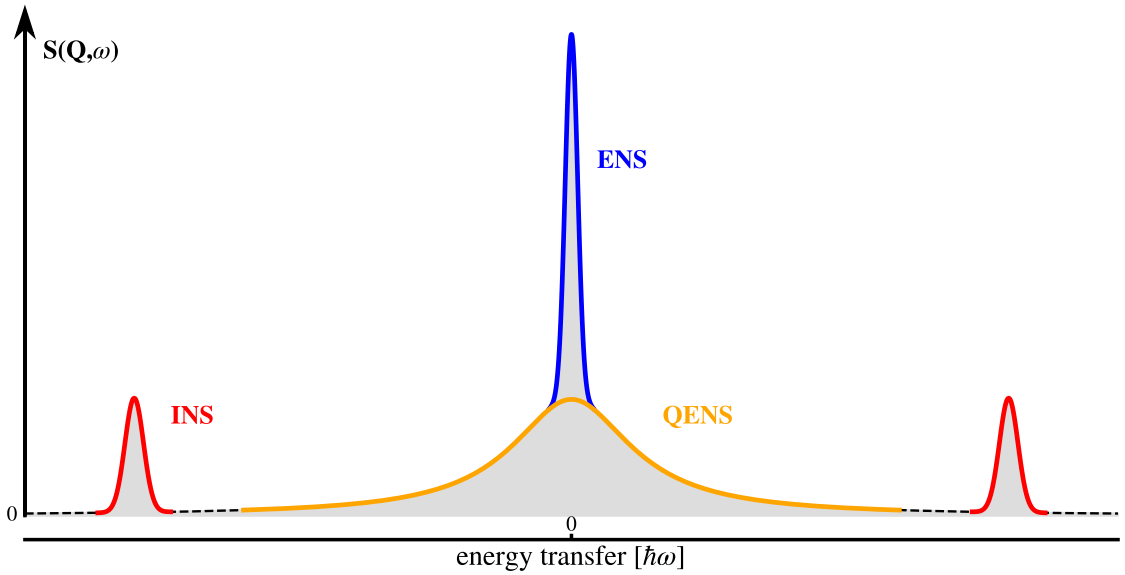


Figure 2.2. - Parts of the dynamical scattering function $S(\mathbf{Q}, \omega)$. Separation of the neutron spectrum $S(\mathbf{Q}, \omega)$ in elastic (ENS), quasi-elastic (QENS) and inelastic (INS) neutron scattering.

2.2. Models for Elastic Incoherent Neutron Scattering

As introduced in the last section, the DSF can be separated in three parts, elastic, quasi-elastic and inelastic neutron scattering. In this thesis, the ENS is investigated for the case of protein powder samples. In this context two important assumptions are made:

1) There is no significant coherent scattering contribution to the DSF so that the total scattering of the sample can be approximated solely by the incoherent DSF $S_{\mathrm{inc}}(\mathbf{Q}, \omega)$. This is in general the case for protein powders or proteins in solution in the investigated momentum transfer range of $Q = 0.3 - 5\,\mathrm{Å}^{-1}$ [17].

2) Hydrogen atoms are considered as the primary contributers to the scattering signal. This is a good approximation for biological samples, like proteins, for the following reason: as shown in section 2.1.2, Table 2.2, hydrogen H has a very large incoherent cross section of about 80 b and in general, biological systems contain $\approx$ 50% H atoms. The remaining atoms mainly consist of C, N and O which have all very small incoherent cross sections ($\sigma_{\text{inc}} < 0.5$ b), thus the majority of the signal comes from the incoherent scattering of H atoms in the protein structure. In most systems, the signal from hydrogen atoms in the surrounding or residual water (H_2O) will add to the signal, therefore often heavy water D_2O is used as medium since its incoherent scattering cross section is 40 times smaller with $\sigma_{\text{inc}} \approx 2$ b. Deuterium can also be used as substitute of hydrogen to mask specific regions in the protein.

Following from the first assumption, ENS spectroscopy is also often called EINS (elastic incoherent neutron scattering). In general, from an EINS experiment the total elastic scattering intensity is related to the mean square displacement (MSD) of the internal motions of the sample. To date, there is no complete model taking into account all possible dynamic contributions to the EINS. As a standard, the Gaussian Approximation (GA) is used to extract the MSD from the elastic data [8]. The limits of applicability of the models are not always clear and respected, such that the way in which they are implemented may sometimes seem arbitrary, with full details lacking in many publications. As long as different samples are analysed in exactly the same way and compared to each other, it may appear to be consistent, but it makes it difficult to compare the results from different samples, spectrometers and experimental groups [18]. In particular, using different neutron spectrometers is important for two main reasons. Firstly, each instrument has a specific energy resolution providing access to different dynamical time scales and secondly, each instrument has a specific range of accessible scattering angles, which define the spatial information probed. For the first effect Doster et al. proposed an analysis via elastic resolution spectroscopy in 2001 [19] and connected it to temperature scans via a scaling function in 2013 [20].

In section 2.2.1 an overview is given on the approximations that form the basis of the models for EINS, and it follows the general neutron scattering theory presented in section 2.1.2. In total four models will be discussed in more detail, the Gaussian Approximation (GA) (in 2.2.2), the Peters and Kneller (PK) model, the Yi et al. (Yi) model (both in 2.2.3) and the Doster (Do) model (in 2.2.4). At the end of this section, the influence of the instrument resolution on the MSD is discussed.

2.2.1. The principles of EINS

For the evaluation of the elastic incoherent neutron scattering signal, one can start with the intermediate scattering function (ISF) defined in Eq. (2.24):

$$I_{\text{inc}}(\mathbf{Q}, t) = \frac{1}{N} \sum_i \langle \exp\{-i\mathbf{Q} \cdot \mathbf{R}_i(0)\} \exp\{i\mathbf{Q} \cdot \mathbf{R}_i(t)\} \rangle \tag{2.29}$$

In the case of elastic scattering, the energy transfer is $\hbar\omega = 0$ which results in $t \to \infty$ for the time-Fourier transformation of the DSF. In the limit of $t \to \infty$ one gets

$$I_{\text{inc}}(\mathbf{Q}, t \to \infty) = \frac{1}{N} \sum_i \langle \exp\{-i\mathbf{Q} \cdot \mathbf{R}_i(0)\} \exp\{i\mathbf{Q} \cdot \mathbf{R}_i(\infty)\} \rangle \tag{2.30}$$

$$= \frac{1}{N} \sum_i \langle \exp\{-i\mathbf{Q} \cdot \mathbf{R}_i\} \rangle \langle \exp\{i\mathbf{Q} \cdot \mathbf{R}_i\} \rangle = \frac{1}{N} \sum_i |\langle \exp\{i\mathbf{Q} \cdot \mathbf{R}_i\} \rangle|^2. \tag{2.31}$$

To obtain line (2.31) it is assumed that in the long time limit: (1) $\mathbf{R}_i(0)$ and $\mathbf{R}_i(\infty)$ are statistically independent and (2) $\mathbf{R}_i(t)$ is translationally invariant such that $\langle \mathbf{R}_i(0) \rangle = \langle \mathbf{R}_i(\infty) \rangle = \langle \mathbf{R}_i \rangle$ are independent of time, which is the case for confined atoms [13]. We use now the cumulant expansion [21]

$$\langle \exp\{\pm i\mathbf{Q} \cdot \mathbf{R}_i\} \rangle = \exp\left\{-Q^2 \rho_2(\mathbf{n}_Q) + Q^4 \rho_4(\mathbf{n}_Q) - Q^6 \rho_6(\mathbf{n}_Q) + ...\right\}, \tag{2.32}$$

where $Q = |\mathbf{Q}|$, $\mathbf{n}_Q$ is the unit vector along $\mathbf{Q}$ and ρ_j are the moments

$$\rho_2(\mathbf{n_Q}) = \frac{1}{2!} \left\langle (\mathbf{n_Q} \cdot \mathbf{R}_i)^2 \right\rangle \tag{2.33}$$

$$\rho_4(\mathbf{n_Q}) = \frac{1}{4!} \left\{ \left\langle (\mathbf{n_Q} \cdot \mathbf{R}_i)^4 \right\rangle - 3 \left\langle (\mathbf{n_Q} \cdot \mathbf{R}_i)^2 \right\rangle^2 \right\} \tag{2.34}$$

$$\text{... .}$$

Here, two important approximations are used: (**I**) the motion of the individual atoms in the system is approximated by the first term of the cumulant expansion and (**II**) the motional distribution of the individual H atoms is isotropic, such that:

$$I_{\text{inc}}(\mathbf{Q}, t \to \infty) \overset{\text{(I)}}{\approx} \sum_i \exp\left\{-Q^2 \left\langle (\mathbf{n}_Q \cdot \mathbf{R}_i)^2 \right\rangle\right\} \tag{2.35}$$

$$= \sum_i \exp\left\{-Q^2 \frac{\langle R_i^2 \rangle}{3}\right\}, \tag{2.36}$$

with

$$\left\langle (\mathbf{n}_Q \cdot \mathbf{R}_i)^2 \right\rangle \overset{\text{(II)}}{=} \frac{\langle R_i^2 \rangle}{3}, \tag{2.37}$$

where $\langle R_i^2 \rangle$ describes the mean deviation of atom i from its equilibrium position or the time independent mean square displacement (MSD). The term MSD is very commonly used by the neutron scattering community. It is important to note that in fact the MSD is a time dependent variable which is defined as:

$$\text{MSD}_i(t) = \left\langle R_i^2(t) \right\rangle = \left\langle [R_i(0) - R_i(t)]^2 \right\rangle. \tag{2.38}$$

Since the atomic motions probed with EINS are confined, the MSD tends to a plateau value for $t \to \infty$ with

$$\mathrm{MSD}_i(t \to \infty) = \left\langle R_i^2(t \to \infty) \right\rangle = \left\langle [R_i(0) - R_i(\infty)]^2 \right\rangle = \left\langle R_i(0)^2 \right\rangle + \left\langle R_i(\infty)^2 \right\rangle = 2 \left\langle R_i^2 \right\rangle . \tag{2.39}$$

Therefore the pre-factor before $\langle R_i^2 \rangle$ is $1/3$ for the time independent (or also called static) MSD and $1/6$ for the time dependent (or also called dynamic) MSD [20]. The static MSD can also be called mean square position fluctuation (MSPF) to avoid confusion between the static and dynamic MSD.

Assuming an ideal instrument, with infinitely high resolution, $I(\mathbf{Q}, t \to \infty)$ equals the DSF at zero energy transfer $S(\mathbf{Q}, \omega = 0)$. This quantity is also called the elastic incoherent structure factor (EISF). To introduce all models, an ideal instrument is assumed. The influence and limitations of a non ideal instrument for neutron scattering experiments will be discussed in section 2.2.6.

2.2.2. The Gaussian approximation: GA model

The Gaussian Approximation (GA) model most commonly used in the literature for analysing data from EINS measurements is based on Eq. (2.36). There the first cumulant term is used to approximate the dynamics of **each** atom. For the GA model the cumulant expansion is applied over the **sum** of all atoms and one gets [7, 8]

$$S_{\mathrm{inc}}(Q, \omega = 0) \approx \exp\left\{ -Q^2 \frac{\langle R^2 \rangle_{\mathrm{GA}}}{3} \right\}, \tag{2.40}$$

where the MSD $\langle R^2 \rangle_{\mathrm{GA}}$ represents the average of the amplitudes of all atoms in the sample. This approximation always holds in case of $Q^2 \langle R^2 \rangle_{\mathrm{GA}} \ll 1$, since in this case the higher cumulant terms evaluate to zero.

This model is by far the most commonly used in publications about protein dynamics investigated with EINS [5] and implemented by fitting $\ln[S(Q)]$ vs Q^2. However, the choice of Q-range used to fit the data and the value of $\ln[S(Q)]$ at $Q^2 = 0\,\mathrm{Å}^{-2}$, is not always specified nor discussed in publications, neither is the often cited limit of $\langle r^2 \rangle_{\mathrm{GA}} Q^2_{\mathrm{max}} \leq 1$ justified [22]. Furthermore, this limit is sometimes surpassed and the usage of a larger Q^2_{max} is justified if the linear behaviour of the data points extends over a wider Q-range. However clear protocols are not well documented.

2.2.3. Models based on heterogeneity

As previously mentioned, the GA is generally only valid in a restricted region of Q, specifically at low Q values, since it neglects any effects from anharmonicity, heterogeneity or anisotropy [23]. Each of these effects alone or any combination can lead to a divergence from the Gaussian behaviour. As shown by Tokuhisa et al. [12] and Vural et al. [13], the heterogeneity of motions is the biggest contribution to the non Gaussianity of the MSD. To account for heterogeneity, all following models assume the validity of the GA for an

individual atom i and the non-Gaussianity arises only from a distribution of the individual MSDs. This assumption is a good approximation beyond low Q out to $Q \gtrsim 4\,\text{Å}^{-1}$ for almost all atoms according to Tokuhisa et al. [12] and Vural et al. [13], as the higher order terms are negligible compared to the Q^2 term. Moreover, Gaussian heterogeneity is able to describe approximately non-Gaussian (for instance rotational) motions, whose importance is well documented by Liu et al. [24]. Having said this, the GA might not be a valid approximation for an individual atom, due to large anharmonic or anisotropic effects. Kneller et al. [23] claim that models could be improved by accounting for anisotropy, but due to lack of precision of today's available data (neither experimental nor with simulations) additional parameters cannot be fit unambiguously.

In the following, two heterogeneity models are discussed which will be tested in this thesis as a comparison to the GA model which in contrast assume a distribution of purely Gaussian motions. Doster also addressed the issue of protein dynamical heterogeneity [25]. His main conclusion was that data can be fitted as successfully when heterogeneity is reduced to two kind of molecular processes, one corresponding to translational and the other to rotational motions (coming mainly from methyl groups), the latter being non-Gaussian scattering processes.

To include different distributions, a generic model distribution $\rho(\langle R_i^2 \rangle)$ can be introduced [10, 11, 26–28]:

$$S_{\text{inc}}(Q, \omega = 0) \approx \sum_i \exp\left\{ -Q^2 \frac{\langle {R_i}^2 \rangle}{3} \right\} = \int \mathrm{d}s \rho(s) \exp\left[-Q^2 s \right], \tag{2.41}$$

where $s = \langle {R_i}^2 \rangle / 3$. There have been several distributions used in publications. Nakagawa et al. [27] tested three different distributions for $\rho(s)$: bi-modal, Gaussian and exponential. For their data they concluded that the bi-modal distribution gave the best fit. Meinhold et al. [11] used a Weibull distribution which could well describe the results they obtained from neutron scattering and molecular dynamic (MD) simulations of thermophilic and mesophilic dihydrofolate reductase. Peters and Kneller [10] used a Gamma distribution and were able to fit elastic neutron scattering data of acetylcholinesterase. The first publication suggesting a Gamma distribution on simulation data by Kneller et al. [26] also introduced a shift of the peak of the distribution, but for experimental studies it could not be determined and therefore was not used due to the worse statistics. Very recently, Vural et al. [28] proposed a new distribution based on two separated Gaussian distributions. It differs from the other models since it is not describing $S_{\text{inc}}(Q, \omega = 0)$ but the ISF $I_{\text{inc}}(Q, t)$ in the time domain.

In this thesis the model by Peters and Kneller [10] (PK) is used as a representative model for heterogeneity described by a distribution and will be discussed now.

PK model: Peters and Kneller - Gamma distribution

Using Eq. (2.41) a Gamma distribution can be introduced for $\rho(s)$. In order to obtain dimensionless units, Q is substituted by $\hat{Q} = lQ$, where $l > 0$ is a scale variable with the

dimension of a length and s is substituted by $\lambda = s/l^2$. This results in [10]

$$S_{\text{inc}}(\hat{Q}, \omega = 0) \approx S_{\text{inc}}(\hat{Q}; \beta) = \int_0^\infty \mathrm{d}\lambda \rho(\lambda; \beta) \exp\left[-\hat{Q}^2 \lambda\right], \tag{2.42}$$

where $\rho(\lambda; \beta)$ is the Gamma distribution,

$$\rho(\lambda; \beta) = \frac{\beta \exp(-\beta\lambda)(\beta\lambda)^{\beta-1}}{\Gamma(\beta)} \qquad \text{for } 0 < \beta < \infty. \tag{2.43}$$

Here, Γ denotes the Gamma function [29]. The integral in Eq. (2.42) can be solved analytically and undoing the substitution from before, the following form for the elastic scattering function is retrieved

$$S_{\text{inc}}(Q; \beta) = \frac{1}{\left(1 + \frac{Q^2 \langle R^2 \rangle_{\text{PK}}}{3\beta}\right)^\beta} \tag{2.44}$$

$$\sigma_{\text{PK}} = \frac{\langle R^2 \rangle_{\text{PK}}}{\sqrt{\beta}}, \tag{2.45}$$

where $\langle R^2 \rangle_{\text{PK}}$ is the mean square position fluctuation (MSPF) of all hydrogen atoms and σ_{PK} its standard deviation. The term MSPF is used in [10] to make clear that the static (time independent) MSD is described. The GA can be retrieved in the limit of $\beta \to \infty$ and corresponds to perfect homogeneity of the atomic motions [10]

$$S_{\text{inc}}(Q; \beta) \overset{\beta \to \infty}{\to} \exp\left(-\frac{1}{3} Q^2 \langle R^2 \rangle_{\text{PK}}\right). \tag{2.46}$$

The PK model is valid over the entire momentum transfer range Q as long as the initial two conditions of the Gaussian approximation are valid, i.e. the displacement of the individual H atoms can be described by the Gaussian term alone and is isotropic.

Yi model: Yi et al - Q^4 extension as description of heterogeneity

The other heterogeneity model discussed in this thesis was proposed by Yi et al. [9] as a simple correction of the GA. It was first introduced by Becker et al. [30] in terms of heterogeneity, but similar expansions were already suggested earlier by Rahman [31] and Sköld et al. [32]. It does not assume any distribution but also recognises heterogeneity as the main reason for non-Gaussianity of $S_{\text{inc}}(Q, \omega = 0)$. In contrast to the methods described above it uses a Q^4 correction term, similar, but not equal, to the 4th cumulant term (second non-vanishing term, Eq. (2.34)). It assumes a derivation of the MSD $\langle R^2 \rangle_{\text{GA}}$

starting from Eq. (2.36) as follows

$$S_{\mathrm{inc}}(Q,\omega=0) \approx \sum_i \exp\left\{-Q^2\frac{\langle R_i^2\rangle}{3}\right\} \tag{2.47}$$

$$= \exp\left\{-Q^2\frac{\langle R^2\rangle_{\mathrm{Yi}}}{3}\right\}\left[\sum_i \exp\left\{-Q^2\frac{\langle R_i^2\rangle - \langle R^2\rangle_{\mathrm{Yi}}}{3}\right\}\right] \tag{2.48}$$

$$= \exp\left\{-Q^2\frac{\langle R^2\rangle_{\mathrm{Yi}}}{3}\right\}\left[\sum_{m=0}^{\infty}\frac{1}{m!}\left(\frac{-Q^2}{3}\right)^m \mu(m)\right] \tag{2.49}$$

$$\approx \exp\left\{-Q^2\frac{\langle R^2\rangle_{\mathrm{Yi}}}{3}\right\}\left(1+\frac{Q^4}{18}\sigma_{\mathrm{Yi}}^2\right), \tag{2.50}$$

where μ_m it the mth central moment of the distribution of $\langle R^2\rangle_{\mathrm{Yi}}$, and σ^2 is the variance, given by $(1/N)\sum_{i=1}^{N}(\langle R_i^2\rangle - \langle R^2\rangle_{\mathrm{Yi}})^2$. The last line is valid if $(-Q^2/3)^m\mu(m) \ll 1$ [9]. This expression is the same as the extension to the 4th cumulant for $\frac{Q^4}{18}\sigma_{\mathrm{Yi}}^2 \ll 1$. The introduction of Q^4 accounts for the heterogeneity of the motions and the spread of the MSD is described via σ_{Yi}, without assuming any functional form for the distribution. Since it does not contain higher order terms further than Q^4, its applicability is limited in Q.

2.2.4. Double well potential: Do model

The last model for EINS introduced in this thesis was published by Doster et al. [2] and in the same way as the PK model, it can be used to describe the entire Q range. It is based on the assumption that anharmonicity and not heterogeneity is the main reason for the non Gaussian terms in the description of EINS data. It is based on a double-well potential model to describe the anharmonicity of atomic motions. Each hydrogen atom can be found in one of two different harmonic wells which are separated by a distance d and by a free energy difference ΔG. There are two important points in comparison to the models described above. First, each well is harmonic and isotropic and therefore leads to the GA defined in Eq. (2.40). Second, there is no assumed heterogeneity of single atoms but each atom is described by anharmonicity. The model will be referred as the Do model from here on and is mathematically expressed as [2]:

$$S(Q,\omega=0) \approx \exp\left(-\frac{1}{3}Q^2\left\langle R^2\right\rangle_{\mathrm{Do},G}\right) \times (1-2p_{12}\,(1-\mathrm{sinc}\,(qd))), \tag{2.51}$$

where the first term with $\langle R^2\rangle_{\mathrm{Do},G}$ describes the Gaussian contribution to the MSD and the second term describes the two state model. p_{12} is the product of p_1 and p_2 which denote the probability of finding an atom in the ground or excited state, respectively, with

$p_2/p_1 \propto \exp(-\Delta G/RT)$. The total MSD is defined as [2]:

$$\frac{\langle R^2\rangle_{\mathrm{Do,tot}}}{3} = -\left(\frac{d\ln\left[S(Q,\omega=0)\right]}{d(Q^2)}\right)_{Q=0} = \frac{\langle R^2\rangle_{\mathrm{Do},G}}{3} + \frac{p_{12}d^2}{3}. \tag{2.52}$$

In the case of only one well, $p_{12} = 0$ or $d = 0$, such that the GA is retrieved.

2.2.5. Investigation of the summed intensities

An alternative way of analysing EINS data is through the evaluation of the elastic neutron intensities summed over all (or a range of) accessible Q-values, I_{sum}. In this way, an evaluation of the dynamics in a sample or a comparison between samples can be made without using a model. The summed intensities are much less affected by errors, so they give a more accurate overview over the dynamics taking place within the timescales probed by the spectrometer. Nevertheless, in the limit of the GA, it is possible to relate the I_{sum} to the inverse of the square of the MSD [33]:

$$I_{\mathrm{sum}} \overset{\mathrm{GA}}{=} \int_{Q'_{\mathrm{max}}}^{Q'_{\mathrm{min}}} \exp\left(-\frac{1}{3}\left\langle r^2\right\rangle' Q'^2\right) \mathrm{d}Q' \tag{2.53}$$

$$= \frac{1}{\sqrt{\frac{1}{3}\left\langle r^2\right\rangle'}} \int_{Q'_{\mathrm{max}}}^{Q'_{\mathrm{min}}} \exp\left(-z^2\right) \mathrm{d}z\,, \qquad \text{with} \quad z^2 = \frac{1}{3}\left\langle r^2\right\rangle' Q'^2 \tag{2.54}$$

$$= \frac{1}{\sqrt{\frac{1}{3}\left\langle r^2\right\rangle'}} \frac{\sqrt{\pi}}{2}\left[\mathrm{erf}\left(Q'_{\mathrm{max}}\right) - \mathrm{erf}\left(Q'_{\mathrm{min}}\right)\right] \tag{2.55}$$

$$= \frac{1}{\sqrt{\left\langle r^2\right\rangle'}} \cdot C\,, \qquad \text{with} \quad C = \frac{\sqrt{3\pi}}{2}\left[\mathrm{erf}\left(Q'_{\mathrm{max}}\right) - \mathrm{erf}\left(Q'_{\mathrm{min}}\right)\right] \tag{2.56}$$

$$\Rightarrow \left\langle r^2\right\rangle' \overset{\mathrm{GA}}{=} \frac{1}{I_{\mathrm{sum}}^2} \cdot C^2\,, \tag{2.57}$$

where $Q' = Q/l$ and $\left\langle r^2\right\rangle' = \left\langle r^2\right\rangle / l^2$ are dimensionless quantities and l is chosen as a typical length scale (1 Å) and erf is the error function [29]. In the limit of experimental precision the I_{sum} is discrete and can be evaluated for each measured temperature T

through:

$$I_{\mathrm{sum}}(T) = \sum_{i=Q_{\mathrm{min}}}^{Q_{\mathrm{max}}} I_i(T) \tag{2.58}$$

$$\left\langle r^2 \right\rangle_{\mathrm{SumI}}(T) \overset{\mathrm{GA}}{\propto} \frac{1}{I^2_{\mathrm{sum}}(T)} \ \text{Å}^2 \tag{2.59}$$

where Q_{min} and Q_{max} are defined either by the validity of the GA or given by the instrument.

2.2.6. The influence of the instrumental resolution

To evaluate the effects of the instrumental resolution, it is convenient to split the ISF into a time-dependent ($I(Q,t)$) and a time-independent ($I(Q,\infty)$) component:

$$I_{\mathrm{inc}}(Q,t) = I(Q,\infty) + [I(Q,t) - I(Q,\infty)] \tag{2.60}$$

$$= I(Q,\infty) + [1 - I(Q,\infty)]\, C(Q,t) \tag{2.61}$$

where $C(Q,t)$ is a relaxation function fulfilling

$$C(Q,0) = 1 \tag{2.62}$$

$$\lim_{t\to\infty} C(Q,t) = 0 \tag{2.63}$$

and therefore describes the two limits at $t = 0$ and $t \to \infty$ for the ISF $I_{\mathrm{inc}}(Q,t)$. In general, the term $[1 - I(Q,\infty)]\, C(Q,t)$ can be attributed to the quasi elastic neutron scattering (QENS) signal after the time Fourier transformation,

$$S_{\mathrm{inc}}(Q,\omega) = I(Q,\infty)\delta(\omega) + S_{\mathrm{QENS}}(Q,\omega), \tag{2.64}$$

where $\delta(\omega)$ is the delta function.

In experiments $I(Q,\infty)$ can not be measured directly since Eq. (2.64) is broadened by the instrument resolution $R(\omega)$, leading to a convolution ($*$) in the energy domain and a multiplication with $R(t)$ in the time domain:

$$S^{\mathrm{R}}_{\mathrm{inc}}(Q,\omega) = S_{\mathrm{inc}}(Q,\omega) * R(\omega) \tag{2.65}$$

$$S^{\mathrm{R}}_{\mathrm{inc}}(Q,\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} I(Q,\infty)\exp(i\omega t) \times R(t) \tag{2.66}$$

$$+ \frac{1}{2\pi}\int_{-\infty}^{\infty} [1 - I(Q,\infty)]\, C(Q,t)\exp(i\omega t) \times R(t)\mathrm{d}t. \tag{2.67}$$

In the case of an ideal instrument, $\Delta\omega = 0$ (infinitely high instrument resolution), we would obtain the EINS signal at zero energy transfer directly, which is also called elastic incoherent structure factor (EISF) such that:

$$S^{\mathrm{R}}_{\mathrm{inc}}(Q,\omega)(Q,\omega=0,\Delta\omega=0) = S_{\mathrm{inc}}(Q,\omega=0) = I(Q,\infty) \equiv \mathrm{EISF}(Q). \tag{2.68}$$

Due to the limitation of the instrumental resolution, $\Delta\omega_{\mathrm{res}} \neq 0$ and thus the accessible observable time scale is not infinity, but approximately equal to $\tau_{\mathrm{res}} = 1/\Delta\omega_{\mathrm{res}}$. If the QENS spectra at $(\omega \approx 0, \Delta\omega_{\mathrm{res}})$ is neglected for the given instrument resolution, one gets the resolution dependent EISF

$$\mathrm{EISF_R}(Q) \equiv \frac{S^{\mathrm{R}}_{\mathrm{inc}}(Q, \omega \approx 0, \Delta\omega_{\mathrm{res}})}{S^{\mathrm{R}}_{\mathrm{inc}}(Q = 0, \omega \approx 0, \Delta\omega_{\mathrm{res}})} = \frac{\frac{1}{2\pi}\int_{-\infty}^{\infty} I(Q,t) \times R(t, \tau_{\mathrm{res}})\mathrm{d}t}{\frac{1}{2\pi}\int_{-\infty}^{\infty} R(t, \tau_{\mathrm{res}})\mathrm{d}t} \tag{2.69}$$

$$\approx I(Q, \tau_{\mathrm{res}}), \tag{2.70}$$

where $R(t, \tau_{\mathrm{res}})$ is truncating the integral over $I(Q,t)$ at $\pm\tau_{\mathrm{res}}$. With the definition of $I(Q, \infty)$ it can be shown that the elastic fraction evaluated in EINS experiments, $\mathrm{EISF_R}$ is $\geq$ EISF. The $\mathrm{EISF_R}$ is larger than EISF if hydrogen atoms appear localised in the range of the instrument resolution τ_{res} and therefore even unconstrained motions can be apparent in the elastic contribution [20]. Thus the instrumental resolution leads to a MSD $\langle R_i^2 \rangle$ which depends on the instrumental time window defined by $\tau_{\mathrm{res}} = 1/\Delta\omega_{\mathrm{res}}$.

3. Instrumentation

In this chapter the neutron instrumentation used for the experiments in this thesis are discussed. In the first section, the production of neutrons using a reactor or spallation source is explained. The second section provides further details about neutron instrument components which are important to perform EINS measurements. Section 3.3 starts with a general overview on neutron spectrometers, followed by the last section with details of the specific spectrometers used in this thesis, backscattering spectrometers. The neutron production and neutron optics sections are inspired by the overview article of H. Schober [34].

3.1. Neutron production: Reactor and spallation sources

Neutrons are omnipresent in the universe but they are almost always in a bound state since the free mean lifetime is only around 15 min (see Table 2.1). For this reason they have to be produced and used before they decay to a proton, an electron and an anti-neutrino. To date there are only two ways of producing large fluxes of free neutrons for experiments, in a reactor and a spallation source. For both, two important steps are required for further use of neutrons in neutron scattering experiments. First, the neutrons need to be extracted from the nucleus and second, the neutron need to be moderated to the appropriate energies. Moderation is a process in which the high energy neutrons are slowed down, normally by scattering neutrons off other atoms or molecules (using moderators). This is necessary due to the high energy of the neutrons produced by nuclear reactions during the extraction process (several MeV).

In case of fission (reactor based source), metastable uranium ^{235}U is used to produce free neutrons. It can be excited by absorbing a slow neutron, ^{236}U*, which then decays in a cascade to several lighter atoms, the so called fission products. During the fission of one ^{236}U* nucleus, 2-3 free neutrons are produced (2.5 in average) which have an energy of around 1 MeV. They cannot be used to induce further fission of ^{235}U and therefore have to be moderated to meV energies. Light atoms are preferred as moderators since they uptake a high amount of the neutron energy during each collision. Often either H_2O or D_2O is used, where mainly the mass of the H or the D atoms accounts for the moderation efficiency and only about 18 (H_2O) or 25 (D_2O) collisions are needed to obtain meV energies [34]. The moderated free neutrons can then induce further fission or can be extracted from the reactor core for scientific use. To maintain a stable reactor power, about one free neutron is needed to fuel the fission and the other 1-2 neutrons of each fission can be extracted for scientific purposes. To allow for an efficient free neutron beam extraction, the moderator

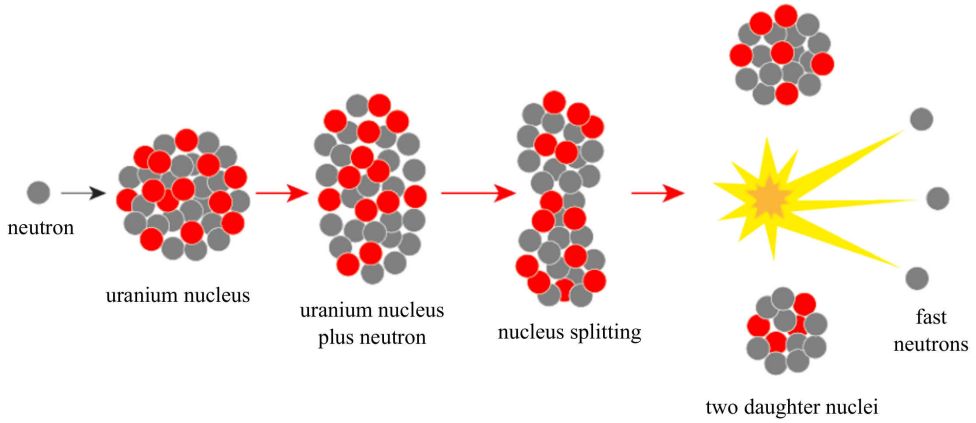


Figure 3.1. - Fission process. Figure taken from [35].

should have a large thermal diffusion length before the neutrons are absorbed. H_2O has only a small thermal diffusion length of about 3 cm, whereas for D_2O it is 1 m. Thus for a very high flux reactor source like that at the Institut Laue-Langevin (ILL), heavy water is used as the main moderator.

In a spallation source, the free neutrons are produced by hitting a heavy target like e.g. uranium, tungsten, lead or mercury, with high energy protons. The reaction in the target nucleus is a two-stage process. In stage one, the primary proton hits a nucleon in the target nucleus, which can either hit another nucleon of the same nucleus (intra-nuclear cascade) or a different nucleus (internucleus cascade). After being hit, the nuclei are highly excited and release their energy mainly by evaporation of neutrons and other particles like protons, deuterons, α-particles, heavier fragments as well as also γ-radiation [36]. Roughly 20 free high energy neutrons (mainly around 2 MeV) are produced for each accelerated proton [34]. As with fission, these neutrons are then moderated. The spallation process is about one order of magnitude more efficient than fission, resulting in roughly the same number of neutrons for a 5 MW particle beam and a fission reactor at 50 MW thermal power. However, the creation of a 5 MW particle beam needs a considerable amount of electrical power, which is in the same order of magnitude as the reactor power [34].

The characteristics of a reactor based source and a spallation source are different. A reactor produces neutrons at a constant rate and thus the flux of neutrons has no explicit time structure and is also called continuous or steady state neutron source. Typical examples are the ILL reactor in Grenoble (France), the the High-Flux Isotope Reactor (HFIR) at Oak Ridge National Laboratory (ORNL) in Oak Ridge, (Tennessee, USA) or the FRM-II reactor in Munich (Germany). In case of the ILL, the unperturbed neutron flux is around 1.5×10^{15} neutrons $s^{-1}cm^{-2}$. The Swiss Spallation Neutron Source (SINQ) at the Paul Scherrer Institut (PSI) in Villingen (Switzerland) is also a continuous neutron source but based on spallation. The flux is about 10^{14} neutrons $s^{-1}cm^{-2}$. [37] In contrast, spallation sources generally work with pulsed neutron beams and are thus called pulsed spallation sources. Typical examples are the ISIS Pulsed Neutron and Muon Source in Didcot (UK), the Spallation Neutron Soucre (SNS) at ORNL in Oak Ridge (Tennessee, USA), the Japanese Spallation Neutron Source (J-SNS) at the Materials and Life science experimental Facility (MLF) in the Japan Proton Accelerator Research Complex (J-PARC)

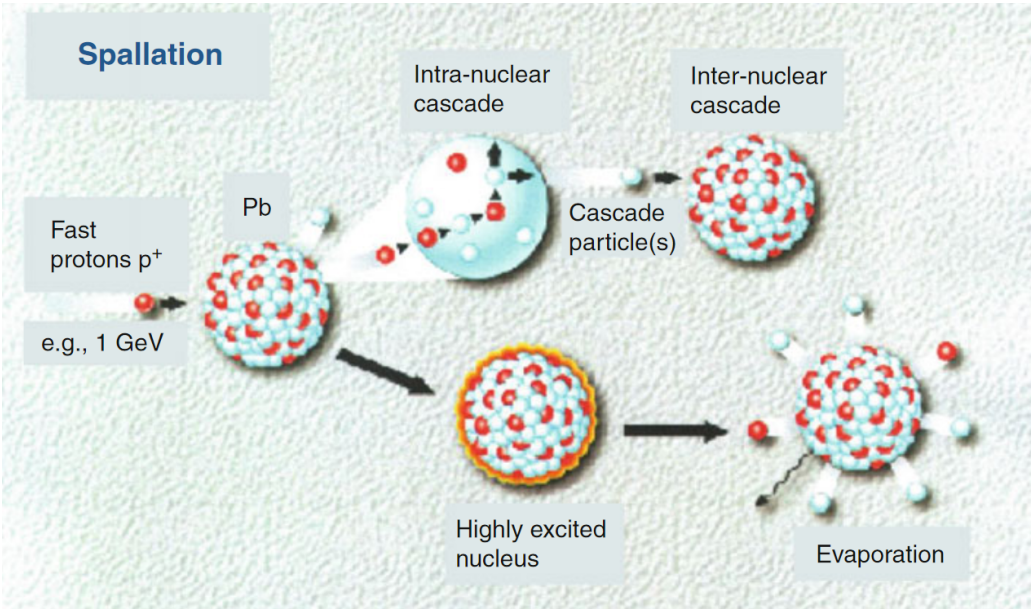


Figure 3.2. - Spallation process. Figure taken from [36].

in Tokai-mura (Japan). In addition, construction is ongoing for the future European Spallation Source (ESS) in Lund (Sweden). The operation at full performance is scheduled for 2025. [38] To achieve the pulsing, protons are bunched in a linear (LINAC) or circular (synchrotron) accelerator and then guided onto the target with high energy ($\sim$ 0.5-3 GeV). Typically, the target is hit with a repetition rate between 10 to 60 Hz and the current of protons delivered on the target is measured in µA (micro-Amps) which is proportional to the produced neutron flux. As an example the ISIS facility has a synchrotron accelerator and protons with an energy of 800 MeV are sent to two different targets; Target Station (TS) 1 and TS 2. TS 1 receives protons at 40 Hz (160 µA) and TS 1 at 10 Hz (40 µA). A summary of the main characteristics of some existing sources are given in Table 3.1.

3.2. Neutron instrument components

In this section three important components employed to manipulate neutrons are presented since they are relevant to the instruments used in this thesis. These components are a small selection amongst those used for many other neutron instruments.

3.2.1. Neutron guides

In order to guide the neutrons from the source to the instruments special neutron guides are needed. The problem is that neutrons are not easily deflected from surfaces and they cannot be guided by electro-magnetic fields as they are electrically neutral. Therefore the guides are coated with a special substance, e.g. Ni and Ti, where neutrons are totally reflected under certain angles. Similar to light scattering, neutrons can have a specular (mirror-like) reflection and can be described via the index of refraction n. For a specific

Table 3.1. - Neutron sources: characteristics. Values taken from [35, 36, 38]. ESS values are the full performance values foreseen in 2025.

	source	beam type	average power [MW]	proton pulse duration [µs]	proton pulse rate [Hz]	average thermal neutron flux [n cm^{-2}s^{-1}]
Fission	ILL	continuous	57	-	-	1.5×10^{15}
	HFIR	continuous	85	-	-	$\sim 1 \times 10^{15}$
	FRM-II	continuous	20	-	-	0.8×10^{15}
Spallation	SINQ	continuous	1.2	-	-	1.0×10^{14}
	SNS	pulsed	2	0.69	60	1.2×10^{14}
	ISIS	pulsed	0.16	0.39	50	0.7×10^{13}
	J-SNS	pulsed	1	1.2	25	1.3×10^{14}
	ESS	pulsed	5	2860	14	-

medium, it is defined as the ratio of the modulus of the neutron wave vector k_m in the medium versus its value k_0 in vacuum. This fraction can be linked to the Fermi pseudo-potential defined in Eq. (2.13) and evaluates to:

$$n = \frac{k_m}{k_0} = 1 - \frac{\lambda}{2\pi} \sum_i N_i b_i, \tag{3.1}$$

where N_i is the number density of scatterers of type i and the quantity $\sum_i N_i b_i$ is called scattering length density (SLD) of the material[34]. For typical values in neutron scattering, e.g. $\lambda = 1\,\text{Å}$, $b = 10^{-15}\,\text{m}$, $N \approx 10^{29}\,\text{m}^{-3}$, the second term evaluates to $\approx 10^{-7}$. Thus n is only slightly smaller than $n_0 = 1$ in vacuum, leading to a very small angle for total external reflection. This angle is called critical or glancing angle θ_c and states that the neutron is totally reflected if it *hits* the boundary from vacuum to material at an angle $\theta < \theta_c$. It can be calculated with[34]

$$\sin\theta_c = \sqrt{\frac{\sum_i N_i b_i}{\pi}}\lambda. \tag{3.2}$$

Since Ni has a very large coherent scattering cross section it is used for neutron guides and its critical angle is $\theta_c = 0.1° \times \lambda[\text{Å}]$. This angle can be enhanced by multi-layering of materials and nowadays so-called supermirrors are used as neutron guides which are made up of Ni and Ti. These guides cannot only be used as neutron transportation system but also to shape the neutron beam, i.e. focusing it by bending the mirrors[34].

3.2.2. Monochromators

In general, the neutrons arriving at the instrument are polychromatic, i.e. consist of many different wavelengths. In order to have monochromatic beams for the experiments the

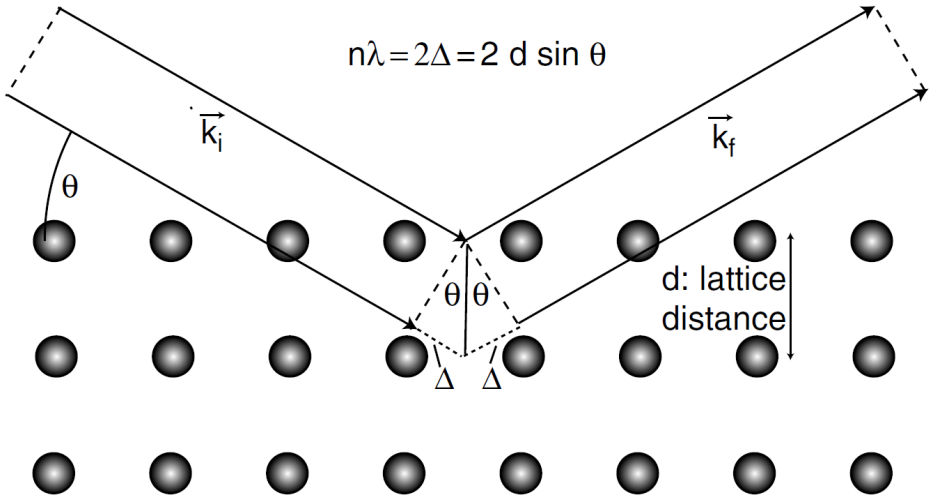


Figure 3.3. - Illustration of Bragg's law. The incoming neutron beam $\mathbf{k}_i$ is scattered to $\mathbf{k}_f$ by a regular lattice obeying Bragg's law, i.e. the distance 2Δ has to be a multiple n of the wavelength $\lambda = 2\pi/k$. Therefore the outgoing wavelength is dependent of the bragg angle θ and the lattice spacing d. Figure taken from [34].

initial beam has to be filtered. This can be done one of two ways: selection of a single wavelength via diffraction using a monochromator, or by time-of-flight method which will be described in the next section.

Diffraction of a neutron beam can be described by Bragg's law

$$n\lambda = 2d\sin\theta, \tag{3.3}$$

where θ is the Bragg angle, n the order of the Bragg reflection and d the lattice spacing in reciprocal space. It is based on positive interference from planes if the optical path difference between the incoming wave vector $\mathbf{k}_i$ and outgoing wave vector $\mathbf{k}_f$, is a multiple n of the wavelength λ. A schematic is shown in Figure 3.3. This allows to select a specific wavelength by changing the angle θ and the lattice spacing d. The resulting monochromatic beam has a finite width $\Delta\lambda$ and is dependent on the divergence of the incoming beam, $\Delta\theta$, and on the uncertainties in the lattice spacing, Δd,

$$\Delta\lambda = 2d\cos\theta\Delta\theta + 2\sin\Delta d = (\lambda\cot\theta)\Delta\theta + \frac{\lambda}{d}\Delta d. \tag{3.4}$$

In terms of the width relative to the incoming λ, the relative wavelength resolution can be calculated as

$$\frac{\Delta\lambda}{\lambda} = \cot\theta\Delta\theta + \frac{\Delta d}{d}. \tag{3.5}$$

Thus two parameters can be changed to achieve a better resolution. $\Delta\theta$ can be decreased by using a perfect single crystal instead of mosaic crystals (numerous perfect crystals which are not aligned) or by reducing the incoming divergence of the beam. Both methods have the disadvantage that the neutron flux will be decreased significantly. A better solution

is to be in perfect backscattering condition, $\theta = 90°$, resulting in $\cot\theta = 0$. Then the divergence of the beam can be neglected and it can be increased to obtain a better neutron flux. The alternative possibility is to decrease $\Delta d/d$. In most cases mosaic crystals are used as a compromise between high resolution and a good neutron flux. The notation for monochromator crystals is normally the chemcial material followed by the reflection used in the reciprocal lattice, e.g. Si 111, Si 311 or CaF_2 422.

3.2.3. Time-of-Flight (ToF) method and neutron choppers

The de Broglie relation (see Eq. 2.2) states that the neutron wavelength is inversely proportional to the speed of the neutron. This can be used to monochromate a white neutron beam. The basic principle is simple: to measure the wavelength the time needed for a neutron to travel a specific length has to be known. For this a starting point in time and space has to be defined for the neutrons, i.e. the neutron beam has to be pulsed and a time t_0 to be defined. This is either intrinsic to the source (spallation source) or in case of a continuous source (reactor) it is done by using so-called neutron choppers, rotating disks. If the rotating axis is parallel to the beam they are called disk choppers and if the rotating axis is perpendicular they are called fermi choppers. In case of a disk chopper, the rotating disk has one or multiple slits which let the neutrons pass and the rest of the disk is made of neutron absorbing material. Fermi choppers consists of a rotating collimator made of absorbing sheets alternating with transparent spacers.

After the pulsing of the beam the wavelength can be related to the travel time t over a distance L using Eq. 2.2

$$t = \frac{m}{h} L\lambda \tag{3.6}$$

and the resolution can be defined with the time width τ of the pulsed beam via

$$\frac{\Delta\lambda}{\lambda} = \frac{\tau}{t} \tag{3.7}$$

This can be used to produce a monochromatic beam by only selecting the wavelength of interest via a second disk chopper. Most instruments often have more than two choppers to define the neutron beams with higher precision. The advantage of the ToF method over using a monochromator is that the neutron wavelength is not correlated to a specific angle. ToF is not only a tool to monochromate a neutron beam but is also used to measure the wavelength of a scattered neutron after it interacts with a sample during an experiment. This is a very important technique to measure the energy gain or loss of the neutron after the interaction.

3.3. Neutron Spectrometers

In general, in a neutron spectrometer three important quantities are measured, the initial neutron energy E_i (= wavelength λ_i), the final neutron energy after the interaction with

the sample $E_\mathrm{f}(\lambda_\mathrm{f})$ and magnitude of the momentum transfer Q. Thus, the DSF, $S(Q,\omega)$, is measured directly. The momentum transfer can be calculated starting from Eq. (2.9) as

$$|\mathbf{Q}| = Q = |\mathbf{k}_\mathrm{i} - \mathbf{k}_\mathrm{f}| = \sqrt{k_\mathrm{i}^2 + k_\mathrm{f}^2 - 2k_\mathrm{i}k_\mathrm{f}\cos\alpha}, \tag{3.8}$$

where α is defined as $\mathbf{k}_\mathrm{i} \cdot \mathbf{k}_\mathrm{f} = k_\mathrm{i}k_\mathrm{f}\cos\alpha$ and is given in the instrument by the position of the detector (see also Figure 2.1). In the case of elastic scattering ($k_\mathrm{i} = k_\mathrm{f}$), Q can then be readily calculated by

$$Q_\mathrm{EINS} = \sqrt{2k_\mathrm{i}^2 - 2k_\mathrm{i}^2\cos\alpha} = \sqrt{2k_\mathrm{i}^2(1-\cos\alpha)} = \sqrt{4k_\mathrm{i}^2\sin^2\left(\frac{\alpha}{2}\right)} = \frac{4\pi}{\lambda_i}\sin\left(\frac{\alpha}{2}\right), \tag{3.9}$$

where for the third step the trigonometric expression

$$(1-\cos\alpha) = 2\sin^2\left(\frac{\alpha}{2}\right) \tag{3.10}$$

was used. For inelastic scattering, $k_\mathrm{i} \neq k_\mathrm{f}$, and Q^2 can be calculated by using Eq. (2.8) in Eq. (3.8)

$$Q^2 = 2k_\mathrm{i}^2 - \frac{2m\omega}{\hbar} - 2k_\mathrm{i}\sqrt{k_\mathrm{i}^2 - \frac{2m\omega}{\hbar}}\cos\alpha. \tag{3.11}$$

Thus the magnitude of the momentum transfer can be calculated if the initial and final neutron energy (wavelength) is known.

Another important quantity of a spectrometer is its finite energy resolution $\hbar\Delta\omega_\mathrm{res}$. It is given by the accuracy of the measured neutron wavelength and thus a combination of the width of the initial and final neutron energy. In general, the resolution of an instrument is expressed as the Full-Width at the Half-Maximum (FWHM) of its resolution function $R(Q,\omega)$. It is important for two different reasons. First, it defines the maximum time of observation, Δt_res, and secondly, it broadens the experimental DSF, $S^\mathrm{R}(Q,\omega)$, around $\omega = 0$. In order to obtain the theoretical DSF $S(Q,\omega)$, which can then be analysed by models, the $S^\mathrm{R}(Q,\omega)$ has to be corrected by $R(Q,\omega)$. In other words, the $S^\mathrm{R}(Q,\omega)$ is a convolution ($*$) between $R(Q,\omega)$ and $S(Q,\omega)$:

$$S^\mathrm{R}(Q,\omega) = S(Q,\omega) * R_\mathrm{res}(Q,\omega) \tag{3.12}$$

For an ideal instrument the resolution is a delta function at $\omega = 0$: $R(Q,\omega) = \delta(\omega)$ and in this case $S^\mathrm{R}(Q,\omega) = S(Q,\omega)$. For QENS measurements the $R(Q,\omega)$ has to be measured or analytically described to obtain $S(Q,\omega)$. In EINS measurements the **shape** of the resolution function is normally neglected since only the value of the DSF around zero energy transfer is used. The important value is the FWHM of $R(Q,\omega)$ since it defines the **width** of the resolution function $\Delta\omega_\mathrm{res}$ which is inversely proportional to the observed timescale, Δt_res, of the experiment. They are related via Heisenberg's uncertainty principle

$$\Delta t_\mathrm{res} \cdot \Delta\omega_\mathrm{res} \geq \hbar. \tag{3.13}$$

The connection between instrumental resolution, the EISF ($S(Q, \omega = 0)$) and the evaluated MSD was already discussed in section 2.2.6.

3.4. Backscattering spectrometers

All experiments in this thesis were performed on so called backscattering (BS) instruments. In section 3.2.2 using Eq. (3.5) it was shown that a monochromator has its highest resolution if it is used in the backscattering position, $\theta = 90°$. This feature is used in backscattering spectrometers to define the wavelength of the neutrons very accurately, i.e. obtain high energy resolution (typically in the order of several µeV).

Owing to the different nature of neutron sources, backscattering (BS) instrumentation is slightly different at each source. In the following, the BS instrumentation is differentiated between Backscattering spectrometers (BSS) and Time-of-Flight backscattering spectrometers (ToF-BSS). In order to compare them, each instrument is split into two parts: the primary and secondary spectrometer. The primary spectrometer defines the incoming neutron beam before the interaction with the sample and the secondary spectrometer analyses the neutrons after the scattering event [39].

BSS use the backscattering condition twice in one measurement: a BS crystal is used as the monochromator (primary spectrometer) and another BS crystal as the analyzer (secondary spectrometer). The measured energy transfer of the neutron is therefore fixed by the properties of the monochromator and the analyzer. In order to observe different energy transfers, either the monochromator or analyzer properties have to be changed. Since the BS geometry needs to be maintained, the scattering angle cannot be changed and therefore two alternative methods are used. Either the crystal properties are changed, typically by changing the lattice spacing d thermally, or a Doppler drive is used. The Doppler drive is a crystal moving parallel to the neutron trajectory, leading to a Doppler broadening which changes the wavelength of the incoming neutron and therefore its energy [6].

For a ToF-BSS the secondary spectrometer is the same as for the BSS, namely a monochromator is used as the analyzer. In contrast, the primary spectrometer uses the ToF technique (see section 3.2.3) to define the incoming neutron beam. Importantly the final energy is defined by the analyzer and the incoming beam is not monochromatic. The ToF is used to analyze the incoming neutron energy by measuring the time between the incoming beam and the detection of the neutron after being analyzed. These kind of instruments are also called indirect (or inverted) geometry spectrometers (does not have to be in perfect BS).

A sketch of the beam path for two BSS, IN13 at the ILL and SPHERES at MLZ, and for two ToF-BSS, OSIRIS and IRIS at ISIS, are shown in Figure 3.4. In the following they will be described in more detail.

3.4.1. IN13 at the ILL

IN13 [40] is a BSS based in the Institut Laue Langevin (ILL), Grenoble, France. It has a uniquely large momentum transfer for a typical BSS instrument with a Q-range of

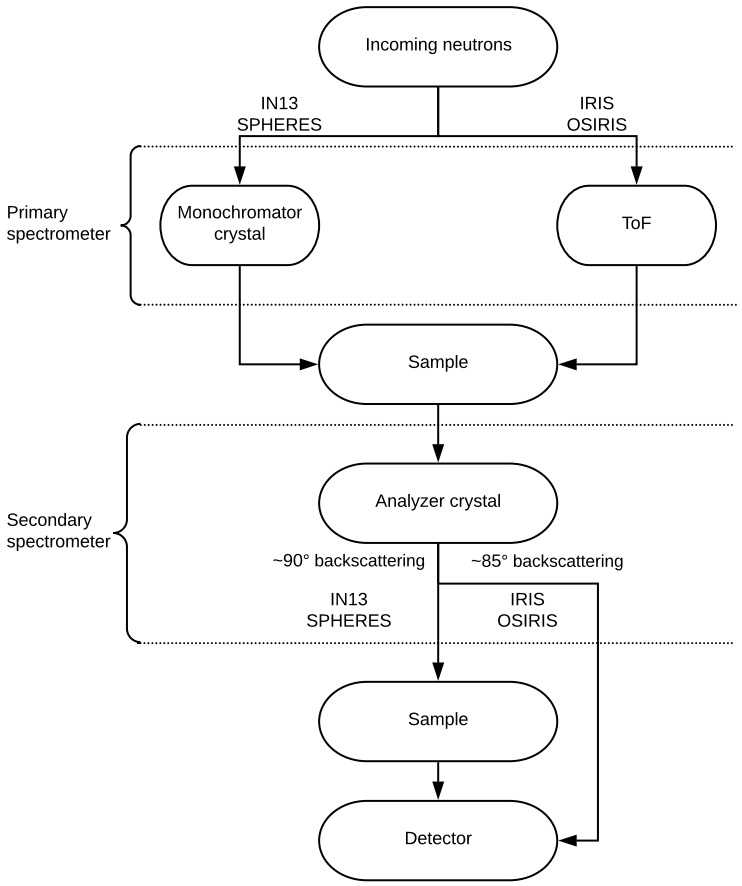


Figure 3.4. - Neutron beam path in backscattering geometry.

$0.2 - 4.9\,\text{Å}^{-1}$. All important instrument characteristics are shown in Table 3.2. The general layout is that of a BSS as described in the previous section. The instrument layout is shown in Figure 3.5 and consists of the following main components:

1. a monochromator CaF_2 422 provides neutrons of incident wavelength 2.23 Å from the reactor beam guide in almost BS direction.
   - the monochromator can be heated and cooled to achieve a dynamic range of -125 to $150\,\mu\text{eV}$
2. a graphite deflector focuses the neutron beam on the sample
3. a chopper cuts the beam into pulses for ToF
   - filters out primary scattered neutrons which were not analyzed by the analyzer crystals and reduces the background (detectors are only counting in a certain time frame given by the travel time of the neutrons between the sample and the analyzer)
4. neutrons are interacting with the sample
5. only neutrons with a wavelength of 2.23 Å are backscattered ($2\theta = 180°$) by the analyzer crystals (CaF_2 422)
6. neutrons pass the sample a second time and are then registered in the detectors.

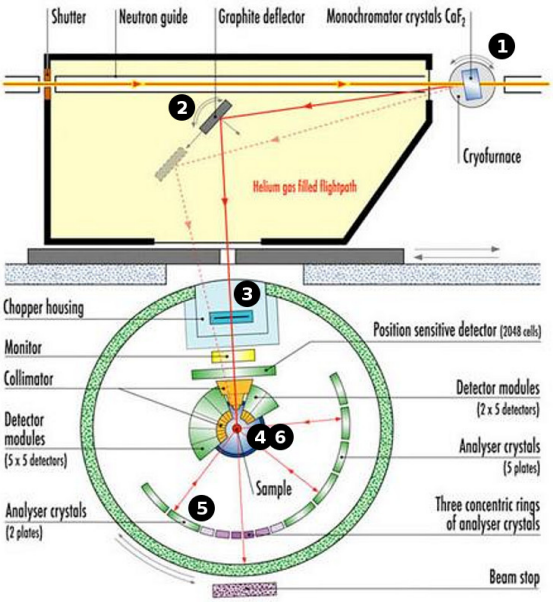


Figure 3.5. - IN13, Instrument layout. For the beam path description see section 3.4.1. Figure taken from [41] and modified.

Owing to its medium flux, IN13 is mainly used for EINS measurements although it is also possible to perform QENS measurements by changing the temperature of the monochromator, but this is much more time consuming. As previously mentioned, neutron pass the sample twice which can lead to a second interaction with the sample, either a scattering event or absorption.

3.4.2. SPHERES at the MLZ

SPHERES [42, 43] is a BSS based in the Heinz Meier-Leibnitz Zentrum (MLZ) in Munich, Germany (FRM II reactor source) and is operated by the Forschungszentrum Jülich. The instrument is similar to IN13 but has three important differences. First, instead of cooling/heating the monochromator to achieve the measurement of inelastic energy transfers, a Doppler drive is used. By changing the speed of the Doppler drive the energy of the incoming neutrons can be slightly changed. Thus the incoming energy of the neutrons can be changed in a very quick way in comparison to IN13. Secondly, a phase space transformation chopper (PST) is used to deflect the incoming neutrons from the monochromator. It accelerates and decelerates the neutrons around the monochromator BS wavelength (6.27 Å) to increase the neutron flux at the cost of increasing divergence. Thirdly, it uses a Si 111 analyser, like the monochromator. The rest of the beam path is equal to IN13. It is a more traditional BSS. The instrument characteristics are shown in Table 3.2 and the layout in Figure 3.6. Additionally, it is important to mention that the first four detectors ($Q = 0.2 - 0.44\,\text{Å}^{-1}$) are not in perfect backscattering condition and suffer from a lower resolution [44]. Their elastic peak is also not perfectly centered around the zero energy transfer. Similar instruments are the IN16B [45] at the ILL and the HFBS [46] at the NIST Center for Neutron Research in Gaithersburg, MD, USA. Both

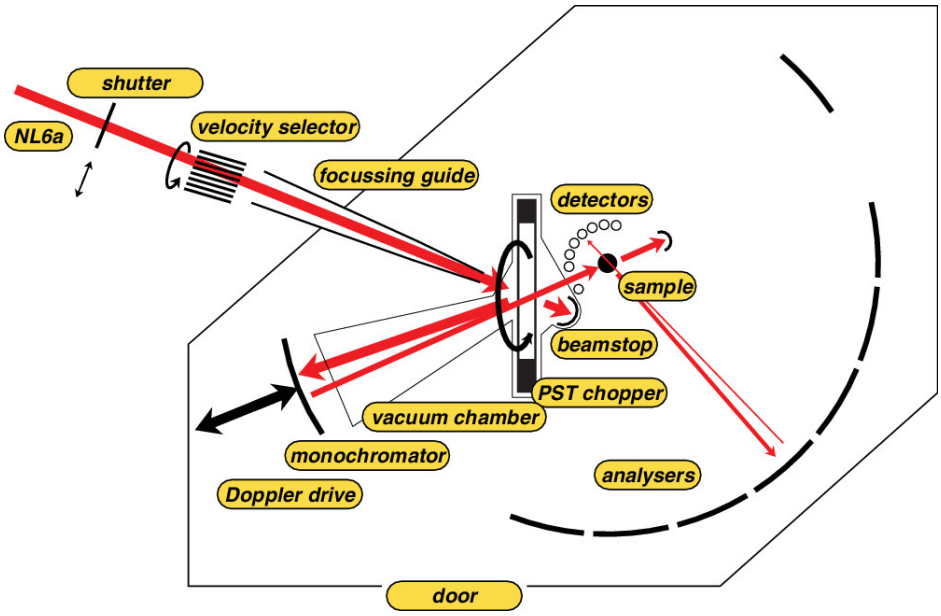


Figure 3.6. - SPHERES, Instrument layout. The beam path is similar to the IN13 spectrometer. Additionally, it has a Doppler drive and a Phase space transformation chopper (PST). Figure taken from [42].

also have a PST and equally suffer from slightly lower resolution at low Q values.

3.4.3. IRIS and OSIRIS at ISIS

Both IRIS [47] and OSIRIS [48] are ToF-BSS based at ISIS Facility in Didcot, UK (spallation source) but their instrumental layouts are very similar (see Figure 3.7 and 3.8) and their specifications are slightly different (see Table 3.2). OSIRIS is the newer instrument and consequently has a higher neutron flux at the sample ($\times 2.7$) and a much higher counting rate ($\times 4.5$ in PG 002 and $\times 11$ in PG 004) due to a larger analyser bank. The incoming beam is defined by two disc-choppers, which allows a high resolution measurement of the incoming neutron energy via ToF: the neutrons are counted in the detector with the time delay in comparison to the incoming beam. Since the final energy is fixed by BS from the analyzer (PG 002 or PG 004) the incoming neutron energy can be calculated. At IRIS and OSIRIS the analyzer is not in perfect BS condition ($2\theta \approx 170°$), leading to a lower resolution in comparison to IN13 and SPHERES. Therefore they are also called near-BSS. In contrast, the neutrons only pass once through the sample. Another advantage is that the energy resolution is the same for all detector angles. Additionally, both ToF-BSS have a diffraction detector bank allowing for simultaneous diffraction measurements which prove useful for many samples. A similar ToF-BSS is BASIS [49] based at the SNS Oak Ridge National Laboratory, that uses Si 111 and a longer primary flight path to obtain a higher resolution of $3-3.5\,\mu eV$ [49].

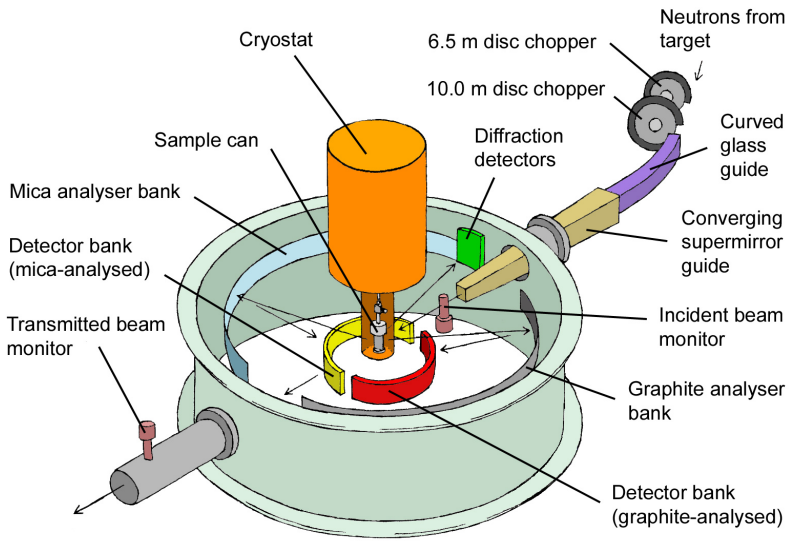


Figure 3.7. - IRIS, Instrument layout. Figure taken from [50].

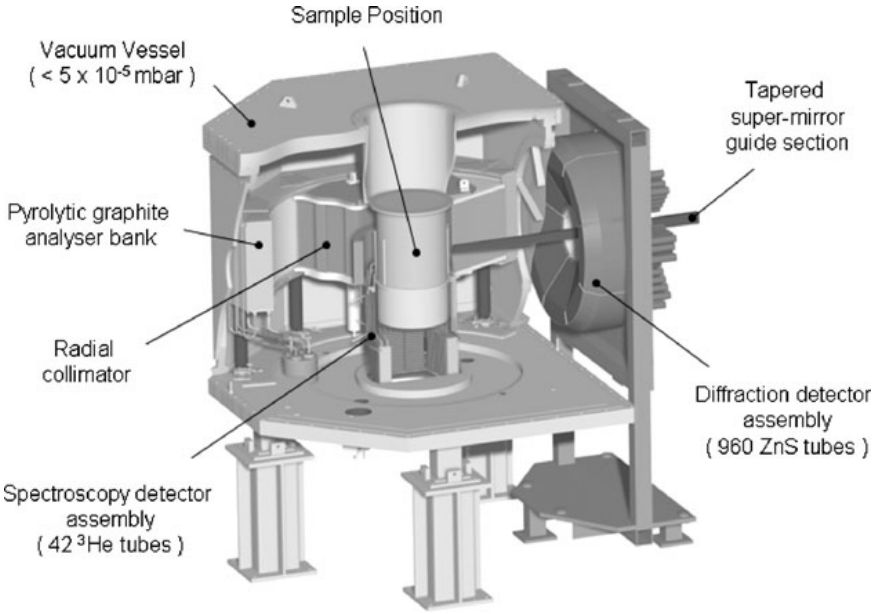


Figure 3.8. - OSIRIS, Intrument layout. Figure taken from [48].

Table 3.2. - Instrument specification of IN13, SPHERES, IRIS and OSIRIS. All specification are given for the case of elastic scattering. $\Delta\lambda_i$ describes how the incoming wavelength is changed and thus the dynamical energy transfer ΔE is achieved. ΔE_{res} is the energy resolution of the respective instrument. The observed timescale Δt_{res} was calculated via the Heisenberg's uncertainty principle (see Eq. (3.13)).

Spectrometer	IN13	SPHERES	IRIS*		OSIRIS	
type	BSS		ToF-BSS			
$\Delta\lambda_i$	thermal	Doppler drive	ToF / two counter-rotating disk choppers			
monochromator	CaF_2 422	Si 111	none / ToF			
analyzer	CaF_2 422	Si 111	PG 002	PG 004	PG 002	PG 004
λ_{f} [Å]	2.23	6.27	6.67	3.33	same as IRIS	
E_{f} [meV]	16.45	2.08	1.84	7.38	same as IRIS	
ΔE	-125 to 150 µeV	± 31 µeV	± 0.4 meV	-3.5 to 4.0 meV	± 0.4 meV	-3.0 to 4.0 meV
ΔE_{res}[µeV]	8	0.7	17.5	54.5	24.5	99.0
Δt_{res}	$\approx$0.1 ns	$\approx$1 ns	$\approx$40 ps	$\approx$10 ps	$\approx$25 ps	$\approx$5 ps
Q-range [$\mathrm{Å}^{-1}$]	$0.2 - 4.9$	$0.2 - 1.8$	$0.4 - 1.9$	$0.8 - 3.7$	$0.2 - 1.8$	$0.4 - 3.6$
flux at sample [$\mathrm{cm}^{-2}\mathrm{s}^{-1}$]	2×10^4	1.8×10^5	$1 \times 10^{7**}$		$2.7 \times 10^{7**}$	

* IRIS can also use Mica as analyzer crystals (in 002, 004, 006).

** Flux at 150 µA ISIS Power.

4. Materials and Methods

The first section of this chapter illustrates the properties of the studied protein Alpha-Lactalbumin (A-L) and is followed by the preparation process for the neutron experiments. In section 4.3 the data reduction and corrections are desribed and in the last section the used Python script for data analysis is explained.

4.1. The sample: Alpha-Lactalbumin

This section gives a brief introduction to the protein Alpha-Lactalbumin (A-L). The biochemical content in this section is mainly based on the information provided by the review article of Permyakov and Berliner [51], which is recommended for more detailed information. A-L is a small milk protein and is produced in the mammary gland. Together with the enzyme β-1,4-galactosyltransferase (β4GalT) it forms a complex which is responsible for the lactose synthase, i.e. transforming galactose and glucose into lactose. A-L from most mammals (e.g. human or bovine) consists of 123 amino acid residues [52] and its weight is $\approx$14.2 kDa[1]. The A-L used in this work always refers to **bovine A-L**. It has a binding site which strongly binds the cation Ca^{2+}. The binding of Ca^{2+} mainly influences the tertiary structure (3D shape of protein) and not the secondary structure (local form of proteins, e.g. alpha-helix or beta-sheet). The binding site can also bind Mg^{2+}, Mn^{2+}, Na^{+} or K^{+} and induce similar but smaller structural changes than Ca^{2+}. However the binding constants are much lower except in the case ofMn^{2+}. In general, the binding of a cation stabilises A-L and increases the thermal denaturation temperature. Furthermore, it was recently shown by Shinozaki et al. [53] that Ca^{2+} and Mn^{2+} accelerate the folding to the native form of A-L, an effect not seen with the other cations. A-L can also bind Zn^{2+} at several other distinct binding sites, but decreases the stability of A-L bound to Ca^{2+}. Another interesting feature of A-L is that it forms a molten globular state either under acidic pH conditions or in the apo-form at elevated temperatures above 25 °C [54–57]. The apo-form refers to A-L which is not bound to Ca^{2+} and will also be called depleted A-L (A-L_{dep}) from now on. Owing to the characteristics described above, A-L is often used as a simple model for Ca^{2+} binding proteins and as a classical example of a molten globule state. Finally of interest is the study A-L in anti-cancer complexes HAMLET and BAMLET (Human/Bovine Alpha-lactalbumin Made LEthal to Tumors) [58, 59].

For the work in this thesis, A-L was chosen as a simple model system to compare the different models explained in section 2.2. To date there are no published temperature dependent elastic intensity scans and in addition, its small size makes it a good system

[1] The unit dalton (Da) is the same as the unified atomic mass unit (u), but is often used for molecular masses of proteins.

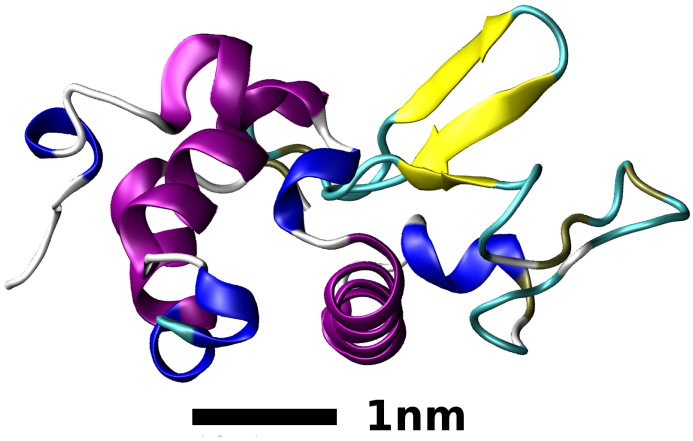


Figure 4.1. - Structure of bovine Alpha-Lactalbumin. Structure is based on Protein Data Bank [60] entry 1F6S.

Table 4.1. - Bovine Alpha-lactalbumin (A-L): characteristics. Values for scattering cross sections σ_{coh} and σ_{coh} are calculated with numbers given in Table 2.2.

number of amino acids	123
weight	14.2 kDa
atoms	$C_{626}\,H_{966}\,N_{162}\,O_{196}\,S_9$
total σ_{coh}	7.78 kb
total σ_{inc}	77.62 kb
number of A-L per 100 mg	$\sim 4 \times 10^{21}$

to simulate using molecular dynamics (MD). Furthermore, it can be produced in large quantities, necessary for neutron experiments which require sample sizes of around 100 mg for good statistics. It can be bought commercially in large quantities and at a reasonable price. The main characteristics of bovine A-L are summarised in Table 4.1, together with important properties relevant to the neutron experiments. The sequence of the 123 amino acid residues and their composition is shown in Table 4.1. An image of the structure of A-L is shown in Figure 4.1, based on the Protein Data Bank[60] entry 1F6S.

Three neutron studies using the QENS technique have been previously performed on Alpha-Lactalbumin [61–63]. Bu et al. [61, 62] compare the dynamics in native bovine A-L_{ca} with that in the globular molten and the denatured states. All measurements were taken at 300 K on four different spectrometers in NIST (FC-TOF, SPINS, HFBS and DCS). The purpose of the study was to gain a deeper understanding of the mechanics of protein folding and concludes that the secondary and tertiary structures are established in the early stages of protein folding.

Simulations on A-L have also given some insights into the dynamics in this protein. Skidar et al. [64] investigated its binding to the enzyme β4GalT and found that the A-L_{ca} form binds favourably to the enzyme, but in case of A-L with Mn^{2+} instead of Ca^{2+} the binding was less stable and not favoured anymore, showing the importance of the metal ion inside A-L. Brotzakis et al. studied the dynamics of the hydration water

around native and misfolded A-L [65]. Glazer et al. simulated A-L along with several other Ca^{2+}-binding proteins to improve structure-based function prediction algorithms that identify Ca^{2+} binding sites [66]. Tarek et. al. [63] compared QENS data for A-L_{ca} and the globular molten state, measured on the instrument DCS at 300 K with molecular dynamics simulations. The instrumental resolution of DCS was 32 µeV corresponding to a time scale of several 100 ps. The simulation was based on the PDB entry 1F6S at 300 K in a water box and the results were derived from a 2 ns simulation. They found a good agreement between experiment and simulation and strong heterogeneity of motions within the protein.

4.2. Sample preparation

All experiments described in this thesis use Alpha-Lactalbumin (A-L), either in its natural form with Ca^{2+} (A-L_{ca}) or Ca^{2+} depleted A-L_{dep}. Two different sample batches were produced, which will be referred to as batches B1 and B2. B1 was prepared only for A-L_{dep} in November 2016 and B2 for both A-L_{ca} and A-L_{dep} in January 2017. The protein was purchased from Sigma-Aldrich in lyophilized powder form (for detailed description see Table 4.2). It may have contained traces of ammonium sulfate and sodium phosphate which have been shown to possibly contribute to small spurious effects at low temperature [20], but these were small enough to not affect the current study. Three different hydration levels were prepared for each batch and protein type, hydrated with heavy water D_2O. The hydration level was determined from the weight with D_2O in comparison to that of the dry lyophilized protein protein and is defined as

$$h = \text{g } D_2O/\text{g dry protein.} \tag{4.1}$$

The different levels of hydration used for this work were $h \approx 0$ (dry), $h \approx 0.4$ and $h \approx 0.8$. The dry sample represents the case where only harmonic motions are present up to room temperature. 0.4h corresponds to around one or two layers of water on the protein surface [67], sufficient to allow for localised dynamical motions, and 0.8h represents a gel state close to full hydration. Hydrated powders were prepared in a desiccator. The purchased lyophilized protein powder was loaded into open flat aluminium sample holders which are standard for neutron backscattering measurements. The holders were then placed inside the desiccator together with silica beads and the desiccator closed so that the protein remained as dry as possible, in case the powder contained traces of residual water (a photo of the holders is shown in Figure 4.2). The sample was dried inside the dry atmosphere of the desiccator for at least one day, after which it was weighed and then vacuum sealed with indium wire of 1 mm thickness. To measure the protein weight, the empty sample holder was weighed before and subtracted from the total weight before closing. The 0.4h and 0.8h samples were obtained in the same way, but before sealing the holders they were put again into the desiccator together with a petri dish filled with pure D_2O instead of the silica beads. The samples were weighed periodically until they achieved the desired uptake of D_2O and then sealed like the dry sample. Details of the weighing process and the final powder hydration can be found in Appendix A. The final

masses and the naming of the samples hereon are given in Table 4.3.

Table 4.2. - Alpha-Lactalbumin, data sheet. Data from Sigma data sheet from bought protein: alpha-lactalbumin from bovine milk in lyophilized powder form. Purification by polyacrylamide gel electrophoresis (PAGE). Batch 1 (B1, November 2016) and B2 (January 2017) were prepared in exactly the same way and the procedure is described in section 4.2.

	depleted Ca^{2+}	with Ca^{2+}
Product Number	L6010	L5385
CAS Number	9051-29-0	
Quality	may contain traces of $(NH_4)_2SO_4$ (ammonium sulfate) and sodium phosphate	
Purity	$\geq 85\%$ (by PAGE)	
Depletion	precipitation with sodium sulfate, Ca$\leq$ 0.3 mol/mol A-L	
Lot for B1	SLBM1466V	
Lot for B2	SLBP6186V	SLBP6187V

The flat aluminium sample holders used for the experiments are almost invisible for neutrons (transmission of $\geq 99\%$) and the sample is distributed over an area of $4 \times 3\,\mathrm{cm}^2$ with a thickness of $0.6 - 0.7\,\mathrm{mm}$. Aluminium is the preferred material for neutron experiments since its total cross section is very small ($\sigma_{\mathrm{tot}}^{\mathrm{Al}} = 1.73\,\mathrm{b}$). D_2O instead of H_2O is used as the solvent so that the majority of the signal of interest arises from the hydrogen atoms in the protein. The total incoherent scattering of one A-L unit is 77 kb. Considering a hydration level of 0.4 and 0.8 D_2O the incoherent scattering contribution of the solvent is around 1.5% and 3% to the total incoherent scattering. Therefore the solvent fraction can be neglected in comparison to the sample scattering.

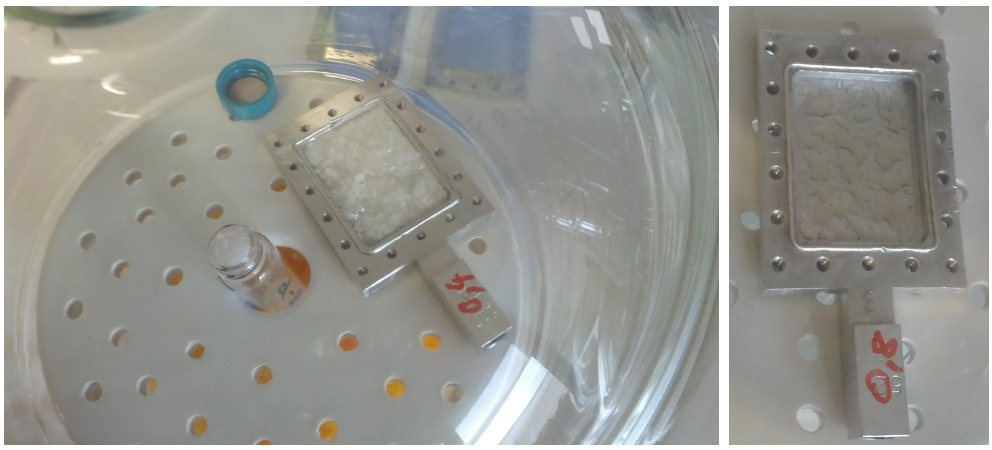


Figure 4.2. - Alpha-Lactalbumin preparation. Left side: Flask and flat aluminium sample holder filled with A-L_{dep} and inside a desiccator to dry the sample. Under the sample the orange silica beads are visible. For the hydration they are exchanged with a petri dish filled with D_2O. Right side: The 0.8h sample of B1 inside the desiccator before closing.

Table 4.3. - Alpha-Lactalbumin, samples. Batch 1 (B1) was prepared in November 2016 and B2 in January 2017. h is defined in Eq. (4.1) and is the uptake of D_2O after the sample has been in the desiccator. Note that the dry sample could still contain some residual water. All transmission measurements were made using IN13 and acquired over a time of 5 min and have a standard deviation of around 1% (assuming Poisson counting statistics).

Name	h	mass protein [mg]	transmission [%]
B1, A-L$_{dep}$			
dry	0	97.8	94
0.4h	0.40	102.5	94
0.8h	0.85	101.5	93
B2, A-L$_{dep}$			
dry	0	110.2	97
0.4h	0.40	115.9	94
0.8h	0.80	111.9	96
B2, A-L$_{ca}$			
dry	0	78.5	97
0.4h	0.40	94.8	95
0.8h	0.80	105.7	95

4.3. Experimental data: acquisition, reduction and corrections

All data on A-L shown in this thesis were collected on the four neutron instruments IN13 (ILL), SPHERES (FRM-II,MLZ), OSIRIS and IRIS (ISIS) (described in section 3.4). In all experiments the samples were cooled to cryogenic temperatures (between $3-20$ K) and then measured on warming until 310 K. On IN13 and SPHERES the data was acquired by collecting data during a slow heating ramp between 0.08-1 K/min (gradient measurement) and on OSIRIS and IRIS the temperature was increased in steps of 10 K or 20 K and after a short equilibrium time of several minutes, the signal was acquired at constant temperature (step measurement). All samples were placed at a 135° angle with respect to the incoming neutron beam to minimise geometric effects of the flat sample holder with respect to the detector angular coverage. The initial data reduction was done with LAMP [68] for IN13, Slaw[2] for SPHERES and Mantid [69, p. v.3.11.0] for OSIRIS and IRIS. Slab can corrections (see next paragraph) for a flat sample holder and normalisation was done with LAMP for the samples measured on IN13 and SPHERES. The measurements on OSIRIS and IRIS were corrected using the empty sample holder and normalised in Mantid. All intensity normalisations were done with the lowest available temperature data of each scan. Therefore, the difference between the slab correction algorithm and the subtraction of the empty sample holder alone are negligible. The reduced and corrected data was

[2]http://apps.jcns.fz-juelich.de/man/slaw.html

Table 4.4. - Alpha-Lactalbumin, measurements and raw data treatment. The list is in chronological order. The data acquisition is done by cooling the sample to T_{low} and then warming it up by ramping or in steps of 10 K to 20 K. On IN13 two measurements were performed: The first with all samples of B1 and the second with B2 (only A-L_{ca} dry and 0.4h, short test run A-L_{dep} 0.4h).

Instrument	Date	T_{low} [K]	Data acquisition	Raw data treatment	Batch
IN13	Nov. 2016*	20	gradient	LAMP	B1
IRIS	Dez. 2016	10	steps	Mantid	B1
SPHERES	Feb. 2017	3	gradient	SLAW, LAMP	B2
OSIRIS	Nov. 2017	10	steps	Mantid	B2
IN13	Mar. 2018**	20	gradient	LAMP	B2

* ILL DOI: [70]
** ILL DOI: [71]

exported for data analysis in Python. A summary of the measurements undertaken is given in Table 4.4.

The slab can correction implemented in LAMP corrects the raw data not only by the empty sample can but also for attenuation due to geometric effects of the flat sample holder in relation with the detector positions. This method is more accurate than a simple subtraction of the empty can, but in fact for data from hydrated powder samples the differences are not significant when normalised to the lowest temperature data set. In order to use this correction the transmission of the sample and the empty can must be known. The algorithm is fully explained in [6, p.142] and the final correction formula is

$$I_{\text{cor}} = \frac{1}{A^{\text{sample}}_{\text{sample+can}}} \left(I^{\text{sample+can}}_{\text{sample+can}} - \frac{A^{\text{can}}_{\text{sample+can}}}{A^{\text{can}}_{\text{can}}} I^{\text{can}}_{\text{can}} \right), \tag{4.2}$$

whereas

- $I^{\text{sample+can}}_{\text{sample+can}}$ is the intensity of the experimental measured spectrum,
- I_{cor} is the intensity of the theoretical spectrum,
- $I^{\text{can}}_{\text{can}}$ is the intensity of the measured empty sample holder (can),
- $A^{\text{sample}}_{\text{sample+can}}$ is the attenuation of the intensity scattered by the sample, when the sample is inside the can,
- $A^{\text{can}}_{\text{sample+can}}$ is the attenuation of the intensity scattered by the can, when the sample is inside the can
- $A^{\text{can}}_{\text{can}}$ is the attenuation of the intensity scattered by the can for the can alone.

The superscript describes which component is actually scattering and the subscript describes which components are present during the scattering event. $A^{\text{sample}}_{\text{sample+can}}$, $A^{\text{can}}_{\text{sample+can}}$ and $A^{\text{can}}_{\text{can}}$ can be calculated by knowing the angle of the sample holder and the transmission of the sample and can separately. The transmission of the sample alone is approximated by the transmission of the sample and the can together (values measured on IN13) as the

transmission of the aluminium sample holder is $\geq 99\%$. Since the data is normalised by the low temperature data, the value $A^{\text{sample}}_{\text{sample+can}}$ is not important. The values of $A^{\text{can}}_{\text{sample+can}}$ / $A^{\text{can}}_{\text{can}}$ evaluated for all used Q values were between 90% and 92% for IN13 and SPHERES. The slab can correction is based on the assumption of single scattering of the neutron either by the container or the sample. It also holds for multiple scattering if cross effects (first scattering in can, second in sample or vice versa) can be neglected. If a second scattering takes place it contributes to the signal as a multiple scattering event, where information about the scattering direction is lost. For transmissions larger than 90% it is commonly neglected [72], but there is in fact no consensus in the literature [73]: In some publications the authors feel safe with a transmission of 99% [74], but as stated above, most are using transmissions of around 90% (e.g. [75, 76]). Others are comfortable with even lower transmissions [77]. Busch et al. [73] investigated multiple scattering effects in QENS and estimated a deviation in intensity of around 3% in their investigated Q range from $0.4 - 1.8\,\text{Å}^{-1}$. Multiple scattering is mainly isotropic [78] and thus it creates a Q independent background [25]. If multiple scattering is not corrected for and takes place in the sample, it leads to an intensity difference at Q=$0\,\text{Å}^{-1}$ compared to the assumed single scattering event (see e.g. [79]).

4.4. Data analysis

The reduced, corrected and normalised data were imported to Python for data analysis. A new script was written to read the data and perform simultaneous fitting with different models. Model parameters can be fixed or floating. The fitting was based on the LMFIT-package [80]. As fitting procedure the least-square method was used, which includes the errors of the individual data and minimises the following equation

$$\chi^2 = \sum_{i=1}^{N} \frac{\left(y_i^{\text{data}} - y_i^{\text{model}}\right)^2}{\epsilon_i^2}, \tag{4.3}$$

where N is the total number of data points, y_i^{data} are the experimental data and y_i^{model} the values obtained with the fitting model. The statistical error ϵ_i of each value y_i^{data} was evaluated by LAMP and is based on Poisson counting statistics in each detector, i.e. error=$\sqrt{\text{number of counts}}$. In order to quantify the quality of the fit, the reduced value $\chi^2_{\text{red}} = (\chi^2/\text{no of free parameters})$ is often used. The workflow of the program is described in Figure 4.3.

There were two major challenges fitting the proposed models to the corrected data. The first relates to statistical uncertainties. The statistical error in the evaluated intensities is large, especially at low Q values and for data collected on IN13 and OSIRIS. In addition, in BSS there are not that many data points covering Q since the number of detectors is usually limited (SPHERES and IN13). This leads to a large variation in the fitting parameter, EISF(0), between the models and has a strong influence on the evaluated MSD, as will be shown in section 5.2. For the linear fits of the GA model, ln[EISF(Q)] vs Q^2, the statistics are also important but ultimately the result is always the same since

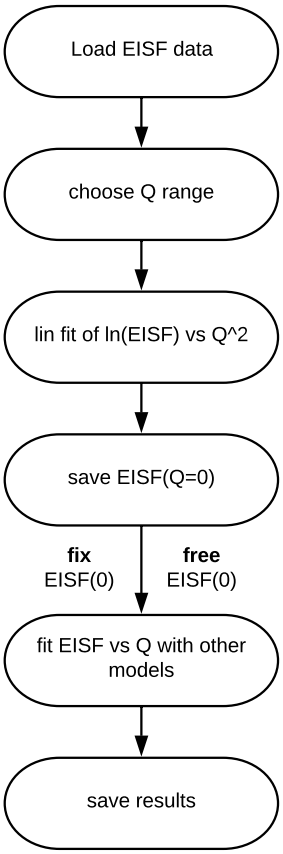


Figure 4.3. - Workflow of fitting program.

there is only one possible solution. For the more sophisticated models, the algorithm may end up trapped in a local minimum rather than the global minimum. Furthermore, the value of χ^2_{red} may be very similar for quite different values of the fitting parameters. This issue can be solved by fixing EISF($Q = 0$) to the value obtained by the GA model and consequently reducing the number of free parameters by one, making the fits overall more stable. The fitting script in Python was developed to facilitate the fitting procedure with a fixed parameter.

The second challenge was the high correlation between the remaining two parameters in the cases of the PK and Yi models. For some data sets, two quite different solutions could have very similar least square values. This effect was also reduced by fixing the value of EISF($Q = 0$). Non-physical results were suppressed by defining lower and upper limits for each parameter. Importantly, every single fit was not done by 'hand' so that the results were not biased by a personal choice of Q-range or the starting values of the fitting parameters, and to create an efficient procedure to deal with large data sets. As an additional check however, the program was made to output plots of all data sets with the used fitting range, the fit itself and the evaluated parameters.

5. Experiments with Alpha-Lactalbumin

This chapter is heavily based on my publication in *JCP* [33]. It evaluates the performance of the four different models described in section 2.2 and the effects on the resulting MSD based on data collected on the powder samples of bovine Alpha-Lactalbumin described in section 4.1 on three backscattering spectrometers described in section 3.4. The first section introduces the motivation for the analysis. The second section describes the limitations of the techniques and the methodology used to tackle them. In the third section, the methods are applied to the calcium depleted Alpha-Lactalbumin A-L_{dep}. The results are compared with A-L_{ca} in section four, and in the last section the overall results are discussed.

Disclaimer: Sections 5.2, 5.5 and sub-sections 5.3.1, 5.3.2, 5.3.3, 5.3.4 are almost entirely taken from my publication in *JCP*, Zeller et al. [33].

5.1. Introduction

MSDs extracted from EINS are commonly used to gain insights into the internal dynamics of a protein. Further and more detailed information can be obtained by measuring QENS spectra (see section 2.1.3), but they are more time-consuming. EINS measurements are often done with hydrated powder samples to suppress whole protein translational and rotational movements and focus on localised or confined atomic motions. This leads to a pronounced elastic peak in the dynamic scattering function which is superposed to the QENS signal, containing information on the mobile fraction of H atoms in the sample. The EINS signal is measured as a function of temperature over a large range of temperatures, from very low values of around 10 K to at least physiological temperature around 320 K [81]. However, EINS does not exclusively have to be a function of temperature. Recently, the influence of pressure on molecular dynamics is equally investigated by neutron scattering [82, 83].

In general, the MSD obtained for protein powders increases linearly in the low temperature regime implying that only small harmonic motions of the atoms are present at these temperatures. At higher temperature anharmonic motions are possible and the flexibility of the protein changes leading to a divergence from the linear regime and a steep increase of the MSD with temperature. This effect is often called the *dynamical transition* and takes place around ≈ 220 K [2, 84, 85]. The origin of this transition is still largely debated and there are several possible explanations [4]. Zaccai [84] suggested a change in the effective elasticity of proteins by introducing a spring constant to describe the low and high temperature regimes by two different pseudo-force constants. Others have suggested

a correlation with the onset of translational motions of bound water molecules [86], with a fragile-to-strong transition of the hydration water [87], or have assigned it to a glass transition in the hydration shell [88]. Others ascribe it to purely instrumental resolution effects [89, 90]. There are indications that the dynamical transition is related to the onset of biochemical activity of the proteins e.g. [85, 91]. The dynamical transition is only observed for hydrated powders, whereas it is absent in dry conditions [84]. It is important to point out that an additional deviation from the linear low temperature regime is observed above 100 K. It is related to the onset of methyl group rotations which are independent of hydration [92, 93] and can contribute significantly to the neutron spectra. For many proteins studied by neutron scattering, methyl group hydrogens contribute around 25% to the total hydrogen signal [92].

In what follows, the models discussed in section 2.2 are applied to EINS data for A-L with and without calcium, A-L_{ca} and A-L_{dep} (see section 4.1 for details). To study different dynamical conditions of the protein, three different hydration levels were measured, h ≈ 0 (dry), h ≈ 0.4 and h ≈ 0.8 [h=gD_2O / g dry protein]. As explained before, the dry sample represents the case where only harmonic motions are present up to room temperature. 0.4h corresponds to about one or two layers of water on the protein surface[67], sufficient to allow for dynamical motions resulting in the onset of the dynamical transition. 0.8h represents a gel state close to full hydration and in fact leads to some ice formation at temperatures below 270 K since not all water molecules are bound to the protein. Three neutron backscattering spectrometers (OSIRIS, IN13, SPHERES, see section 3.4) were used to have access to motions from a few picoseconds to a few nanoseconds (see section 3.3) and to investigate the influence of the instrumental resolution.

5.2. Methodology

In order to compare the models to the experimental data, the broadening of the DSF $S^{\mathrm{R}}_{\mathrm{inc}}(Q, \omega \approx 0, \Delta\omega)$ due to the resolution function of the instrument has to be taken into account, as explained in section 2.2.6. The $S^{\mathrm{R}}_{\mathrm{inc}}(Q, \omega \approx 0)$ has to be normalised by the instrumental resolution. This is done by dividing the experimentally measured DSF with the lowest temperature $\mathrm{T_{low}}$, $S_{\mathrm{inc}}(Q, \omega \approx 0, T = \mathrm{T_{low}})$, where the mobility of the sample is assumed to be very close to zero:

$$\frac{S^{\mathrm{R}}_{\mathrm{inc}}(Q, \omega \approx 0, T)}{S^{\mathrm{R}}_{\mathrm{inc}}(Q, \omega \approx 0, T = \mathrm{T_{low}})} = S^{\mathrm{R}}_{\mathrm{norm}}(Q, \omega \approx 0, T) \equiv \mathrm{EI}(Q, T) \tag{5.1}$$

The models are applied to this normalized DSF which for simplicity is labeled as elastic intensity EI(Q) from now on. As explained in more detail in section 2.2.6 this quantity is not the same as the EISF(Q). The EI(Q) is only developed over a finite amount of time $\tau_{\mathrm{res}} = \hbar/\Delta\omega_{\mathrm{res}}$ which is defined by the instrument resolution (see section 3.3). This leads to an evaluated MSD that is dependent on the used instrument resolution. All the data was analyzed with the Python script described in section 4.4 which fits all models by the least-square method. The value $\chi^2_{red} = (\chi^2/\text{number of free parameters})$ can then be used to quantify the quality of the fit.

In section 2.2 four different models were introduced in detail and are now used to fit the data: the Gaussian Approximation (GA), the Peters and Kneller model (PK), the model by Yi et al. (Yi) and the Doster model (Do). The GA model only describes the Gaussian part of the EISF, whereas the PK and Yi model also describe the non-Gaussian part by including heterogeneity. The Do model also corrects the Gaussian part of the GA by including anharmonicity. The models have different Q ranges of applicability. The PK and Do models can be used to fit the entire available Q-range for all instruments. The PK model will be applied here to all instruments whereas the Do model is only shown for IN13 as it has the broadest Q-range. The Yi model can also fit a broad Q-range, but in cases of fast decay of the EISF with increasing momentum transfer Q, the expansion of the GA to Q^4 is limited as it neglects higher order terms. Therefore, a cut-off to the Yi model similar to the GA is introduced, after which the EISF at high Q-values will not be fitted anymore. To use a consistent Q-range between the three different hydrated A-L samples, the following procedure is used:

First, the minimum Q value, Q_{min}, is chosen to be the same for fitting all models to a given data set from a given instrument. Then the maximum Q value, Q_{max}, is evaluated from the data for the most hydrated sample (0.8h A-L) at the highest temperature T_{max}, since the decay of the EISF with increasing Q is the largest. For the GA, Q_{max} is determined as the last Q-value where the fit $\ln[\mathrm{EI}(Q)]$ vs Q^2 is linear. The Yi model can describe the entire available Q-range for the OSIRIS and SPHERES instruments. However, for IN13 (up to 4.5 Å^{-1}) the model reaches its limit as mentioned before and a cut-off value at $Q_{max,Yi} = 2.5\,\text{Å}^{-1}$ is introduced. It provides the best compromise between including the widest range of Q-values and yet still describing the experimental data satisfactorily for the 0.8h sample. The Q_{max} evaluated for each model and each instrument was then the same for all samples and temperatures. The consequences due to the limited number of data points in the available Q-range are elaborated in section 5.2.2. A summary of all the different Q-ranges used for the GA, Yi, PK and Do models can be found in Table 5.1. A final important point is the relevance of analysing different Q-ranges, which is covered in section 5.2.3.

Table 5.1. - Different Q-ranges used for the various models and instruments.

	Q-range [Å^{-1}]			
Model	GA	Yi	PK	Do
IN13	0.52-1.7	0.52-2.5	0.52-4.5	0.52-4.5
	1.7-4.5	-		-
SPHERES	0.34-0.6	0.34-1.8		-
	0.60-1.2	0.60-1.8		-
	0.96-1.8	-		-
OSIRIS	0.30-1.5			-

5.2.1. The question of the intercept EI(Q = 0)

All models should start with the same value of the EI(Q) at zero momentum transfer. In theory the value should be 1 for the normalised DSF EI(Q). Due to instrumental and experimental effects, like multiple scattering or coherent effects, the starting value is often lower than 1, especially at higher temperatures. For this reason, the value of EI($Q = 0\,\text{Å}^{-1}$) is introduced as a fitting parameter ≤ 1 and its consequences evaluated here. In most publications about EINS data, the value at EI($Q = 0\,\text{Å}^{-1}$) is not clearly defined. On the contrary, in the case of the normally used linear fit of $\ln[\text{EI}(Q)]$ vs Q^2, only the slope ($\propto$ MSD) is reported and not the intercept with the y-axis ($= \ln\left[\text{EI}(Q = 0\,\text{Å}^{-1})\right]$). This is surprising as will be apparent later since this value has a strong impact on the resulting MSD value (see section 5.2.2). If the GA is used this value is unique since a linear fit has a global minimum and therefore only one solution but in the case of more complex models, a change in the value at EI($Q = 0\,\text{Å}^{-1}$) results in a significant variation in MSD. Furthermore due to the limited experimental information at Q values close to zero given by instrument geometry constraints, and the statistical error of experimental data, it is often not possible to get an unique solution for EI($Q = 0\,\text{Å}^{-1}$). Since it is also not possible to fix this value to 1 in all cases, it was fixed to be the same for all models and to the one obtained by the GA as result of several trials (see section 5.2.1).

To illustrate the importance of the axis intercept EI($Q = 0\,\text{Å}^{-1}$), two representative examples are shown in Figure 5.1. Both data sets are measured on SPHERES at around 295 K. The first one corresponds to A-L_{dep} at 0.4h (Figure 5.1a,c,e) and the second one to A-L_{dep} dry (Figure 5.1b,d,f). In order to qualify the differences between two fits on the same data set, they are compared in terms of the least-square error, evaluated by χ^2_{red} (see Eq. (4.3)). The evaluated MSD and χ^2_{red} for the different cases are given in Table 5.2. Figure 5.1a shows a visual under-evaluation of the EI($Q = 0\,\text{Å}^{-1}$) for the PK model in comparison to the GA if one does not constrain the fit. The gray dashed line indicates the maximum Q-value used for the fit of the GA. The PK and Yi model consider all available Q-values from 0.3 to 1.8 Å^{-1}. The smallest Q-values, 0.3 and 0.46 Å^{-1}, are not well described by the fit of the PK model because EI($Q = 0\,\text{Å}^{-1}$) $= 0.86 \pm 0.03$ is much smaller than the GA EI($Q = 0\,\text{Å}^{-1}$) $= 0.93 \pm 0.02$ which fits these points well. In contrast the Yi model has a slightly higher value of EI($Q = 0\,\text{Å}^{-1}$) $= 0.95 \pm 0.01$ than the GA, leading to a difference of almost a factor of two in the MSD between the PK and Yi model. If the EI($Q = 0\,\text{Å}^{-1}$) of these two models is fixed to the value obtained by the GA (see Figure 5.1c), the lowest two Q-values are now well described by both models, as are the higher Q-values. The differences in the resulting MSD and χ^2_{red} for the fixed and free case are shown in Table 5.2. For the PK model, the $\chi^2_{red,\,fixed} = 12$ for the fit where the intercept was fixed is $\approx 20\%$ larger than that of the free fit result $\chi^2_{PK,\,free} = 10$. For the Yi model, the $\chi^2_{red,\,fixed} = 5.1$ for the fixed fit is also $\approx 20\%$ larger than for the free fit $\chi^2_{red,\,free} = 4.0$. In addition there are big differences in χ^2_{red} between the PK and the Yi model even if the fits are visually quite similar for the fixed case (see Figure 5.1c). The reason for these is the very small error bars of the counting statistics by which the χ^2 is weighted (see Eq. (4.3)). The main differences are visible in the Q-range $1.0 - 1.8\,\text{Å}^{-1}$ where the Yi model follows the data better (see Figure 5.1c). However, more importantly is the effect on the evaluated MSD of the PK model. In the free case it is $1.3 \pm 0.3\,\text{Å}^2$

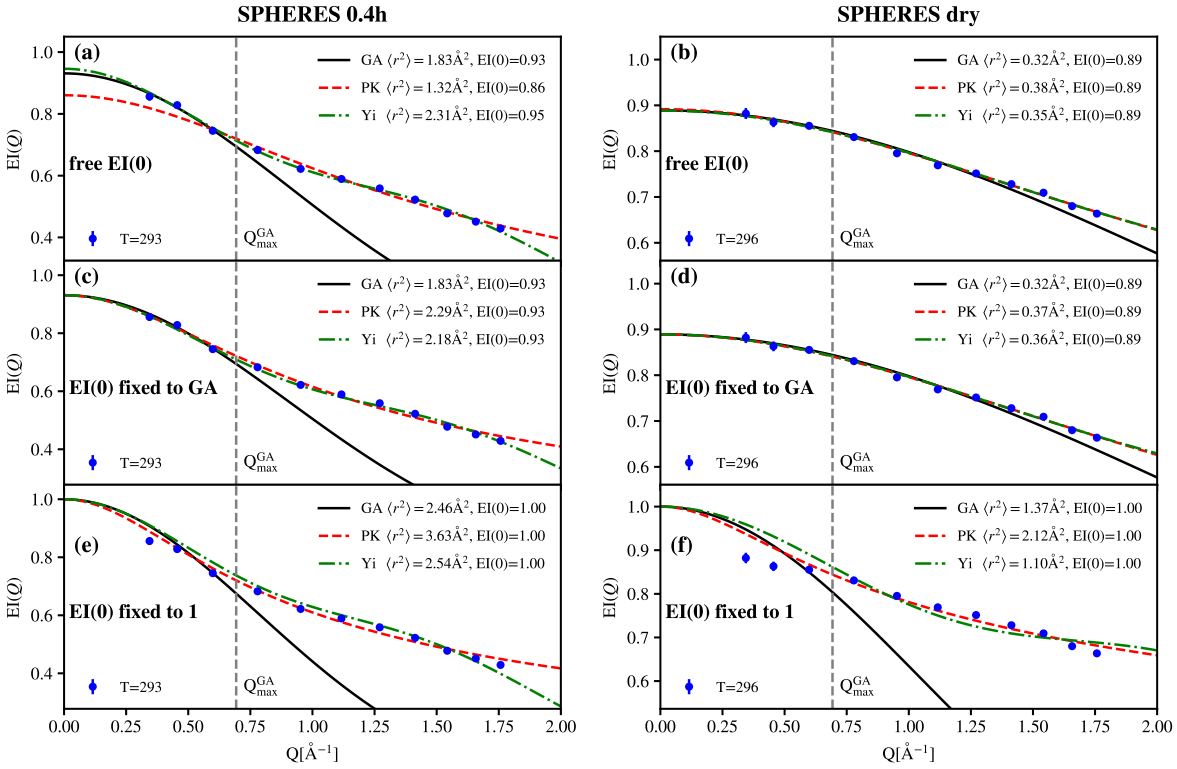


Figure 5.1. - Example of the effect of EI($Q = 0\,\text{Å}^{-1}$) for the SPHERES instrument. The left graphs (a,c,e) show an example for A-L_{dep} with 0.4h at 293 K, the right graphs (b,d,f) show an example for dry A-L_{dep} at 296 K. In graphs a) and b) EI($Q = 0\,\text{Å}^{-1}$) is a free fit parameter. In graphs c) and d) EI($Q = 0\,\text{Å}^{-1}$) is a fixed to the value evaluated with the GA. In graphs e) and f) EI($Q = 0\,\text{Å}^{-1}$) is fixed to 1 for all models. [*This figure was published in [33].*]

while in the fixed case it is almost a factor of 2 larger with $2.3 \pm 0.2\,\text{Å}^2$. The changes of the MSD in the Yi model are not as large (2.3 vs 2.2 Å^2) since the value EI($Q = 0\,\text{Å}^{-1}$) only changed by a small amount. This example is shown to illustrate that even though the least square chi statistical value may be better for a free fit, the small Q-values can be under-evaluated. This is due to: 1) The statistical error being smaller for the higher Q-values leading to a higher weight on them, and 2) the larger Q-values can be better described by the models with a lower value at EI($Q = 0\,\text{Å}^{-1}$). This under-evaluation of the first Q-values is not only a problem for the PK model. It is not shown here, but it happens for all models describing higher Q-ranges.

Figure 5.1e shows what happens if the EI($Q = 0\,\text{Å}^{-1}$) is fixed to 1. Visually, the GA does not describe the first data point well, the PK model is describing the range better and the Yi model is worse. The resulting MSD values are much larger than in the two cases before (see Table 5.2). To emphasise that a EI($Q = 0\,\text{Å}^{-1}$) = 1 is not only problematic for hydrated samples, an example for the A-L_{dep} dry sample at the same temperature is shown in Figure 5.1f. Here the fits show that no model is able to describe the data in the low Q-range. In contrast, they perform well if a free and a fixed value of EI($Q = 0\,\text{Å}^{-1}$) to the GA is chosen (see Figure 5.1b,d). It also shows that in this case fixing the offset is

Table 5.2. - Effect of EI($Q = 0$) Å^{-1} on χ^2_{red} and $\langle r^2 \rangle$. Values for χ^2_{red} and MSD $\langle r^2 \rangle$ for the PK and Yi model with the value of EI($Q = 0\,\text{Å}^{-1}$) as free parameter, fixed to the value obtained by the GA and 1.

	Example	A-L_{dep} 0.4h 294K			A-L_{dep} dry 296K		
	EI($Q = 0$)	free	fixed to GA	fixed to 1	free	fixed to GA	fixed to 1
PK	$\langle r^2 \rangle$ [Å^2]	1.3±0.3	2.3±0.2	3.6±0.5	0.38±0.05	0.37±0.02	2.1±0.4
	χ^2_{red}	10	12	17	1.1	0.98	9.7
Yi	$\langle r^2 \rangle$ [Å^2]	2.3±0.1	2.2±0.1	2.6±0.2	0.35±0.03	0.36±0.02	1.1±0.2
	χ^2_{red}	4.0	5.1	28	1.2	1.06	29

unnecessary since all models evaluate to the same value and that the GA is also closer to the larger Q-values than for the case of EI($Q = 0\,\text{Å}^{-1}$) = 1.

The results can be summarised as follows: 1) In general, a higher EI($Q = 0\,\text{Å}^{-1}$) leads to a higher MSD; 2) the differences between the models can be large if they are allowed to have different values of EI($Q = 0\,\text{Å}^{-1}$) and should be kept the same for all models; and 3) fixing EI($Q = 0\,\text{Å}^{-1}$) = 1 is not always possible, therefore when comparing models, the value EI($Q = 0\,\text{Å}^{-1}$) should be fixed to the same value, which can be the value evaluated by fitting the GA in the low Q-regime where $\langle r^2 \rangle\, Q^2 \leq 1$. It may be the case that many experimenters use this method to fix the value in their publications, specifically where the authors state that they normalised to $Q = 0$, but many do not explain how they achieved that without the knowledge of EI($Q = 0\,\text{Å}^{-1}$).

5.2.2. Differences between considered Q-ranges within the GA and the influence of statistics

A second consideration when fitting models to EINS data to extract MSD is the definition of the Q-range to be fitted. Typically this is vague and deserves further consideration. Many authors in publications cite the criterion defining the upper value Q_{max} and some discuss arguments why it is reasonable to surpass this limit [94]. This becomes important because even the value of EI($Q = 0\,\text{Å}^{-1}$) from the GA, which will be taken in the following as starting value for all models, depends on the chosen Q-range and the statistical error of each measured intensity data point. This can be best shown through data collected on the backscattering spectrometer IN13 at the ILL. In order to compare accurately the differences of the MSD of similar proteins, usually the same Q-ranges are chosen for the fit over all temperatures. Figure 5.2 illustrates two problems if a limited Q-range is chosen. Figures 5.2a and b show fits to the GA which include data points up to $Q^2_{\text{max}} = 2.8\,\text{Å}^{-2}$, evaluated as the largest Q-range for the A-L_{dep} sample from the highest temperature (in total five distinct Q-values with a statistical error of around $5 - 10\%$). This region is similar to the Q-range of the other two spectrometers SPHERES and OSIRIS. By fitting only the very first few points where the GA is strictly valid, large differences result in the MSD even over a temperature range of just 10 K, see Figure 5.2a and 5.2b for the dry and 0.4h A-L_{dep}Within the statistics, the intensities EI(Q) at a given momentum transfer

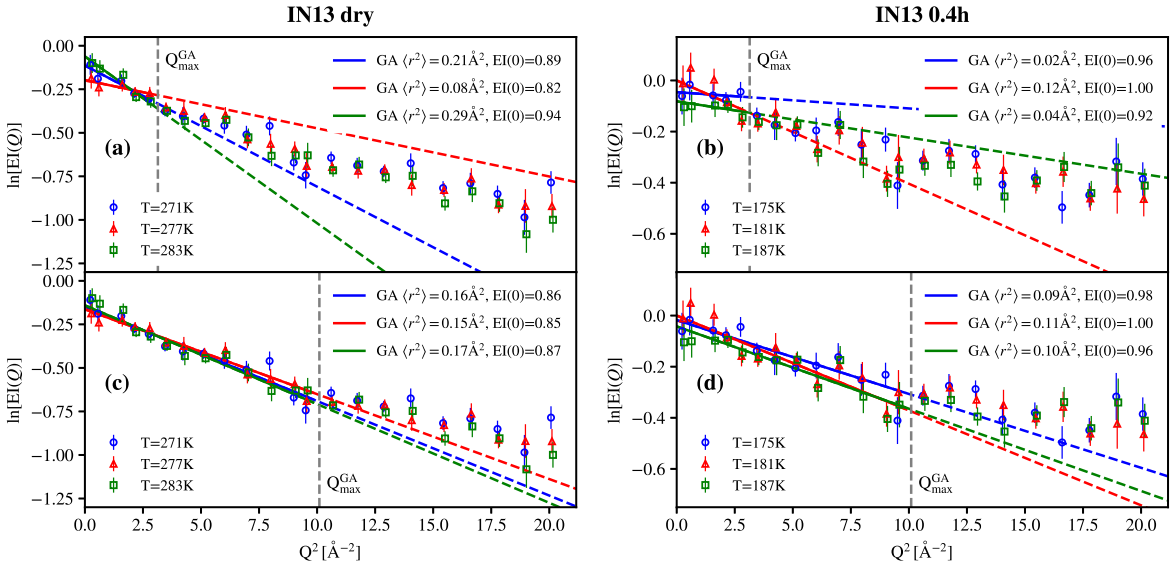


Figure 5.2. - Effect of statistics on the value of the EI($Q = 0\,\text{Å}^{-1}$) and the MSD evaluated with the GA model. For the dry A-L_{dep} an example is shown between 271 K and 283 K taken on IN13: a) If only the low Q-range (gray line) is considered, the results differ from 0.08 to 0.29 Å^2 which explains the drop of the MSD at 277 K shown in Figure 5.5a. c) Considering Q-values within the validity of the GA model up to $Q^2_{max} = 10\,\text{Å}^{-2}$ gives similar results for the MSD since the low Q-values are less weighted. b)+d): Similar example for the 0.4h A-L_{dep} between 175 K and 187 K: (b) low Q-range for GA (d) $Q^2_{max} = 10\,\text{Å}^{-2}$ (gray line). [*This figure was published in [33].*]

are almost the same for all temperatures, but due to the variation in absolute height of the first five Q-values, the resulting MSD can vary by a factor of 3 (from 0.08 to 0.29 Å^2) (see Figure 5.2a) or even 5 (from 0.02 to 0.12 Å^2) (see Figure 5.2b). On the contrary, if the Q-range is extended and the validity of the GA is assumed until $Q^2_{max} = 10\,\text{Å}^{-2}$ at all temperatures (a region that still reasonably corresponds to a linear fit region and is just slightly larger than the theoretically proposed limit $\langle r^2\rangle_{GA}\, Q^2_{max} \leq 1$), the resulting MSD is almost the same for all models within error bars: Figure 5.2c, $0.15 - 0.17$ Å^2 and Figure 5.2d, $0.09 - 0.11$ Å^2. A summary of the obtained $\langle r^2\rangle$ and χ_{red} is shown in Table 5.3.

Table 5.3. - Values for χ^2_{red} and $\langle r^2\rangle$ for the GA model with $Q^2_{max} = 2.8\,\text{Å}^{-2}$ and $Q^2_{max} = 10\,\text{Å}^{-2}$.

Q^2_{max} [Å^{-2}]	Example	A-L_{dep} dry			A-L_{dep} 0.4h		
	T [K]	271	277	283	175	181	187
2.8	$\langle r^2\rangle$ [Å^2]	0.21±0.04	0.08±0.03	0.29±0.06	0.02±0.04	0.12±0.04	0.04±0.04
	χ^2_{red}	0.38	0.21	0.92	0.30	1.3	0.24
10	$\langle r^2\rangle$ [Å^2]	0.16±0.02	0.15±0.01	0.17±0.02	0.09±0.02	0.11±0.01	0.10±0.02
	χ^2_{red}	0.98	0.48	1.1	0.75	0.96	0.75

This example shows two important aspects of using the largest available Q-range: 1) A larger Q-range results in more precise and consistent results since more data points can be included and 2) it also leads, in general, to a different EI($Q = 0\,\text{Å}^{-1}$) and MSD values, even when both should give the same result since they are still in the limit of the GA. Therefore is is important to include the highest possible Q-range to be as precise as possible, but also to stick with the same Q ranges to compare data accurately.

5.2.3. Two regimes - high Q-range

As already mentioned, the GA model is strictly valid for Q$\rightarrow 0\,\text{Å}^{-1}$, but some experimental and instrumental issues arise specifically at low Q values. Multiple scattering [78] or non-negligible coherent contributions may occur which result in EI(0) < 1. Additionally, detectors can have different resolutions for low momentum transfers as it is the case for SPHERES [44] and the very similar instruments IN16B and HFBS. In contrast, IN13 and OSIRIS do not suffer from detector resolution effects, but in general the counting statistics are also worse at lower Q-values.

To illustrate this, an example for each instrument and hydration level at common temperature between 305 K and 310 K, is shown in Figure 5.3 as plots of $\ln[\text{EI}(Q)]$ vs Q^2. As can be seen, the data appears to have different linear regimes. For IN13 and SPHERES a second linear regime at high Q is clearly visible, whereas OSIRIS shows only one. The

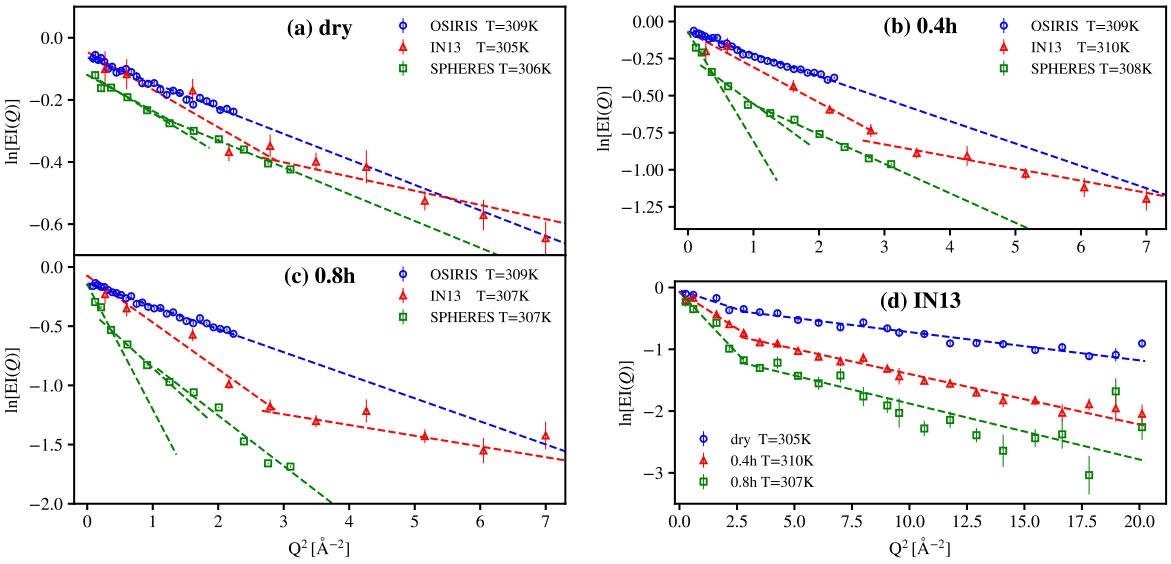


Figure 5.3. - Normalized $\ln[\text{EI}(Q)]$ vs Q^2 for all instruments around 310 K. Normalized reduced data $\ln[\text{EI}(Q)]$ vs Q^2 for the three different hydration levels (a-c) and the three different instruments, OSIRIS (blue), IN13 (red) and SPHERES (green). The dashed lines indicate linear fits to the data in their respective range. For OSIRIS at all hydration levels mainly one linear regime is visible. For IN13 clearly two regimes are visible, separated around $3\,\text{Å}^{-2}$. For SPHERES two low Q regions and one high Q region can be identified. Figure d shows the entire Q-range of IN13 and the linearity of the second regime up to $20\,\text{Å}^{-2}$. [*This figure was published in [33].*]

reason are the different dynamical processes which are visible at different time scales (see section 3.3 and 3.4). On SPHERES, the first two Q-values used at 0.34 and 0.45 Å^{-1} are clearly higher for the hydrated samples, which is probably due to a slightly reduced resolution of these two detectors [44]. Therefore two different low Q regimes are fitted for SPHERES, the first three detectors, including the two Q detectors with lower resolution ($0.34 - 0.60\,\text{Å}^{-1}$) and then the next four detectors after these two detectors ($0.60 - 1.2\,\text{Å}^{-1}$). To factor out effects at low Q and to evaluate the information that can be obtained in the second linear regime, the GA is also fitted to high Q values, even though it strictly falls out the $Q^2 \langle r^2 \rangle_{\text{GA}} \ll 1$ condition. A similar approach was taken in previous publications of IN13 data (e.g. Lehnert et al. [95] or Zanotti et al. [96]). Figure 5.3 illustrates using dashed lines all the different linear regimes fitted in this paper (for Q-ranges see Table 5.1).

In addition to the second linear regime at high Q-values, an ordering between the different instruments, and therefore resolution, can be clearly observed; the break between regions of linearity is moving to smaller momentum transfers with increasing resolution.

5.3. Results and Discussion for A-L_{dep}

In the following, the validity of the aforementioned models is discussed in more detail, comparing them in turn to data sets from the three spectrometers IN13, SPHERES and OSIRIS. In addition, the summed intensities versus the MSD evaluated with the PK model for the three instruments and hydration levels are shown and discussed. Note that even though a data set was collected on IRIS at ISIS, the data is not shown or discussed here since it was an incomplete set and added no new insights into the discussion.

5.3.1. Analysis of data from IN13

The Q-ranges used for IN13 data evaluation are given in Table 5.1. Representative fits are shown at three different temperatures for the A-L_{dep} at 0.4h and 0.8h in Figures 5.4a,b. All models describe the data points well within their specific Q-ranges. At high temperatures, for the 0.8h sample the Yi model follows the general behavior of the experimental data but does not fit as well as the other models. For this data a lower Q^{Yi}_{max} would be needed to obtain better agreement but a smaller Q-range would not include much more Q information than that already considered using the GA model. As explained above, all samples should be compared within the same Q-range. This results in $Q^{Yi}_{max} = 2.5\,\text{Å}^{-1}$ which is the best compromise between including the largest Q-values possible and a good description of the data with a given model.

The MSD results of the fits to the elastic normalized DSF, EI(Q) are shown in Figure 5.5. Specifically, Figure 5.5a shows the difference between the GA, PK and Yi model for the dry protein. The differences between the models are very small and all models show a similar behaviour. The same plot is shown for 0.4 hydration in Figure 5.5b. Here, the differences between the models are also small, but overall the PK and Yi model give rise to slightly higher MSD values due to the inclusion of data at higher Q values. At 0.8 hydration (see Figure 5.5c) the MSD values are clearly higher at high temperatures for both models. Both hydrated samples show an increase in the MSD at around 230 K compared to the dry protein as expected at the dynamical transition temperature, commonly observed

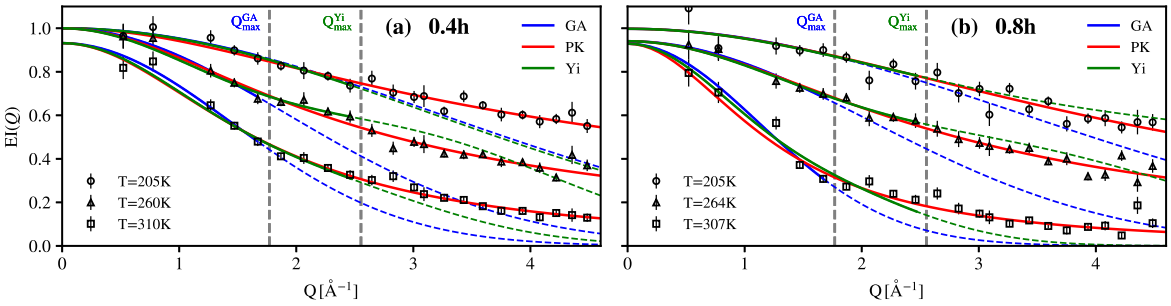


Figure 5.4. - Representative fits for the three models at three different temperatures for IN13 data. Figure a) corresponds to 0.4h and Figure b) to 0.8h, respectively. The vertical gray lines indicate the Q_{max} used for each model: $0.5 - 1.7\,\text{Å}^{-1}$ for GA and $0.5 - 2.5\,\text{Å}^{-1}$ for Yi. [*This figure was published in [33].*]

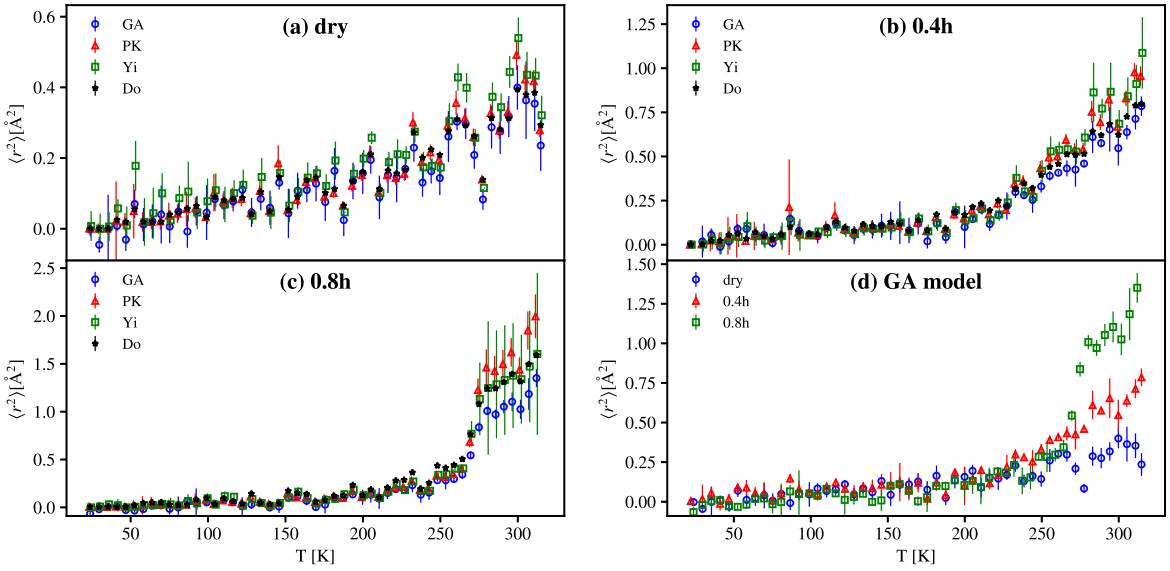


Figure 5.5. - IN13: Results of the MSD for all models. MSD values extracted from the GA, PK, Yi and Do models for dry, 0.4 and 0.8 hydrated A-L_{dep} depleted samples (a to c); IN13 data. Figure d) shows the MSD of all three hydration levels evaluated with the GA model. [*This figure was published in [33].*]

in hydrated proteins. For the 0.8h sample, around $270 - 280\,\mathrm{K}$ a very steep increase in the MSD is visible that can be attributed to the melting point of heavy water at 278 K. Probably for 0.8 hydration some free water exists that can freeze and therefore inhibit the motion of the protein. Therefore the MSD shows an abrupt rise at the melting point. After 280 K the increase in the MSD for all models is similar to the 0.4h sample, but the absolute values are higher. The MSD for the 0.8h sample is higher than for the 0.4h sample since the higher hydration decreases the crowding in the sample and thus allows more motions to take place. For both hydrated samples the error bars of the GA are smaller than for the other models. This gives the illusion of a higher accuracy, which is only due to the cut-off effect and to the exclusion of certain amplitudes. In the case of the Yi model at $270 - 320\,\mathrm{K}$, very large errors are evaluated since the fit is not describing the data as well as for the other cases (e.g. see Figure 5.4b at 307 K). This can be confirmed by the higher χ^2_{red} shown in Figure 5.6.

Finally, the performance of the models that include heterogeneity is compared with the double well potential model (Do model) that takes into account anharmonicity. Since IN13 has by far the largest available Q-range, the Do model is only evaluated on this instrument to show that it also gives good results. Figures 5.5a-c show the results of $\langle r^2 \rangle_{\mathrm{Do,tot}}$ as defined in Eq. (2.52). They compare rather well with the results from the other three models. The evaluated values for the enthalpy ΔG and d are $6.1\mathrm{kJmol}^{-1}$ and $= 1.7\,\text{Å}$ (at 0.4h) respectively, similar to values found for myoglobin at 0.38h ($\Delta G = 12\mathrm{kJmol}^{-1}$, d = 1.5 Å) [2]. The values for the other hydration levels can be found in Table 5.4. Figure 5.6 shows the reduced χ^2_{red} for the different fits averaged over the entire temperature range.

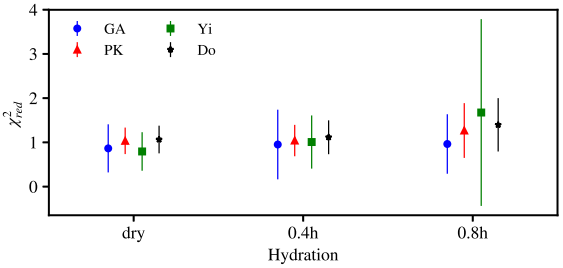


Figure 5.6. - IN13: Reduced χ^2_{red} for the fits, averaged over all temperatures. It is important to note that the reduced χ^2_{red} values are calculated with the respective Q-values used for the fit. Therefore, the GA and Yi model take less values into account. All four models have a χ^2_{red} value around 1, but the value for the Yi model at 0.8 hydration is around 2 and has a much bigger standard deviation. The reason for this was already mentioned in the results of the evaluated MSD. The fit of the Yi model is indeed worse at high temperatures and therefore also its χ^2_{red}. [*This figure was published in [33].*]

hydration	d[Å]	$\Delta H[\frac{\text{kJ}}{\text{mol}}]$	$\Delta S[\frac{\text{J}}{\text{mol K}}]$
dry	1.58 ± 0.16	4.7 ± 0.7	−2.7 ± 2.8
0.4h	1.73 ± 0.12	6.1 ± 0.7	7.7 ± 2.6
0.8h	2.28 ± 0.18	13.7 ± 1.3	36.5 ± 5.2

Table 5.4. - Parameters evaluated by the Do model: the distance between the two wells d, the change in enthalpy ΔH and entropy ΔS.

The results on IN13 lead to the conclusion that for the dry A-L_{dep} all models are equal since the system is still close to a harmonic system (no dynamical transition visible) and the heterogeneity does not influence the evaluated MSD. When the hydration increases the anharmonicity grows and quantitative differences can be seen between the models. Especially at 0.8 hydration the three models diverge at high temperatures, showing the influence of heterogeneity on the MSD. The quality of the fits is similar for all three models in their respective Q-range. Only in the case of the Yi model it is worse at high temperatures for the A-L_{dep} at 0.8h.

IN13 - only high Q-range

Taking advantage of the fact that IN13 covers a wide Q-range and following on from the discussion in section 5.2.3, fitting the high Q-range alone using the GA model is now considered to evaluate the effect on the dynamical transition temperature and the MSD. Specifically, the Q-range $1.7 - 4.5\,\text{Å}^{-1}$ is fit. The result is shown in Figure 5.7 and is compared to the GA at low Q-values in Figure 5.5d. The absolute value of the MSD from the high Q-range is lower by up to a factor of 5 for the 0.8h sample. This is a huge effect, but might be reasonable when separating motions of large and small amplitudes, for instance, atoms of the side chains of the amino acids and fluctuations within the backbone of the amino acids. The relative change in the MSD with increasing temperature between the different hydrations is however similar. All hydrations follow the same trend until

220 K and then deviate from each other. In the high Q-range, the 0.4h and 0.8h curves are then superposed until the second splitting due to the melting of heavy water, whereas for the small Q-range the MSD of the higher hydrated sample lies even below the 0.4h sample.

In the paper of Combet and Zanotti [97], the authors study a protonated protein hydrated with D_2O and the same protein in a perdeuterated form hydrated with H_2O on two different instruments, IN13 and the spectrometer MIBEMOL at the LLB in Saclay, France, with a resolution of $\approx 140\,\mu eV$. The corresponding short time window of around 10 ps reveals a weak dynamical transition, observed both for the protein and its hydration water. In contrast, the larger time window of IN13 permits a separation of the experimental data into large and local motions with a cross over at around $1.2\,Å^{-1}$; clear differences in the motions of water and protein molecules being visible since larger amplitude motions can be probed at low Q. Their approach of using H/D-contrast evidences directly the existence of two different dynamic populations. The findings here, without contrast variation, confirm that a division in two population is indeed reasonable.

Another approach in this same line of thought involves fitting the whole Q-range using a bimodal fitting model, where $EI(Q) \approx p_1 \exp(-\langle r^2\rangle_{large} Q^2) + p_2 \exp(-\langle r^2\rangle_{local} Q^2)$, with $p_1 + p_2 = 1$. It assumes the presence of two well-separated MSD defined within the GA $\langle r^2\rangle_{large} > \langle r^2\rangle_{local}$ in the sample, as proposed by Nakagawa et al. [27]. This approach works well for the protein staphylococcal nuclease with an instrumental resolution of 1 meV, obtaining $\langle r^2\rangle_{large} \approx 0.7\,Å^2$ and $\langle r^2\rangle_{local} \approx 0.15\,Å^2$ at 300 K.

In conclusion, the results suggest that by analysing the high Q range only on an instrument like IN13, one can access small amplitudes of motion in the proteins. These follow quite closely the temperature and hydration behaviour of the large amplitudes in case of hydrated powders.

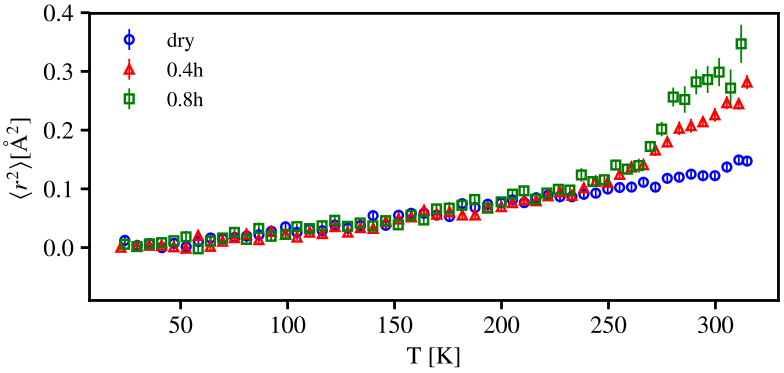


Figure 5.7. - IN13: MSD for all three hydration levels evaluated with the GA model at high Q range ($1.7 - 4.5\,Å^{-1}$). [*This figure was published in [33].*]

5.3.2. Analysis of data from SPHERES

For the SPHERES spectrometer the fitting Q-range is $0.34 - 1.8\,\text{Å}^{-1}$. The first two detectors at $Q = 0.34$ and $Q = 0.45\,\text{Å}^{-1}$ have a lower energy resolution [44]. To see the influence of this resolution effect two different low Q-ranges were fitted in which the GA is still valid, i.e. $\ln[\text{EI}(Q)]$ vs Q^2 linear for all temperatures. One which only includes the first three available Q-values $0.34 - 0.6\,\text{Å}^{-1}$(see section 5.3.2) and one excluding the first two Q-values with lower resolution, i.e. $0.6 - 1.2\,\text{Å}^{-1}$(see section 5.3.2). In both cases the PK and Yi model are fitted to the same Q-values as the respective fit of the GA and using in addition the larger available Q-values until $1.8\,\text{Å}^{-1}$. A third Q region is fitted to only high Q-values as it was done for IN13 in section 5.2.3. There, only the GA was used in the Q-range $0.95 - 1.8\,\text{Å}^{-1}$, neglecting the first four Q-values.

Low Q-range I (GA, $0.34 - 0.6\,\text{Å}^{-1}$)

In Figures 5.8a and 5.8b representative fits for three different temperatures are shown for the 0.4h and 0.8h A-L_{dep} samples. The fits show that it is not possible to include more Q-values for the GA if the small Q-values are to still be well described by the fit. A comparison to a larger Q-range which does not describe the lowest two Q-values is shown in section 5.3.2. The SPHERES spectrometer has ten times better resolution than IN13, such that larger motions are included and the resulting MSD becomes larger, leading to a lower Q-range accessible by the GA. These larger motions are possibly movements of the side chains of A-L_{dep}. Also small differences are visible between the samples in comparison to IN13.

The results of the fits of the $\text{EI}(Q)$ are shown in Figure 5.9 in the same way as for IN13. For all samples a linear increase of the MSD is visible until 200 K where the dynamical transition sets in.

For the dry sample the evaluated MSD is very noisy which can be explained by the low statistics implied in using only the first three Q-values. Higher Q-values could be included but in order to compare the same Q-ranges between the different hydration levels, the

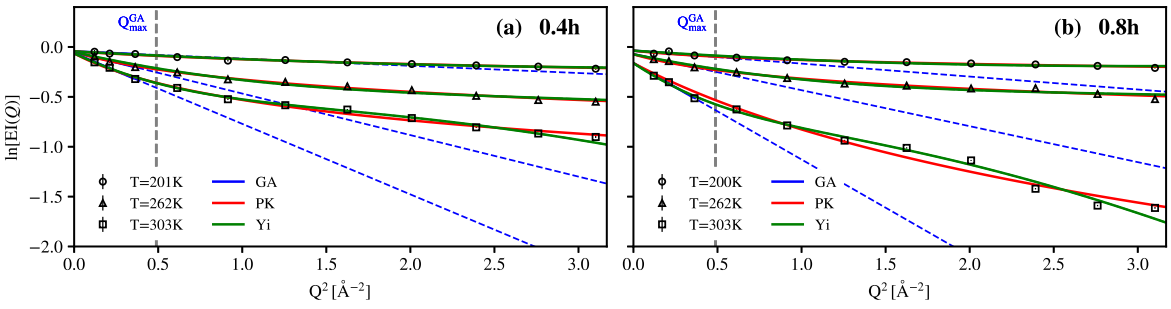


Figure 5.8. - Representative fits for the three models at three different temperatures for SPHERES data. Low Q-range I. Figure a) corresponds to 0.4h and Figure b) to 0.8h, respectively. The vertical gray line indicates $Q_{max} = 0.6\,\text{Å}^{-1}$ used for the GA. [*This figure was published in [33].*]

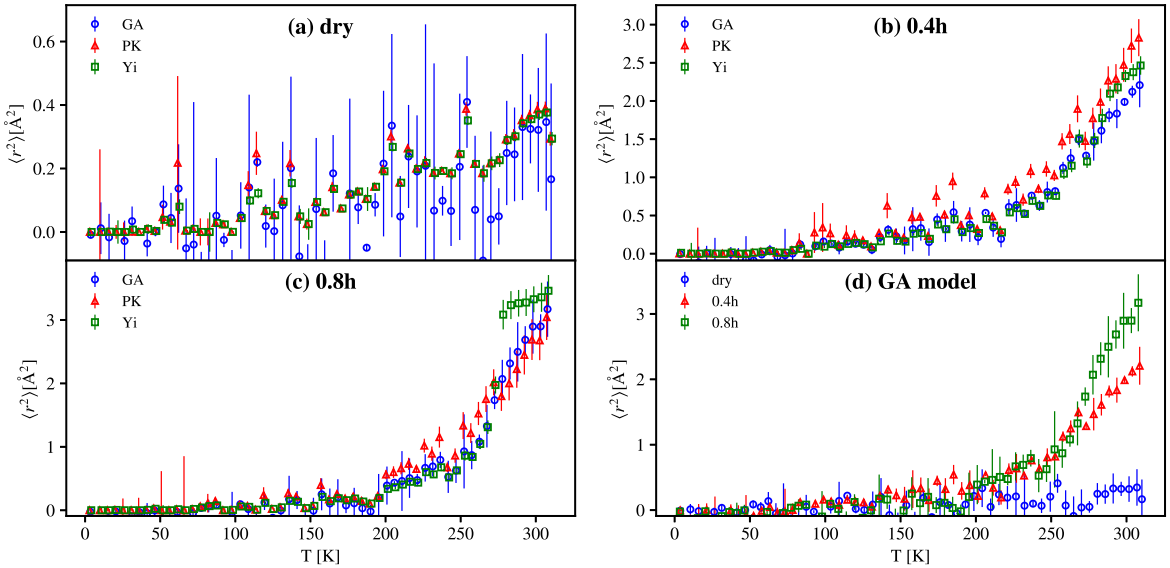


Figure 5.9. - SPHERES: Results of the MSD for all models. Low Q-range I. MSD values extracted from the GA, PK and Yi models for dry, 0.4 and 0.8 hydrated A-L_{dep} depleted samples (a to c); SPHERES data. Figure d) shows the MSD of all three hydration levels evaluated with the GA model. For the GA, only the first three Q-values ($0.34-0.6\,\text{Å}^{-1}$) were used like shown in Figure 5.8. [*This figure was published in [33].*]

Q-range evaluated for 0.8h is taken (for a larger Q-range see next section 5.3.2). The problem with using the first three data points only is emphasised by the small decay of the EI(Q) for the dry sample. The Yi and PK model fit the data nicely and with a similar goodness, resulting in similar values of MSD, and with a much better accuracy than the GA since they are including all experimental data points.
At 0.4h hydration the dynamical transition is visible around 200 K. The MSD of the PK and Yi model are higher than the MSD of the GA model following the trends seen on IN13. At 0.8 hydration two changes in slope are visible at $\sim$ 200 K and $\sim$ 270 K. The first change of slope is again attributed to the dynamical transition and the second to the melting of free heavy water which enhances the movements of the protein. The highest MSD values are around 3.5 Å^2. The PK model has the same MSD as the GA model, whereas the Yi model has higher MSD at temperatures above 270 K. This coincides again with the melting point of D_2O. The MSD from 270 to 280 K increases in the Yi model by almost 1.5 Å^2. This transition is smoother for the two other models. On the other hand, after the jump, the increase of the MSD in the Yi model is much slower in comparison to the other models. Therefore, at 310 K the three models reach a similar value in MSD. The jump of the Yi model can be explained by the counterbalancing between the Q^2 and Q^4 term (see Eq. (2.50)), which likely results in a mathematical but not a physical solution. Also the reduced χ^2_{red} statistics indicate that the fitting of the PK and Yi model for 0.8 hydrated samples is much worse than for the dry and 0.4 hydration samples (data not shown).

Low Q-range II (GA, $0.6 - 1.2\,\text{Å}^{-1}$)

In the section above, the GA was only fitted to the first three Q-values $[0.35 - 0.6\,\text{Å}^{-1}]$ to take into account the limit of validity. This leads to a good description of the data in the low Q-range, but also to a high statistical error. In addition, like stated before, the resolution is larger for the first two low Q detectors in comparison to the other detectors. If the GA is fitted in the range of $0.6 - 1.2\,\text{Å}^{-1}$, it leads to a more consistent linear fit with a smaller error. Examples for the fits are shown in Figures 5.10a and 5.10b. The resulting MSDs are shown in Figure 5.11, together with the Yi and PK model using the Q-range $0.34 - 1.8\,\text{Å}^{-1}$.

Figures 5.11a-c) summarises the three different hydration levels of A-L_{dep}.For all three hydration levels, the GA, PK and Yi model evaluate to similar MSD values. The values of the MSD for the dry protein are very similar between the two different Q-ranges of the GA. For the 0.4h and 0.8h hydration the MSD is lower by a factor of ≈ 2 compared to fits including the lowest Q-value. Figure 5.11d shows the results of the GA for all three hydration levels. A clear difference in the increase of the MSD is visible between the dry and the two hydrated samples at around 220 K. The two hydrated samples are then following the same pattern until around 270 K. At higher temperatures the 0.8h sample has a much larger MSD than the 0.4h sample. As explained before, the reason is the melting of frozen D_2O. Nevertheless, it is interesting that both curves have the same MSD until 270 K which could be due to the non frozen water shell around the protein being the same at both hydration levels.

The comparison between the two different Q-ranges at low Q for the GA (see Figures 5.9d and 5.11d shows that the quantitative value of the MSD is different depending on which range is chosen. The difference of a factor of 2 in the MSD for hydrated samples originates mainly from the difference in the value of the $\text{EI}(Q = 0\,\text{Å}^{-1})$. This is especially visible for the PK and Yi models. They both use the entire available Q-range at high Q, but fixing $\text{EI}(Q = 0\,\text{Å}^{-1})$ to the respective value obtained by the GA, changes their quantitative results dramatically.

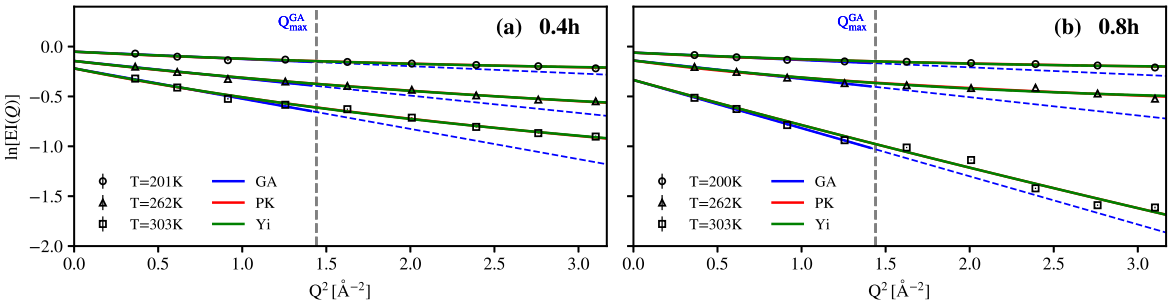


Figure 5.10. - Representative fits for the three models at three different temperatures for SPHERES data. Low Q-range II. Figure a) corresponds to 0.4h and Figure b) to 0.8h, respectively. The vertical gray line indicates $Q_{max} = 1.2\,\text{Å}^{-1}$ used for the GA. Here, the lowest two Q-values are neglected in contrast to Figure 5.8. [*This figure was published in [33].*]

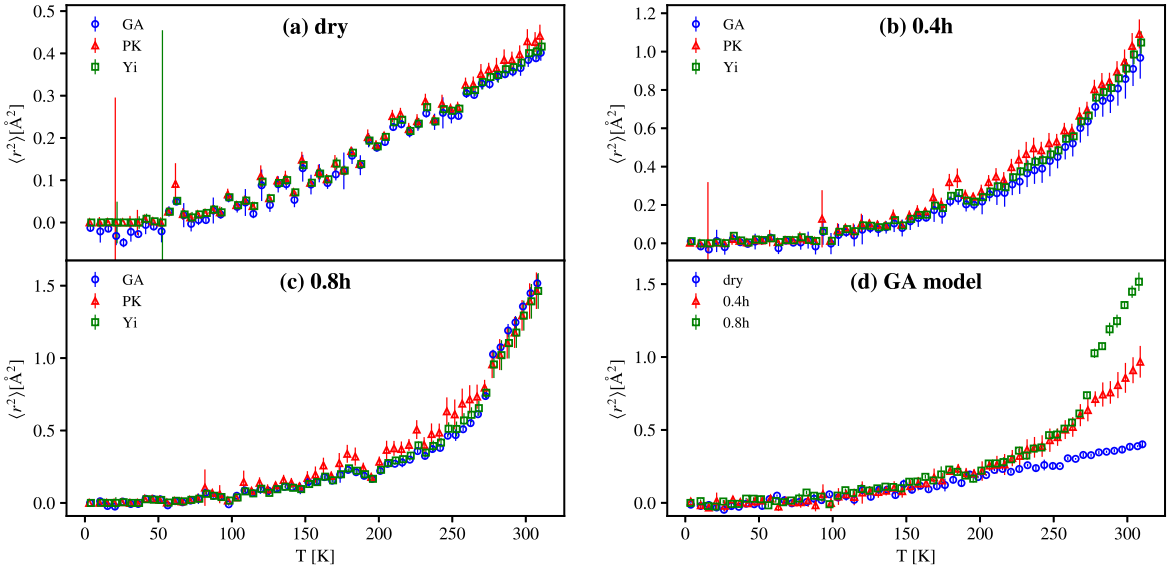


Figure 5.11. - SPHERES: Results of the MSD for all models. Low Q-range II. MSD values extracted from the GA, PK and Yi models for dry, 0.4 and 0.8 hydrated A-L_{dep} depleted samples (a to c); SPHERES data. Figure d) shows the MSD of all three hydration levels evaluated with the GA model. For the GA the Q-range $0.6 - 1.2\,\text{Å}^{-1}$ is evaluated, for the PK and Yi model the Q-range is $0.6 - 1.8\,\text{Å}^{-1}$. [*This figure was published in [33].*]

High Q-range for GA: $0.96 - 1.8\,\text{Å}^{-1}$

As shown in section 5.2.3, it is possible to describe only the high Q range of SPHERES using the GA. The fitted Q-range in which $\ln[\text{EI}(Q)]$ vs Q^2 is linear is $0.96 - 1.8\,\text{Å}^{-1}$. The resulting MSD are shown in Figure 5.12. Comparing to the low Q fits without the small angle detectors (Figure 5.11d), the MSD are smaller but only by up to a factor of 2 in the case of the 0.4h sample. More interesting are the changes in the behaviour between the different samples. Firstly, it appears that the dry powder shows a slightly larger MSD in the temperature range $150 - 240\,\text{K}$ compared to the hydrated samples which have the same MSD in this temperature range. Such an observation has been reported by Nickels et al. [98] for green fluorescent protein (GFP) (0.4 D_2O hydration vs. dry, Figure 2a in their publication), and measured on the very similar instrument, HFBS. They suggest that the frozen hydration shell reduces the MSD at low T by suppressing fast ps fluctuations. A similar behavior was also found in molecular dynamics simulations of GFP by Hong et al. [93]. Secondly, the MSD of the 0.8h sample is the same as for the dry sample between $240 - 270\,\text{K}$ and lower than for the 0.4h sample. It seems that the MSD is inhibited by the frozen D_2O for the 0.8h sample. Above 270 K the MSD of the 0.8h sample increases steeply. This again coincides with the melting point of heavy water at 278 K. Finally, in the inset of Figure 5.12 one can see that the dynamical transition still takes place around 200 K but since the dry sample has a higher MSD at lower temperature, the crossover between the hydrated and dry sample takes place at 250 K.

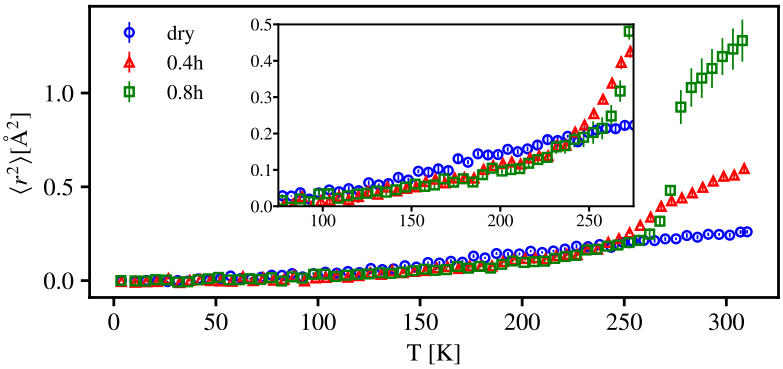


Figure 5.12. - SPHERES: Results of the MSD for all models at high Q range. MSD for all three hydrations evaluated with the GA model at high Q range ($0.96 - 1.8\,\text{Å}^{-1}$); SPHERES data. The inset shows a zoom of the low T region. [*This figure was published in [33].*]

5.3.3. Analysis of data from OSIRIS

For the OSIRIS spectrometer, the chosen Q-range is $0.29 - 1.5\,\text{Å}^{-1}$. Note that OSIRIS has more detectors that cover this Q-range as compared to the other two spectrometers, allowing for better Q-resolution. The instrumental time resolution is three times larger than on IN13 such that faster motions up to around 25 ps can be probed. This may explain why it is possible to use all three models until $1.5\,\text{Å}^{-1}$. Mainly small localised movements are observed and not larger side-chain motions. The resulting values of the MSD are thus small in comparison to IN13 and SPHERES. Representative fits of the EI(Q) for the two hydrated samples are shown in Figure 5.13.

The MSD for the different models are illustrated in the same way as for IN13 and SPHERES in Figure 5.14. For each hydration, the GA, PK and Yi model yield almost the same MSD values. At 310 K the PK and Yi model evaluate slightly higher MSD values for the dry and 0.4 hydration samples and for the 0.8 hydration sample above 280 K. This behaviour confirms that all models give similar results if they use the same Q-range on this instrument. The dry sample shows a linear increase of the MSD with increasing temperature. For the 0.4 and 0.8 hydration the dynamical transition is visible and starts around 250 K, similar to that found on IN13.

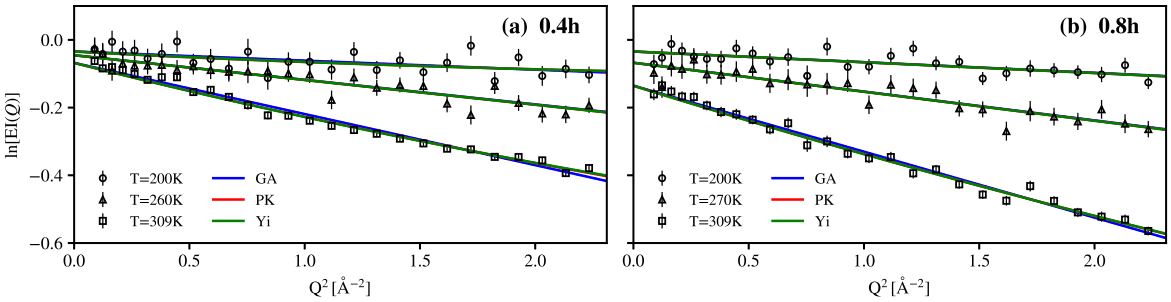


Figure 5.13. - Representative fits for the three models at three different temperatures for OSIRIS data. Figure a) corresponds to 0.4h and Figure b) to 0.8h, respectively. [*This figure was published in [33].*]

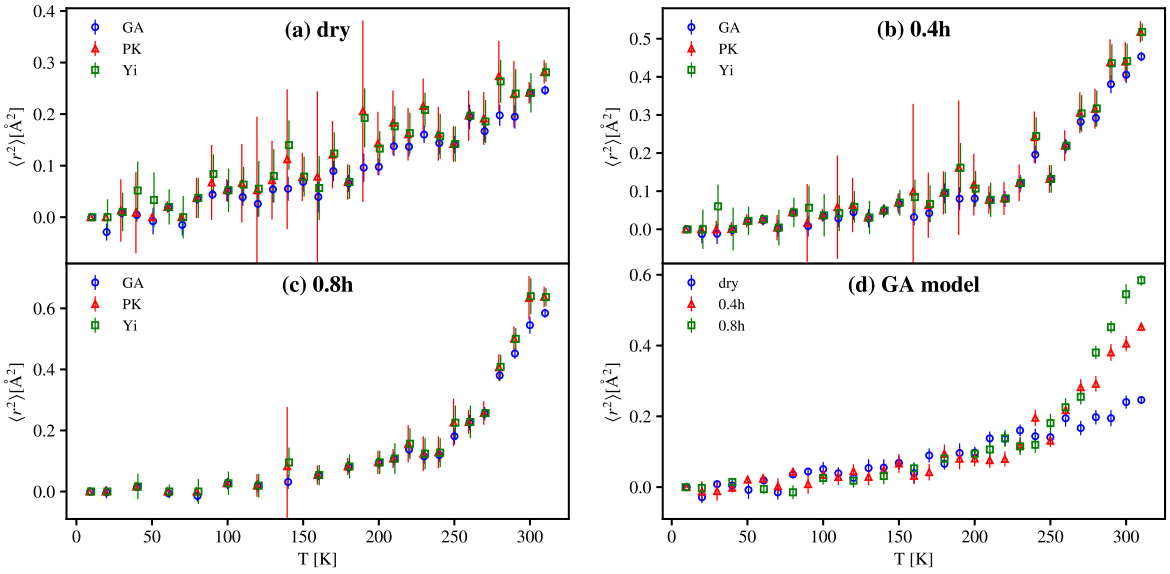


Figure 5.14. - OSIRIS: Results of the MSD for all models. MSD values extracted from the GA, PK and Yi models for dry, 0.4 and 0.8 hydrated A-L_{dep} samples (a to c); OSIRIS data. Figure d) shows the MSD of all three hydration levels evaluated with the GA model. All models use the same Q values ($0.29 - 1.5\,\text{Å}^{-1}$). [*This figure was published in [33].*]

5.3.4. Comparison between summed intensities and MSD

It should be noted that the behaviour of the MSD with temperature is not always the best way to look at data when small differences are expected between samples. In such cases the comparison of the intensities summed over all (or a range of) available scattering angles, I_{sum}, can be much more insightful. If required, in the limit of the GA they can moreover be related to the MSD (see Eq. (2.59)). In Figure 5.15, the I_{sum} for the three different instruments and hydration levels is compared to the MSD evaluated by the PK model using the same Q-range. The results demonstrate the inverse hierarchy for the I_{sum} and the MSD between the samples. As the error bars of the summed intensities are much smaller, it allows to better separate the curves and the differences between the hydration levels become more visible in the summed intensities. For instance, the kink for the 0.8h sample at $\approx 275\,\text{K}$ clearly indicates the melting of ice. Moreover, one distinguishes on IN13 that in the case of I_{sum} (see Figure 5.15b) the curve corresponding to the dry sample lies below the curves of the hydrated samples in the temperature domain from 150 to 210 K. This effect was discussed earlier in section 5.3.2. This behaviour cannot be observed in the MSD (see Figure 5.15e) where this information is lost due to data fitting. However, for the MSD evaluated on SPHERES, a small bump at 180 K can be seen for the hydrated A-L_{dep} samples (see Figure 5.15f), which is not visible in the I_{sum}(see Figure 5.15c) and which shows up also when applying the GA (see Figure 5.11d). It might be a real effect, eventually due to an ice phase transition [99], which appears only through the evaluation of the slopes. Both data analyses might therefore furnish complementary information.

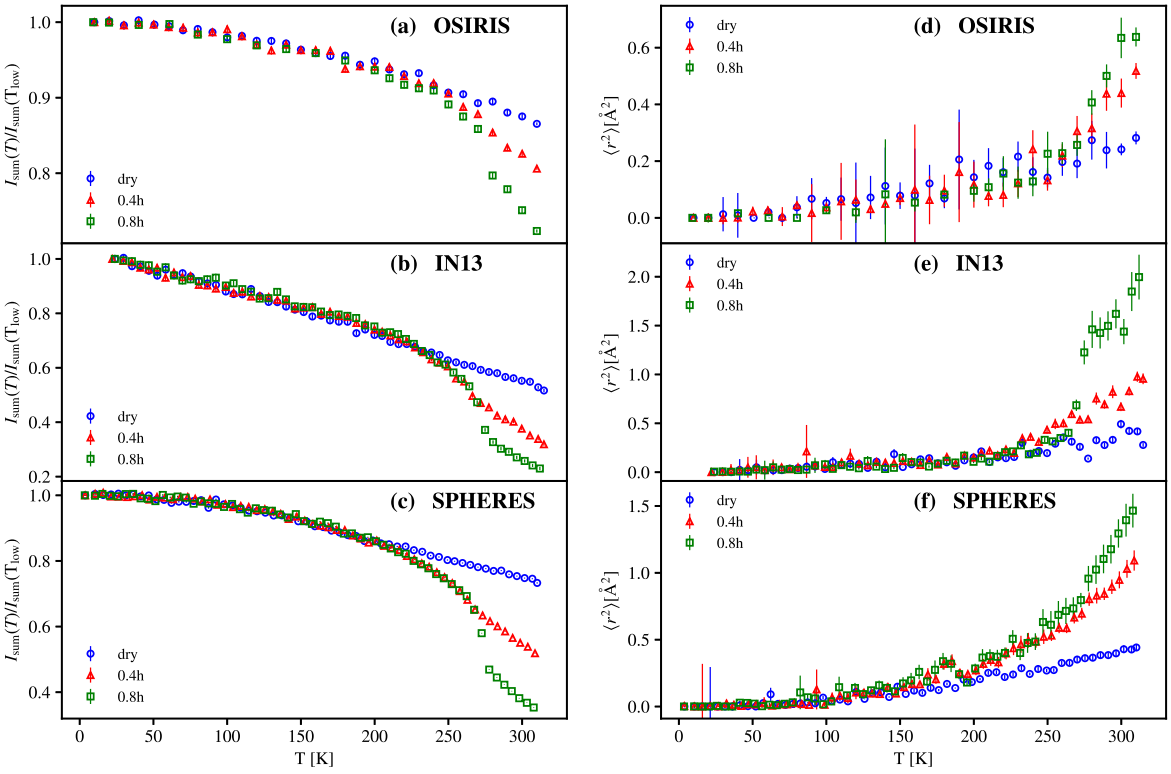


Figure 5.15. - Comparison: Summed intensities vs. MSD (PK model). Left: Summed intensities over the entire available Q-range of each instrument. Right: Respective evaluated MSD within the PK model over the same Q range. [*This figure was published in [33].*]

5.3.5. The value of EI(Q = 0)

The value of the resulting elastic intensity at $Q = 0$ is discussed in what follows. In Figures 5.16a,b,c the EI($Q = 0\,\text{Å}^{-1}$) is shown for the evaluated low Q ranges for OSIRIS, IN13 and SPHERES. It is not possible to find a clear relation between the hydration level and EI(0), but for all samples it is slowly decreasing at higher temperatures. As presented in Table 4.4 in section 4.3 the A-L_{dep} for IN13 is not exactly the same sample as for OSIRIS and SPHERES (B1 vs. B2). The samples in B2 contain $\approx 10\%$ more dry protein powder than B1 but the hydration levels are the same (except for 0.8h where B1 has 0.85h instead). Looking only at OSIRIS and SPHERES data (B2) gives different tendencies. For OSIRIS the 0.8h sample has the lowest normalised intensity EI(0) around 0.90 at 300 K and the dry and 0.4h samples have more or less the same value around 0.95. In contrast, in the case of SPHERES the 0.4h sample has also a value of 0.95, but the dry and 0.8h samples have a similar EI(0) around 0.87 at 300 K. At IN13 (B1) no clear difference between the three hydration levels is visible.

In comparison in Figure 5.16d the EI(0) is shown for the larger 'low Q range' without the first two detectors (see sub-section 5.3.2). There, a clear difference between the three hydration levels is noticeable. The curves look indeed very similar to the summed intensities in Figure 5.15. This again confirms the importance of the chosen Q range for

the evaluated results. In addition, it shows that on SPHERES the intensity loss is large in the low Q regime and thus a small change in the Q range has a large impact on the EI(0) which can influence the evaluated MSD. Therefore it might be reasonable to also report the EI(0) for evaluated data sets to have a better comparison between samples, and not only report the MSD.

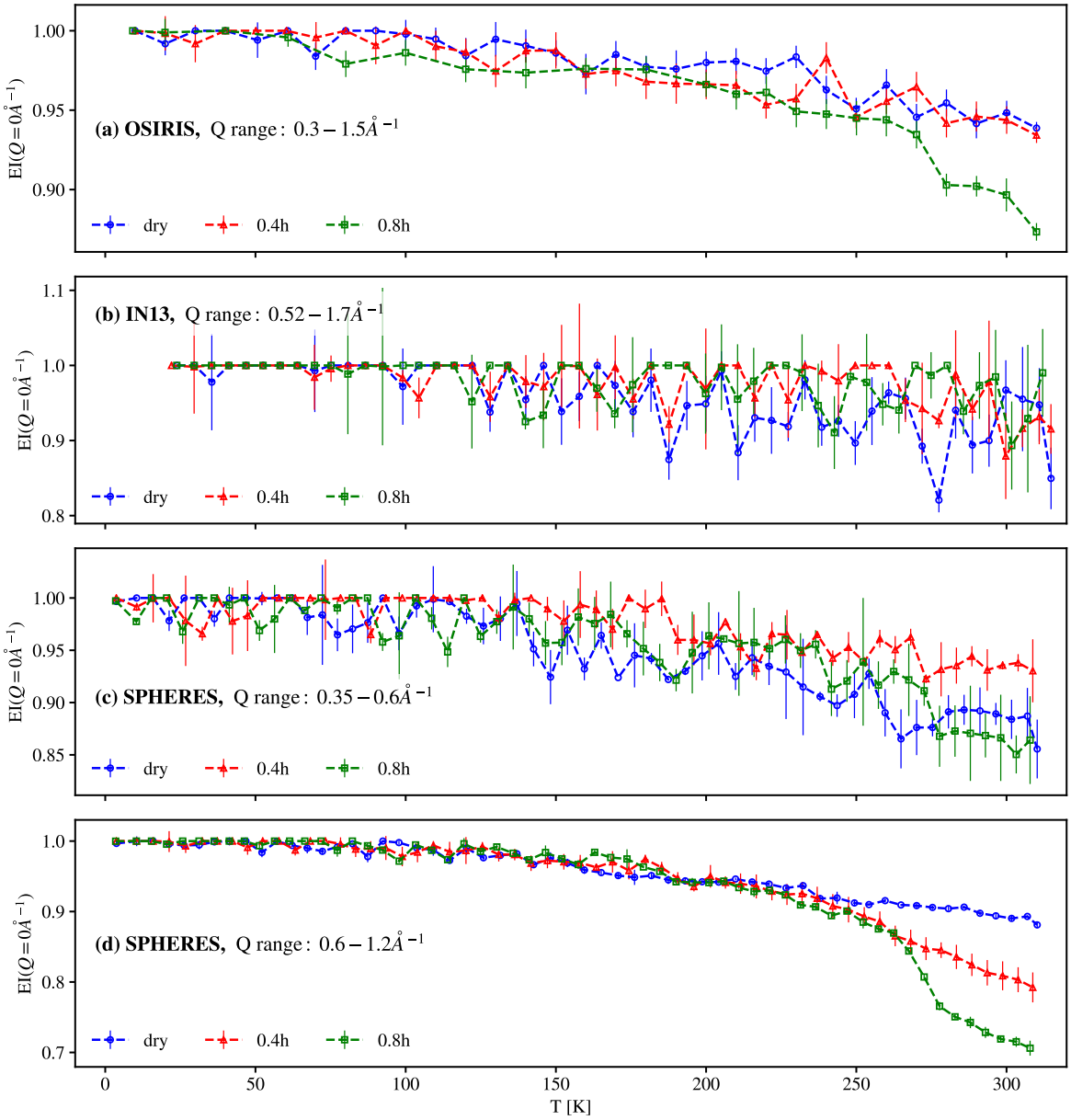


Figure 5.16. - The value of EI(Q = 0) for all three instruments. (a-c) EI(0) evaluated with the GA model for the low Q range fits and in (d) EI(0) evaluated for SPHERES at the larger 'low Q range' (0.6 – 1.2 Å−1).

5.4. Comparison between A-L_{dep} and A-L_{ca}

In this section, the results of A-L_{dep} discussed up to now will be compared to those obtained on A-L_{ca} on the same instruments. The data analysis and respective Q-ranges (at low Q values) and hydration levels are the same as used for A-L_{dep}. First the MSD for the three different models are compared for each instrument and then the summed intensities. For SPHERES the Q range $0.6 - 1.2\,\text{Å}^{-1}$ for the GA model and $0.6 - 1.8\,\text{Å}^{-1}$ for the PK and Yi models was chosen. Figure 5.17 shows the MSD of all A-L samples, sorted by instrument (rows) and models (columns).

No clear differences between the A-L_{dep} and A-L_{ca} samples measured on OSIRIS (1st row) are visible for all three hydration levels. The dry samples are the same (blue curve) within the statistics and no trend is visible. The samples at a hydration level of 0.4 (red curve) are also almost the same within the statistics. Only for the last three temperature values ($290 - 310\,\text{K}$) small differences are visible and in all three models the A-L_{ca} sample has a slightly smaller MSD, indicating less dynamics, but within the statistics this effect is not conclusive. The same can be said for the sample at 0.8h (green curve).

The data for IN13 (2nd row) shows similar results compared to OSIRIS. No difference for the dry protein samples and small differences for the 0.4h samples. Here again, the A-L_{ca} sample has slightly larger MSD values in all three models at temperatures above 280 K, but within the statistics no clear statements can be made. Nevertheless, the A-L_{ca} values are mainly above the A-L_{dep} MSD values, indicating that there is a small difference between the two samples. For the 0.8h no data could be measured for the A-L_{ca} sample at IN13 and thus no MSD could be analyzed.

For SPHERES (3rd row) clearer differences between the samples are obtained. The A-L_{ca} sample at zero hydration has a slightly higher MSD than the A-L_{dep} sample above 250 K. This difference is emphasised in both models describing a larger Q range, the PK and Yi model, suggesting that the A-L_{ca} sample is more dynamic. At 0.4h no difference between the samples is observed and no clear trend is visible over the entire temperature range. For the highly hydrated samples (0.8h) the MSD is larger for the A-L_{dep} above 270 K. This is the case for all three models; in the GA model there is a clear difference, whereas for the PK and Yi models the difference is withing the statistics of the fitting routine. Nevertheless the trend is the same and the A-L_{dep} sample always has a larger MSD. Thus SPHERES indicates that with regards to the nanosecond motions, in dry conditions, the A-L_{ca} is more dynamic than A-L_{dep}, whereas at higher hydration levels it is less dynamic than the A-L_{dep}. This indicates a stabilisation of the Alpha-Lactalbumin with calcium ion binding.

The example of A-L shows that often the GA model is enough to distinguish differences between similar samples, but there are cases where the models describing the intensities at higher momentum transfers Q can give more or clearer information, e.g. the dry A-L samples for SPHERES.

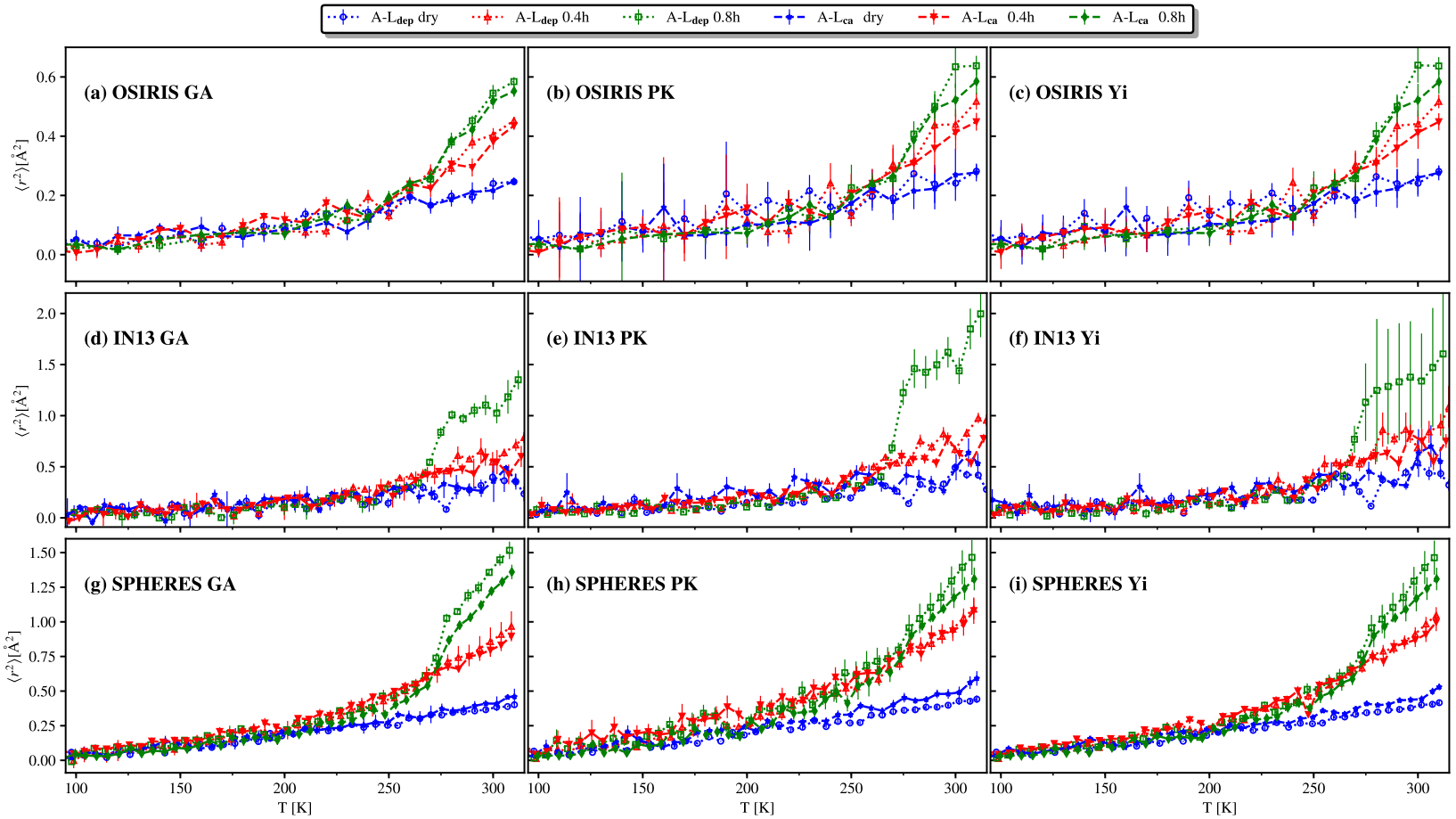


Figure 5.17. - Difference of MSD between A-L$_{dep}$ and A-L$_{ca}$ for OSIRIS, IN13 and SPHERES. The three rows are sorted by instrument from top down: OSIRIS, IN13 and SPHERES. The columns are sorted by model from the left to right: GA, PK and Yi model. All data is shown from 100 K to 310 K. The values under 100 K are all the same within the statistics and are not shown. The MSD values of the A-L$_{dep}$ (A-L$_{ca}$) samples are connected by dots and have open symbols (dashes, filled symbols). The dry samples are blue, the 0.4h samples are red and the 0.8h samples are green.

Like for A-L_{dep}, the MSD of the PK model is compared to the summed intensities. Figure 5.18 depicts the summed intensities on the left side and the MSD on the right side including both the A-L_{dep} and A-L_{ca} sample from 200 K to 310 K (for the entire temperature range $3 - 310$ K see Figure 5.15). As before the used Q ranges are the same for the MSD and the summed intensities. As discussed in sub-section 2.2.5 in the limit of the GA the summed intensities are inversely proportional to the MSD. Here this is not the case since the intensities were summed over the entire available Q range and not only in the validity of the GA. Nevertheless, it is normally expected that the summed intensities are qualitatively the same as the MSD (inversely proportional to the MSD). In short, if the summed intensities of A-L_{dep} are smaller than the summed intensities of A-L_{ca}, the MSD should be larger. For IN13 (Figures 5.18b,e) the qualitative differences between the A-L_{dep} and A-L_{ca} samples are similar. In the case of OSIRIS (Figures 5.18a,d) a small but clear difference between the samples is visible which is not the case in the MSDs which have a larger statistical uncertainty. Also looking at the trends of the temperature

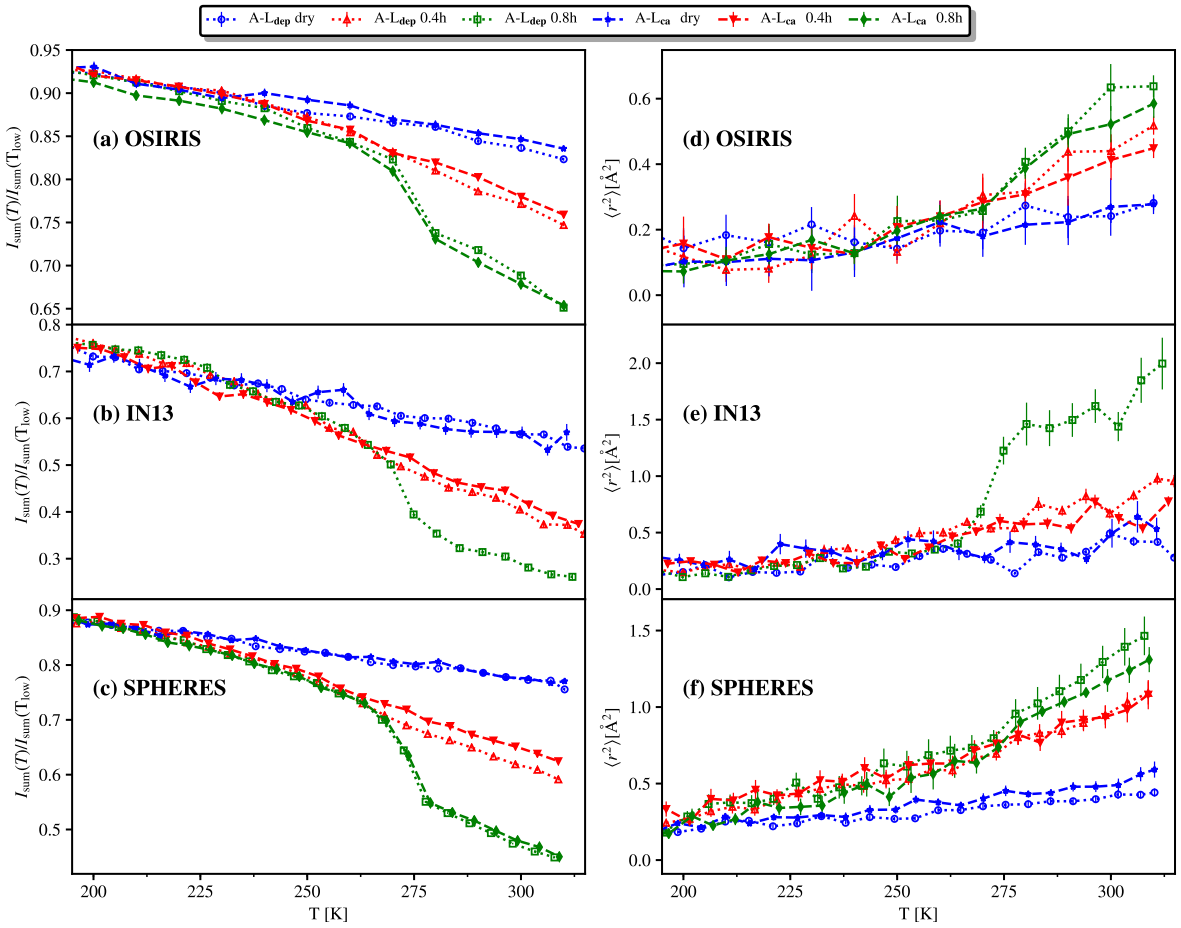


Figure 5.18. - Comparison: Summed intensities vs. MSD (PK model) between A-L_{dep} and A-L_{ca} for OSIRIS, IN13 and SPHERES. Left: Summed intensities over the entire available Q range of each instrument. Right: Respective evaluated MSD within the PK model over the same Q-range. The values are shown in the temperature range 200 K to 310 K (for the entire temperature range $3 - 310$ K for A-L_{dep} see Figure 5.15).

development of the MSD, the summed intensities are not showing the same results as the MSD. In the case of the summed intensities, A-L_{dep} has a smaller intensity for the dry and 0.4h hydrated sample, whereas in the MSD the A-L_{dep} has the trend to surpass the A-L_{dep} sample slightly. But within the statistics no real discrepancy can be inferred. For SPHERES (Figures 5.18c,f), a clear discrepancy between the summed intensities and the MSD are shown. In the case of the summed intensities the 0.8h samples are almost the same and for the MSD the A-L_{dep} sample has a larger MSD. At 0.4h hydration level the summed intensities show a significant difference between the two samples, whereas for the MSD the values are the same. In the dry case, the summed intensities are the same and the MSD of A-L_{ca} are larger.

To investigate the reasons behind these discrepancies the 0.4h hydration samples are looked at in more detail in the following. A reason for the difference in the summed intensities and the MSD must be the starting point of the fit at EI(0). Thus Figure 5.19a depicts the EI(0) for the SPHERES instrument. In the case of the 0.4h sample, EI(0) is significantly smaller for A-L_{ca} than for A-L_{dep}. For the other two hydration levels, EI(0) is almost the same. Inspecting the reduced and normalised raw data, also a large intensity gap between the two 0.4h samples is apparent in EI(Q). It seems that even having a different absolute intensity, the EI(Q) of A-L_{ca} and A-L_{dep} have a similar slope evaluating to almost the same MSD. To prove this an example is shown in Figure 5.19b, depicting the fits and resulting MSD of the GA and PK models at 293 K for A-L_{ca} and A-L_{dep}. Here, the figure is plotted for ln[EI(Q)] vs. Q^2. The difference in ln[EI($Q = 0$)] is striking and for the GA model the slopes of the linear fits are almost the same indicated by the almost parallel blue lines. The evaluated MSD are $\langle r^2 \rangle^{ca}_{GA} = 0.81 \pm 0.09\,\text{Å}^2$ and $\langle r^2 \rangle^{dep}_{GA}\, 0.77 \pm 0.03\,\text{Å}^2$. At higher Q values the intensity difference is larger, but the non Gaussianity part behaves similar to the Gaussian part, giving rise to very similar MSD for the PK model with $\langle r^2 \rangle^{ca}_{PK} = 0.92 \pm 0.07\,\text{Å}^2$ and $\langle r^2 \rangle^{dep}_{PK}\, 0.90 \pm 0.05\,\text{Å}^2$. Figure 5.19b also shows the points which have been excluded (in gray). As already discussed before, they

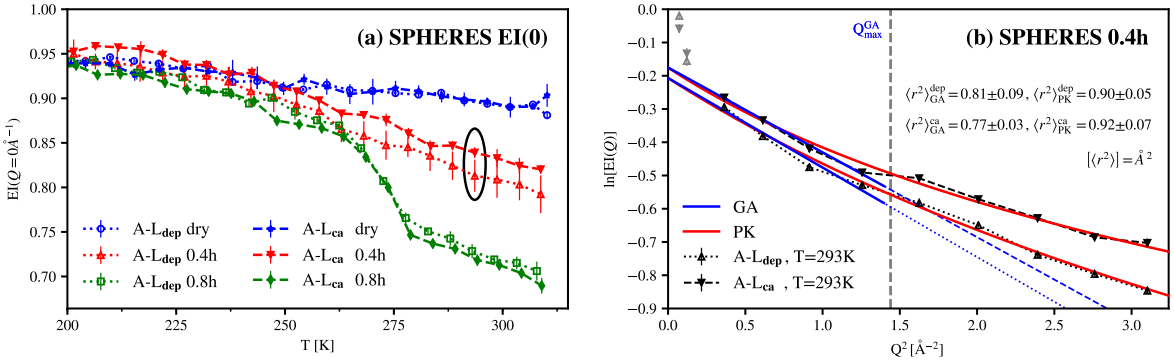


Figure 5.19. - EI(0) of A-L_{dep} and A-L_{ca} for SPHERES and example of fit differences for the 0.4h sample. (a) Evaluated values of EI(0) with the GA model for SPHERES instrument and both A-L samples. A-L_{dep} (A-L_{ca}) has open (closed) symbols and has dotted (dashed) lines. (b) Example of ln[$EI(Q)$] vs. Q^2 fits with GA and PK model for the A-L_{ca} and A-L_{dep} samples at 0.4h - both at 293 K indicated by black circle in a). In gray the two neglected Q values in the low Q range are shown.

are much larger in intensity than the other Q values. There, a larger decay for A-L_{dep} sample is visible (open symbols) which is in agreement with the lower intensity of A-L_{dep} at higher Q values.

5.5. Summary and Conclusions

A number of models were tested and the resulting MSDs were evaluated for A-L_{dep} and A-L_{ca} at different hydration levels, each measured on three instruments. It becomes clear that the intercept of EI($Q = 0\,\text{Å}^{-1}$) is very important in fitting the data to any given model. Fixing it to the theoretical value of EI($Q = 0\,\text{Å}^{-1}$) = 1 is often not possible for the GA, as shown in section 5.2.1, nor for the models that take into account a larger fitting Q-range. Since all models should converge in the limit of Q $\rightarrow$ 0 to the same value, but can present significant variations when leaving it a free parameter, it is advised to fix EI($Q = 0\,\text{Å}^{-1}$) to a common value for the same data set. Here, this value was obtained by fitting the data with the GA model. With such settings the qualitative results between the models are similar, concerning the dynamical transition and dynamical changes as a function of hydration level, which leads to the conclusion that the simplest and most commonly used model, the GA, is indeed a good standard, if a consistent protocol is followed, i.e. respecting a linear behaviour of ln[EI(Q)] vs Q^2 and comparing the same Q-ranges.

The inclusion of the smallest Q-values in traditional reactor-based neutron backscattering spectrometers can change the quantitative MSD dramatically as shown on SPHERES, i.e. resulting in a difference of a factor of two in the MSD. As mentioned already, the first low angle detectors on SPHERES have a poorer resolution than the other detectors, but this is unlikely the only reason for the large differences in MSD. As the temporal range of SPHERES allows to access larger amplitudes of motion, it could also arise from the onset of movements of molecular sub-groups. If the same Q-range is respected when comparing hydration levels or different samples, the GA is an adequate model and the inclusion of non-Gaussian terms is not needed in order to have a good estimation on the dynamics. In contrast, it is shown that for a quantitatively accurate analysis the MSD can be quite different between models, especially at a higher hydration level. In fact, even the PK and Yi models do not provide much higher precision, mainly because they are too dependent on the value of EI($Q = 0\,\text{Å}^{-1}$). If the experimenter is expecting small differences in dynamics between samples, then the extended models could give more accurate values for the MSD since they include information over a larger Q-range. However, the data will be subject to larger errors than the GA since they also have more free parameters to fit, as it can be seen from the MSD on OSIRIS (see section 5.3.3). The comparison of the summed intensities I_{sum} may be more helpful to determine small differences and they can deviate from the behaviour of the MSD even when both use the same raw data and Q-ranges (see section 5.3.4).

Looking at the comparison between A-L_{dep} and A-L_{ca} small changes in the MSD could be revealed on the SPHERES spectrometer, whereas for OSIRIS and IN13 the changes were not conclusive within the statistics. It seems that the depleted A-L is becoming more dynamic than A-L_{ca} at higher hydration levels. In contrast, for the dry protein,

A-L_{dep} appears slightly less dynamic, although the difference in the MSD is not significant, especially for IN13 and OSIRIS. The analysis also showed that the summed intensities and MSD values evaluated by the PK model are not always in agreement and especially on SPHERES there was a clear difference between the A-L_{dep} and A-L_{ca} samples at 0.4 hydration level. The reason was the value at EI($Q = 0\,\text{Å}^{-1}$). The two samples have different absolute intensities but the slope above $Q = 0.6\,\text{Å}^{-1}$ is the same. Therefore the main difference in the dynamics must be at very low Q values which cannot be analyzed properly with only two data points and which suffer from having a different resolution. This emphasises the importance of the evaluated Q range and the value at EI($Q = 0\,\text{Å}^{-1}$).

As the EINS method is often applied to compare samples of a given type, which have, similarly to A-L_{dep} vs. A-L_{ca} only minor differences (enzymes in presence or absence of an inhibitor[100], a wild-type biomolecule vs. a mutant[101], proteins at different concentrations[102] or with various co-solutes[103] etc.), the comparison of such a data set requires highest consistency concerning:

- use of the same hydration level
- use of the same instrumental resolutions
- use of the same Q-values for extraction of MSD or summed intensities
- setting of EI($Q = 0\,\text{Å}^{-1}$) to the same value if comparing different models for the same data set.

In conclusion, this investigation demonstrates that, despite many efforts to improve the quantitative results for the MSD, significant questions remain, and it was not possible to establish a reliable method on how to treat the data to get results with highest accuracy or precision. However,the analysis was able to show that a prevailing, although often ignored, point is the correct treatment of the point at EI($Q = 0\,\text{Å}^{-1}$). Interestingly, this study gives hints that in some cases a bi-modal approach might be sufficient and helpful to distinguish smaller and larger motions as already suggested by e.g. Nakagawa et al. [27], Doster and Settles [104] or Combet et al. [97] in the past. On the one hand, such treatment could help to distinguish large motions, which are sometimes associated with movements of hydration water if H_2O is used or with local translational diffusive displacements, whereas smaller motions account for localized dynamics or vibrations within the biomolecules. Doster very recently also showed that a bi-modal treatment of data taken on myoglobin clearly identifies at least two molecular processes which might be sufficient to describe the neutron scattering spectra of proteins[25]. On the other hand, as discussed in the present work, if a clear separation of linear regimes is no longer appropriate, the models using a continuous distribution of individual MSD describe the dynamics very well without further assumptions.

For a reliable conclusion, the exact knowledge of EI($Q = 0\,\text{Å}^{-1}$) is of utmost importance, but most instruments do not give direct access to this information. Three important effects could contribute to the measured EINS intensity: multiple scattering, coherent scattering contributions and neglecting QENS effects. From the literature [73] it is well known that multiple scattering effects are present and can play a role as a constant background [25, 78], but how large this influence actually is and if it is Q independent is not well documented. In most cases the effects are deemed negligible if the transmission of the sample is above 90% (see section 4.3). The contribution to the overall signal from coherent scattering

was shown by Gaspar et al. [17] for D_2O hydrated myoglobin. For most of the Q-range covered by backscattering neutron spectrometers, it manifests as a flat background in protein powders, but at low Q values it increases dramatically especially for hydrated powders. This may also rise to problems for the investigation of very low Q values and probably spin polarisation techniques are needed to distinguish between both distributions for more accurate measurements of EI(0). Finally neglecting the influence of the QENS signal was discussed by Doster et al. [20] and for elastic scans with low mobility samples the contribution is expected to be small. It is possible to correct for it but requires QENS measurements at each temperature step and thus long measuring times.

In the future, further studies by spin-echo spectroscopy, which provides access to Q-values as small as $10^{-2}\,\text{Å}^{-1}$ (e.g. IN15[1] at the ILL), together with a polarisation analysis to separate the coherent and incoherent scattering signals should be performed to get a more accurate value of EI(0). In addition, a careful analysis of the multiple scattering contribution has to performed, e.g. via Monte Carlo methods. This approach would also help to determine the exact reasons for the deviations of $\text{EI}(Q = 0\,\text{Å}^{-1})$ from 1 and how this deviations can be estimated. In addition, it may enable more quantitative precision in determining the values of MSD in the future.

[1] https://www.ill.eu/fr/users-en/instruments/instruments-list/in15/characteristics/, accessed 25.07.2018

6. Simulations of Alpha-Lactalbumin protein powder

In this chapter simulations of bovine A-L with (A-L_{ca}) and without calcium (A-L_{dep}) are compared to the experimental data obtained on OSIRIS, IN13 and SPHERES that have been discussed in the previous chapter. The simulations will be analysed by two different methods: 1) The MSD will be directly calculated through the movements of the H atoms from the simulations and 2) the MSD will be calculated by applying a neutron scattering experiment analysis, i.e. calculating the resolution broadened dynamic structure factor (DSF), $S^{\mathrm{R}}_{\mathrm{inc}}(\mathbf{Q},\omega)$. From this quantity, the EISF can be calculated by integration over the elastic line and the MSD can be extracted like for the experimental data.

In the first section 6.1 the basics of molecular dynamics simulations are explained. The second section gives an overview of the simulated system and the methods applied to extract the MSD of the H atoms. In the third section the results from the simulations are shown and in the fourth section they will be compared to the experimental data. Finally, the results will be summarised and discussed.

6.1. Molecular dynamics simulations - An overview

This section gives a brief introduction to empirical molecular dynamics (MD) simulations. As basis the review of M.A. González [105] was used and is also recommended as reference for more detailed information.

MD simulations are an important tool to investigate biological systems. They link experimental biology with analytical theory by using numerical approximations of biological systems. The calculations are based on physical models and can give new and testable insights into intra- and intermolecular movements [106]. Here, only classical molecular mechanical (MM) simulations will be discussed. They are based on empirical force fields (FF) which describe the interactions between atoms with classical mechanics. Also quantum mechanical (QM) simulations or a combination of MM and QM simulations can be performed but they have a much higher computational cost. They are needed if for example, the formation or breaking of bonds, electron excitation or charge transfers are important to describe the system properly.

There are several programs which can perform MD simulations, e.g. Charmm [107], Amber [108], NAMD [109] or Gromacs [110]. To start a MD simulation the initial coordinates of the atoms in the system must be defined, as well as details on the bonding between them (molecules). To evaluate the dynamics in the system, the forces between the different atoms and/or molecules have to be calculated. The involved forces are all

described in the force field (FF). In general, the FF is a mathematical expression that describes the dependency of the energy of a system on the coordinates of the atoms. Therefore it is the most important part of the simulation since it defines all possible interactions between the atoms. A FF describes two kind of interactions: bonded (intra-molecular) and non-bonded (inter-molecular) and typically consists of at least of the following terms [105, 111]

$$E = \sum_{\text{bonds}} k_b(b - b_0)^2 + \sum_{\text{angles}} k_a(\Theta - \Theta_0)^2 + \sum_{\text{dihedrals}} \frac{V_\Phi}{2}(1 + cos(n\Phi - \delta)) \tag{6.1}$$

$$+ \sum_{\text{improper}} \frac{V_\Psi}{2}(\Psi - \Psi_0)^2 + \sum_{elec} \frac{q_i q_j}{\epsilon\, r_{ij}} + \sum_{LJ} \epsilon_{\text{LJ}} \left[\left(\frac{r_{\min}}{r_{ij}} \right)^{12} - 2 \left(\frac{r_{\min}}{r_{ij}} \right)^{6} \right] \tag{6.2}$$

The first term describes the energy for the stretching of a bond of length $b = |\mathbf{r}_i - \mathbf{r}_{i+1}|$, between two atoms i and $i+1$, around an equilibrium value b_0. The second term describes the energy due to angle bending Θ between 3 bond atoms. The third and fourth term describe the energy due to the dihedral (torsional) angle Φ and improper angle Ψ between 4 bond atoms. k_b and k_a are force constants, V_Φ and V_Φ energy potentials and n is the multiplicity or periodicity of the dihedral angle and δ is the phase shift. The last two terms are the non-bonded terms, describing the electrostatic Coulomb interactions and the Van der Waals interactions with a 12-6 Lennard-Jones potential. ϵ_{LJ} is the Lennard-Jones well depth and $r_{\min}$ is the distance at the Lennard-Jones minimum, q_i is the partial atomic charge, ϵ is the effective dielectric constant, and r_{ij} is the distance between atoms i and j. Common empirical FFs for biomolecules are for example AMBER [112], CHARMM [111, 113, v.27] [114, 115, v.36], GROMOS [110] or OPLS [116]. They are based on the energies above but can include additional terms. Their parameter sets for each energy term can be different, too. Thus they may influence the results of a MD simulation significantly.

An important ingredient for the simulation of biomolecules is the surrounding water. There are a vast amount of different water models, each with their particular strengths and weaknesses [117]. Some of the most typical water models are TIP3P, SPC/E, TIP4P and TIP4P/2005, the latter giving the best agreement with real water according to Vega et al. [118]. However some FFs are optimised for specific water molecule models and the interactions also depend on the simulated system, e.g. TIP3P for the CHARMM FF [115].

An empirical MD simulation produces trajectories for a system with N atoms based on Newton's equation of motion. To start, the position and velocity of each atom i has to be known (initial condition). Then the forces acting on each atom can be calculated by the differentiation of the potential energy $U(\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_N)$

$$m\frac{d^2\mathbf{r}_i}{dt^2} = \mathbf{F}_i = -\frac{\delta}{\delta \mathbf{r}_i} U(\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_N). \tag{6.3}$$

Since U is dependent on the coordinates of all N particles, the above formula is a system of N coupled second order non-linear differential equations and cannot be solved exactly.

Thus the dynamics over time have to be discretized in small time steps of size Δt

$$\mathbf{r}_i(t_0) \rightarrow \mathbf{r}_i(t_0 + \Delta t) + \rightarrow \mathbf{r}_i(t_0 + 2\Delta t) \rightarrow ... \rightarrow \mathbf{r}_i(t_0 + n\Delta t). \tag{6.4}$$

There are several algorithms to calculate the next step in the trajectory but the most common are the velocity-Verlet, Verlet integrator or leap-frog algorithm, which all create the same trajectory [105]. As an example, the velocity-Verlet algorithm calculates the next step via

$$\mathbf{r}_i(t_0 + \Delta t) = \mathbf{r}_i(t_0) + \mathbf{v}_i(t_0)\Delta t + \frac{1}{2}\mathbf{a}_i(t_0)\Delta t^2, \tag{6.5}$$

where $\mathbf{v}_i(t_0)$ and $\mathbf{a}_i(t_0)$ are the velocity and acceleration of atom i at time t_0, respectively. The acceleration can be calculated via the FF (approximation of U) and the velocity for the next step can be calculated via

$$\mathbf{v}_i(t_0 + \Delta t) = \mathbf{v}_i(t_0) + \frac{1}{2}\left[\mathbf{a}_i(t_0) + \mathbf{a}_i(t_0 + \Delta t)\right]. \tag{6.6}$$

This algorithm is simple but efficient, stable and reasonably accurate. Importantly, it is time-reversible and symplectic, i.e. the energy in the simulated system is conserved. The time step has to be chosen such that the fastest motion in the simulation is modelled correctly, which is normally around 1 fs for biomolecules to account for the fastest vibrational motions, normally C–H stretches.

The method described creates a trajectory in the micro-canonical or NVE ensemble. Often a MD simulation is executed either in the canonical ensemble NVT or isobaric-isothermal ensemble NPT. The captial letters reference to the quantities which stay constant, with N = number of particles, V = volume, E = energy, T = temperature and P = pressure. Thus in comparison to the NVE ensemble, in the NVT or NPT ensemble the temperature has to be held constant in the system, leading to a modification of the Newtonian equations. The most commonly used method to achieve this is by coupling the simulation to a heat bath. This can be done in different ways, e.g. the Berendsen [119] thermostat, the Andersen thermostat [120], Nosé-Hover thermostat [121–124] or velocity rescaling algorithm [125] which is similar to the Berendsen thermostat but adds a stochastic term. As an example, the modification to the equation of motion (see Eq. 6.3) using the Berendsen thermostat leads to:

$$\mathbf{a}_i = \frac{\mathbf{F}_i}{m} + \frac{1}{2\tau_T}\left(\frac{T_0}{T(t)} - 1\right)\mathbf{v}_i, \tag{6.7}$$

where τ_T is the coupling constant T_0 the desired temperature and $T(t)$ the actual temperature of the system at time t. For $\tau_T \rightarrow \infty$ the NVE ensemble is recovered. For the NVT ensemble the simulation box size is fixed, whereas for the NPT ensemble the size can be varied and the pressure is applied via a barostat. This is done in a similar way as for the temperature coupling. Examples are the Berendsen [119] or Andersen [120] barostat or Parrinello-Rahman algorithm [126].

Another important aspect of MD simulations is setting the right boundary conditions of the simulated system. Rigid walls could be used but they would impose physical constraints

on the surface of the biomolecule which would prevent it from diffusing freely and would constrict it to one location. Thus, in general, periodic boundary conditions (PBC) are used: the initial simulation box is surrounded by an infinite number of replicates in all three dimensions, thereby simulating infinite space. For crystals this is not a problem, but for disordered systems like the protein powders under study here, this can introduce an artificial periodicity. For a large enough initial box size this is normally not a problem as long as no quantity is investigated which depends on long range correlations.

With all these ingredients, a simulation can be conducted in four distinct steps:

1. **Starting structure**
   A starting structure can be a protein from the PDB, which can then be solvated by water molecules in a cubic box or any other geometrical shape
2. **Preparation/Minimisation**
   The energy of the starting structure is minimised to fix atomic positions which are energetically not possible, e.g. too large or small bonds, to obtain a stable initial system.
3. **Equilibration**
   The simulation is started. To obtain reasonable results the system has to be equilibrated first to make sure that their are no major drifts in the total energy and an equilibrium position is reached. This step may take a long time and is highly dependent on the simulated system.
4. **Production**
   After the system is in equilibrium, the simulation trajectory can be saved in small time steps (normally around 1 ps) for later data treatment. This is the part of the simulation which is analyzed.

6.2. Applied methods

6.2.1. Simulated system and its properties

Molecular dynamics simulations of hydrated protein powder were used, representing the interactions of proteins with a small amount of water to be able to compare them directly with experimental data. The MD simulations were performed by Pan Tan from the group of Prof. Liang Hong in Shanghai Jiao Tong University, China. However, all the analysis of the production trajectories of 20 ns after proper equilibration were performed by me. The simulations were started from two different protein structures which can be found in the Protein Data Bank (PDB [60] (www.rcsb.org)): 1) bovine A-L with calcium (A-L_{ca}), PDB ID: 1F6S and 2) bovine A-L without calcium (A-L_{dep}), PDB ID: 1F6R. Both structures were published by Chrysina et al. [127]. Each PDB structure consists of six distinct A-L proteins (=chains), allowing to calculate an average dynamics of a single A-L chain.

The protein molecule was centered in a cubic box of size 8.39 nm at first, with CHARMM 27 force field [111, 113], and TIP4P-ew water model [128], using GROMACS 5.0.7 (GPU version) as MD engine [129, 130]. The boxes were filled with water molecules to start with, which were then deleted (starting from the outside) until the number of water

Table 6.1. - Simulation protocol. Summary about the most important settings in the MD simulations.

PDB	1F6R (A-L_{dep}), 1F6S (A-L_{ca})
T [K]	280,300
hydration [h]	0.05 (dry), 0.4 (hyd)
Force-Field	CHARMM 27
Equilibrium:	300 ps NVT, then 50 ns NPT with 0.5 fs time steps
Production:	500 ns (dry), 500 ns (hyd), NPT, 2 fs time steps
Production recorded:	last 20 ns, every 2 ps
NPT pressure	1 bar, coupling: Parrinello-Rahman algorithm, τ=1 ps [126]
Temperature coupling	velocity-rescale algorithm, τ=1 ps [125]

molecules around the protein met the desired hydration level h [$g_{water}/g_{protein}$]. The box for a hydration level of 0.4h contained 1824 (dep)/1834 (Ca) water molecules and 232 (both) for the dry system (0.05h). All systems were electrically neutralised by adding NaCl. Van der Waals interaction was truncated at 1.2 nm with the Lennard-Jones potential switched to zero gradually at 1.0 nm. Particle Mesh Ewald [131] with Coulomb cutoff of 1.2 nm was used to calculate electrostatic interaction. All bonds involving hydrogen atoms were constrained with LINCS [132] algorithm. The systems were firstly energetically minimised using steepest descent steps with a maximum force of 10.0 $kJmol^{-1}nm^{-1}$ and a maximum of 50000 steps, then equilibrated in the NVT ensemble at T=280 K (and 300 K) for 300 ps and in the NPT ensemble at p=1 bar for 50 ns, with 0.5 fs time step to slowly release the unreasonable atom contact and suppress vacuum. The temperature coupling was performed using the velocity-rescale algorithm with a coupling time of τ=1 ps [125]. The pressure coupling was performed using the Parrinello-Rahman algorithm with a coupling time of τ=1 ps [126]. The production MD simulations for hydration level=0.4h were conducted in the NPT ensemble for 100 ns, with a 2 fs time step, while, those of the dry systems were conducted for 500 ns. Only the last 20 ns trajectories recorded at every 2 ps were used for the analysis. For such a dense system, the global translation and rotation of the protein molecules was strongly suppressed [93, 133].

Since the NPT ensemble (constant particle number, pressure and temperature) was used during the simulation, the density of the system can be changed by varying the simulation box size. This can influence the movements of the proteins and therefore it is important to have a stable density in the system. The evolution of the protein density in the simulation over the last 20 ns used for the analysis is shown Figure 6.1. The density is in all simulations more or less stable. Only for the dry samples with calcium at 280 K and 300 K and the hydrated sample without calcium at 300 K a small increase is visible within the first 5 ns, but the increase in all cases is lower than 2%. The fluctuations of the density around the mean value are around 1%. In addition the dry simulations (0.05h) have a 20% smaller protein density in comparison with the hydrated simulations (0.4h). This can be explained by the fact that the volume of the box is smaller to obtain the same pressure. A visualisation of the simulations and the difference in hydration level is shown in Figure 6.2.

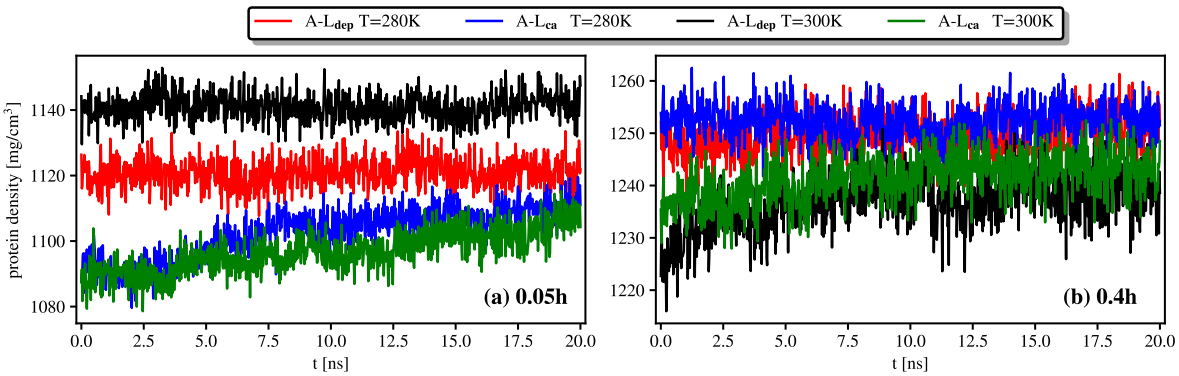


Figure 6.1. - Evolution of protein density in simulation. a) Dry environment (0.05h), b) hydrated environment (0.4h). The density of all simulations stay roughly the same, indicating that an equilibrium state was reached. The density can change since the simulation box size is not fixed due to the NPT ensemble. The density of the protein was sampled every 20 ps.

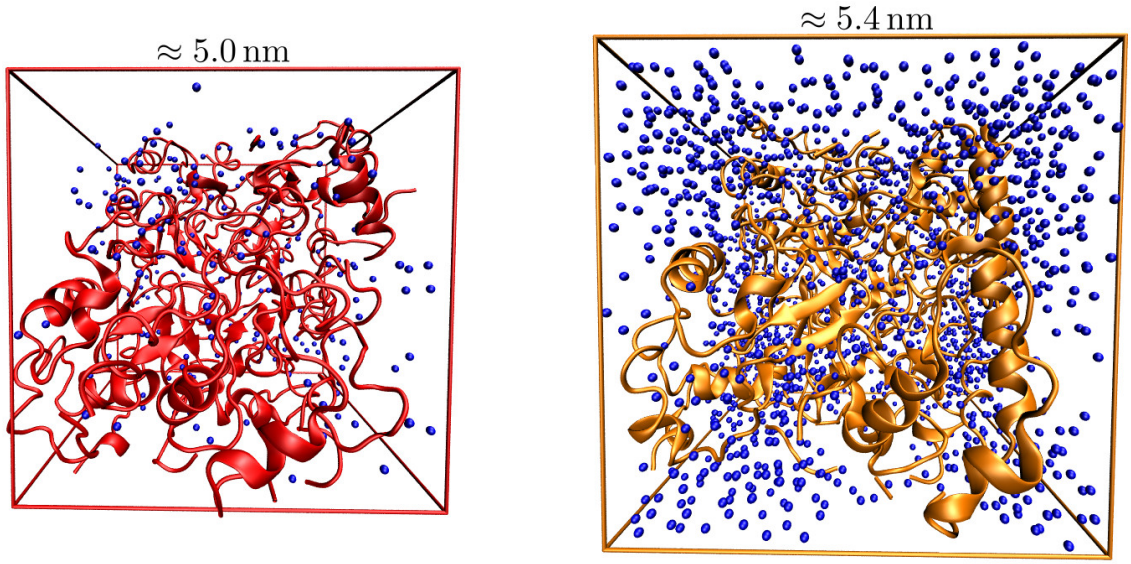


Figure 6.2. - Visualisation of the dry and hydrated simulations of A-L$_{dep}$. Left) dry environment (0.05h) in red. Right) hydrated environment (0.4h) in orange. The lines show the simulation box size and inside the 6 chains of A-L$_{dep}$ at 300 K are visualised. The blue spheres indicate a water molecule.

6.2.2. Direct calculation of the MSD

The atoms of interest in this analysis are the H atoms in the protein since they account for the main contribution to the signal for neutron scattering experiments. The A-L_{ca} (A-L_{dep}) simulations consist in total of 11512 (11457) protein atoms. The number of atoms is different in the two forms, because some chains are missing some amino-acids (residues) at the end of the A-L chain since they were not resolved in the PDB structure. To be consistent in the evaluation of the MSD of the H atoms, only H atoms which are in all chains are considered: every single A-L consists of at least 922 H atoms which are of the same type for all A-L protein chains. Therefore with 6 single A-L chains in each simulation, in total $6 \times 922 = 5532$ H atoms have been evaluated to calculate the MSD and thus to analyze the local movements of the protein. The MSD of a single atom α at location $\mathbf{R}(t)$ at time step t in the simulation is calculated via

$$\mathrm{MSD}_\alpha(t) = \left\langle (\mathbf{R}_\alpha(t_0) - \mathbf{R}_\alpha(t_0 + t))^2 \right\rangle_{t_0}, \tag{6.8}$$

with $\langle\rangle_{t_0}$ is the average over all t_0 defined by the time steps of the simulation. From these individual atoms a mean $\mu(t)$ and standard deviation $\sigma(t)$ of the MSD can be calculated. To estimate the error of the mean and standard deviation of the MSD due to different conformations, the 20 ns simulations were truncated in four equally time spaced parts of 5 ns. The result of the four independent parts was then averaged to obtain a mean MSD $\bar{\mu}(t)$ and its sample standard deviation $s(t)$ taken as an estimation of the error:

$$s(t) = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N=4} (\mu_i(t) - \bar{\mu}(t))^2}. \tag{6.9}$$

The result of each part and its mean $\bar{\mu}(t) \pm s(t)$ can be seen in Figure 6.3. The values of the MSD are shown until $t = 2.5$ ns for which the statistics are still good enough.

In order to compare the MSD $\bar{\mu}(t)$ of the simulations with the MSD $\langle R^2 \rangle$ calculated by the models which describe the EISF, it has to be divided by 2; indeed for the models the MSD $\langle R^2 \rangle$ is defined as a time independent quantity due to the confined motion resulting in (see Eq. (2.39)):

$$2 \left\langle R^2 \right\rangle = \left\langle \mathrm{MSD}(t \to \infty)^2 \right\rangle. \tag{6.10}$$

For convenience, in the following, the MSD obtained with this method will be labeled as direct MSD $\Delta_{\mathrm{dir}}(t)$ with the following definition

$$\Delta_{\mathrm{dir}}(t) = \frac{1}{2} \bar{\mu}(t). \tag{6.11}$$

The $\Delta_{\mathrm{dir}}(t)$ can be directly compared to the MSD extracted from the fits of the EISF. The same procedure is done for the standard deviation of the MSD $\sigma_{\mathrm{dir}}(t)$.

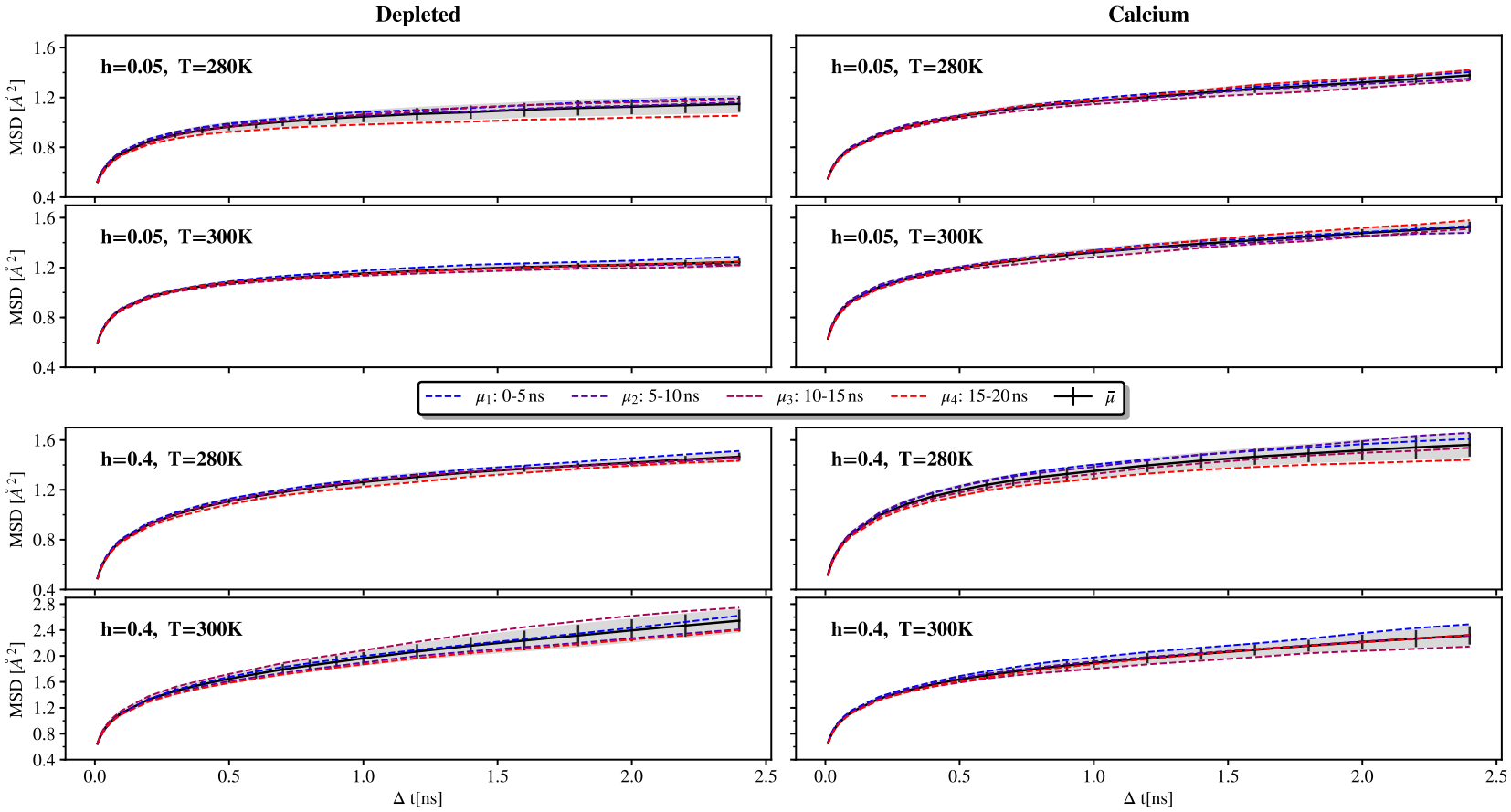


Figure 6.3. - Evolution and average of mean MSD of H atoms in each simulation. Each graph shows the evolution of the mean MSD from 0ns to 2.5ns for each 5ns interval. The color from the first to the last interval changes from blue to red and the black curve shows the average and its sample standard deviation. The left column shows the A-L_{dep} simulations and the right column the A-L_{ca} simulations, whereas the first two rows are at a hydration level of 0.05h (dry protein) and the lower two at 0.4h.

6.2.3. Indirect calculation of the MSD: A theoretical neutron experiment

In order to compare the simulation results with the experimental data, the MSD can also be estimated by the convolution of the instrument resolution $R(\omega)$ with the theoretical dynamic incoherent structure factor (DISF) $S_{\mathrm{inc}}(\mathbf{Q},\omega)$ calculated with the help of the simulation data. The DISF can be calculated via Eq. 2.22 and $R(\omega)$ is given by the instrument. The resulting property is the resolution broadened DISF, $S^{\mathrm{R}}_{\mathrm{inc}}(\mathbf{Q},\omega)$ (see also Eq. (2.65)). This property was calculated with the program MDANSE [134, v.1.1] (www.mdanse.org). The resolution function for each instrument was approximated by a normalised Gaussian function with a Full Width at Half Maximum (FWHM) equivalent to the resolution of the instrument:

$$G(\omega,\sigma_{\mathrm{res}}) = \frac{1}{\sigma_{\mathrm{res}}\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{\omega}{\sigma_{\mathrm{res}}}\right)^2\right\}, \tag{6.12}$$

whereas

$$\mathrm{FWHM} = \sigma_{\mathrm{res}} 2\sqrt{2\ln(2)} = \sigma_{\mathrm{res}}\sqrt{8\ln(2)} \approx 2.35\sigma_{\mathrm{res}}. \tag{6.13}$$

The FWHM of each instrument was obtained by matching the above defined Gaussian function to vanadium scans summed over all momentum transfers Q for IN13 and OSIRIS. For SPHERES the resolution function found in the literature for the large angle detectors was used [44, Fig.2] (Voigt profile with $\sigma_{\mathrm{res}} = 0.244\,\mu\mathrm{eV}, \gamma_{\mathrm{res}} = 0.052\,\mu\mathrm{eV}$). The results are shown in Figure 6.4. The approximated Gaussian functions are in good agreement with the experimental data around the elastic peak. At the edges differences are visible and also the asymmetrical shape of the resolution curve of IN13 and OSIRIS. That should not significantly influence the results for the elastic peak as it is mainly the FWHM of the resolution curve is important. The resolution functions shown are then convoluted with the DISF which is obtained from the simulation. For each DISF calculated with an absolute momentum transfer Q_m, $N_q = 50$ Q-vectors $\mathbf{Q}_i$ with a randomised direction and an absolute length of $|\mathbf{Q}_i| = Q_m + \Delta Q$, with $\Delta Q \leq 0.05\,\mathrm{\AA}^{-1}$ are averaged. In total, the DISF is then calculated in MDANSE as

$$S_{\mathrm{inc}}(Q_m,\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} I_{\mathrm{inc}}(Q_m,t) \times \exp\left(i\omega t\right) \mathrm{d}t \tag{6.14}$$

$$I_{\mathrm{inc}}(Q_m,t) = \frac{1}{N_\alpha}\sum_\alpha \left\langle \frac{1}{N_q}\sum_i \exp\left\{-i\mathbf{Q}_i\cdot\mathbf{R}_\alpha(t_0)\right\}\exp\left\{i\mathbf{Q}_i\cdot\mathbf{R}_\alpha(t_0+t)\right\}\right\rangle_{t_0}, \tag{6.15}$$

where N_α is the amount of H atoms in the simulation and $\mathbf{R}_\alpha$ their location. t and t_0 are defined by the time steps of the trajectory.

From the resolution broadened DISF, $S^{\mathrm{R}}_{\mathrm{inc}}(Q_m,\omega)$, the EISF($Q_m$) is computed by summing up the intensities in the range $\omega = \pm\mathrm{FWHM}/2$. The resulting EISF(Q) is normalised

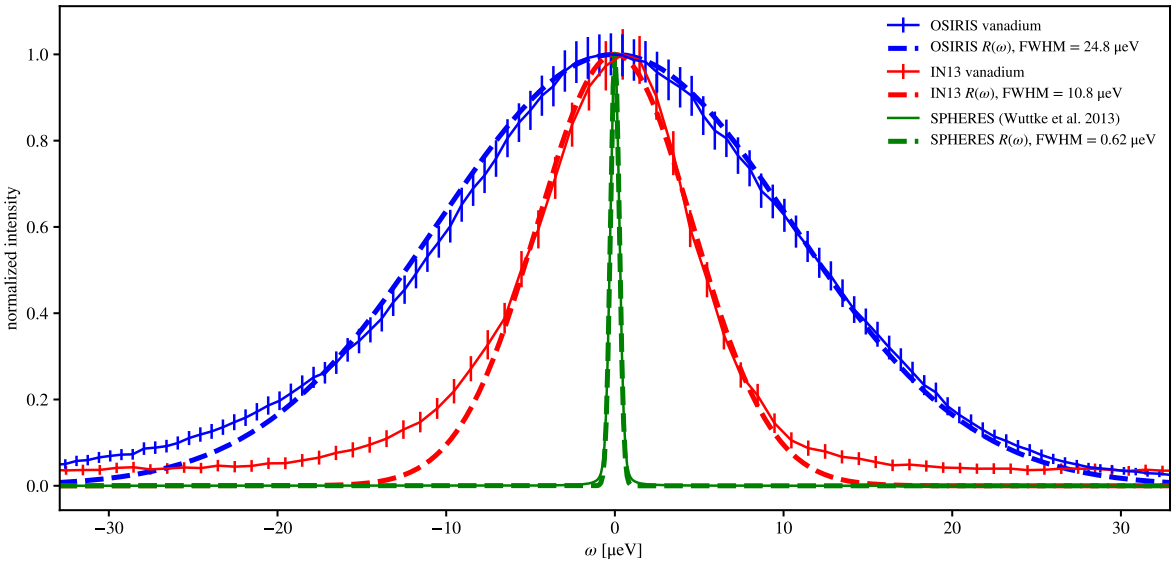


Figure 6.4. - Gaussian resolution function used for simulations. Comparison of Gaussian resolution used for the simulation to experimental resolution of OSIRIS, IN13 and SPHERES. For OSIRIS and IN13 vanadium scans summed over all momentum transfers Q are shown and for SPHERES the resolution function published by Wuttke and Zamponi for large angle detectors is shown [44, Fig.2] (Voigt profile with $\sigma_{\mathrm{res}} = 0.244\,\mu\mathrm{eV}, \gamma_{\mathrm{res}} = 0.052\,\mu\mathrm{eV}$).

by EISF(Q_m=0), so that:

$$\mathrm{EISF}(Q_m) = \sum_{\omega=-\frac{\mathrm{FWHM}}{2}}^{\omega=+\frac{\mathrm{FWHM}}{2}} S^{\mathrm{R}}_{\mathrm{inc}}(Q_m, \omega) \tag{6.16}$$

$$\mathrm{EISF}(Q) = \frac{\mathrm{EISF}(Q_m = Q)}{\mathrm{EISF}(Q_m = 0)}. \tag{6.17}$$

The obtained EISF(Q) can be fitted in the same way as the experimental data to calculate the MSD (see section 2.2 for the models). It is important to mention that for the experimental data the lowest temperature scan was used for the normalisation, whereas here the value obtained at $Q_m = 0$ was taken due to the lack of a simulation at very low temperature.

As models for the EISF(Q), the GA, Yi and PK model have been used with the following Q ranges

- GA: $0 - 1\,\mathrm{Å}^{-1}$
- Yi: $0 - 3\,\mathrm{Å}^{-1}$
- PK: $0 - 4\,\mathrm{Å}^{-1}$.

Figure 6.5 shows the resulting fits for A-L_{dep} and A-L_{ca} at 280 K and 300 K for the dry and hydrated protein with the Gaussian resolution function equivalent to SPHERES. The simulation of the dry sample (0.05h) shows only minor differences in the EISF between 280 K and 300 K, whereas for the hydrated protein (0.4h) the EISF at 300 K is clearly

smaller at $Q \neq 0$.

The differences resulting from the three resolutions of SPHERES, IN13 and OSIRIS are shown in Figure 6.6 at a temperature of 300 K. For the dry protein the resolution has only a small impact on the behaviour of the EISF. Also, in the case of the hydrated protein, the EISFs from IN13 and OSIRIS are not much different but the EISF obtained from the narrow FWHM of the SPHERES resolution function, is well separated. Overall, it can be seen that the EISF decreases with higher resolution (smaller FWHM). This is the anticipated behaviour since the resolution is inversely proportional to the observed time frame. Larger time frames result in the observation of larger localised movements and therefore a decrease in elastic intensity.

The models fit the EISF reasonably well within the fitted Q range indicated by the dashed gray lines in Figures 6.5 and 6.6. After their respective ranges the GA and Yi model under evaluate the EISF (dashed lines) whereas the PK model over evaluates the EISF. None of the models can fit all Q values to 7 Å^{-1} properly.

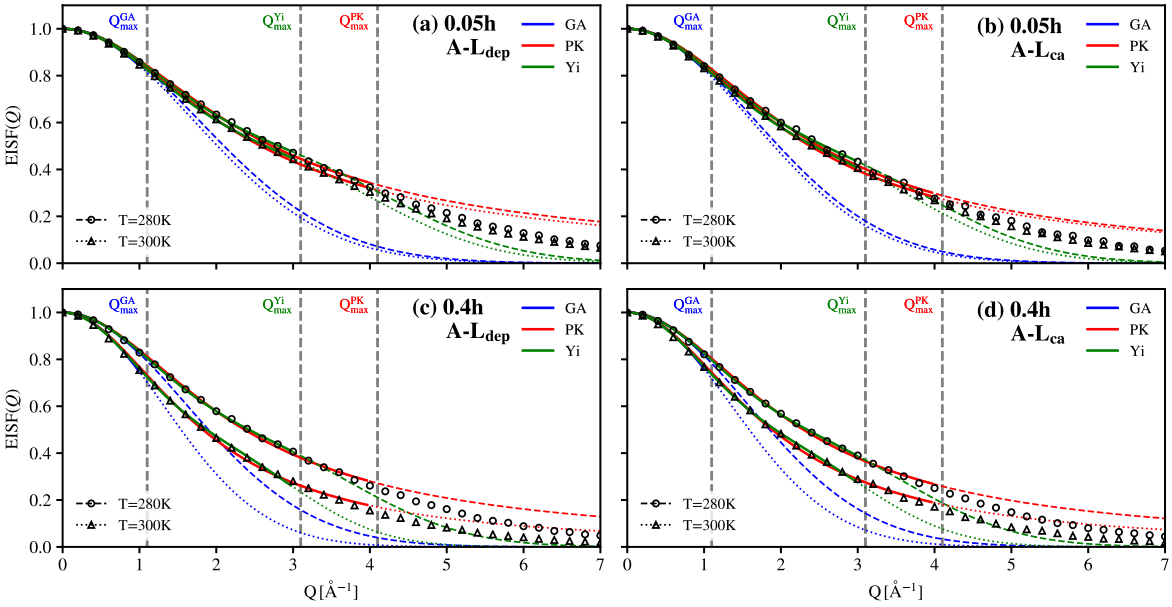


Figure 6.5. - Fits of EISF obtained with simulations for the SPHERES resolution function. Comparison between the three fit models GA, PK and Yi for the EISF obtained by the simulations of A-L$_{dep}$ (left) and A-L$_{ca}$ (right) with the resolution function of SPHERES. Figures a) and b) show the simulations with hydration level of 0.05h (dry) and figures c) and d) with hydration level of 0.4h.

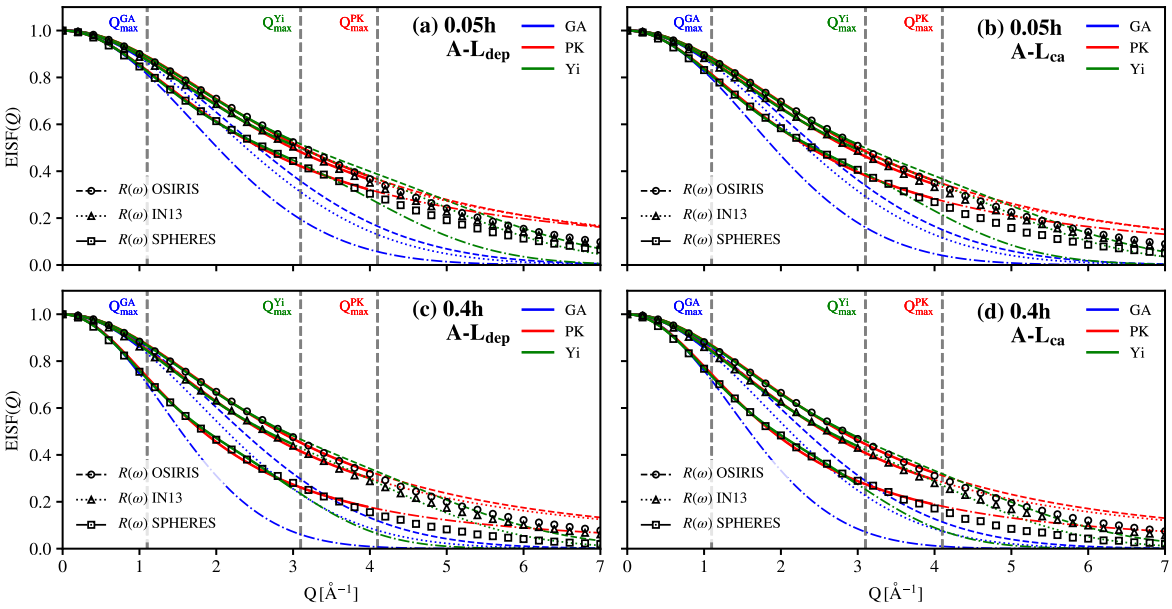


Figure 6.6. - Fits of EISF obtained with simulations with different resolution functions at 300 K. The figures show the EISF at 300K for A-L$_{dep}$ (left) and A-L$_{ca}$ (right) with the resolution functions of SPHERES, IN13 and OSIRIS. Figures a) and b) show the simulations with hydration level of 0.05h (dry) and figures c) and d) with hydration level of 0.4h. The coloured lines indicate the fit results of the different models.

6.3. Results

Here the results for the MSDs obtained directly (see section 6.2.2) and indirectly (see section 6.2.2) from the simulations, are compared. To relate the time dependent direct MSD $\Delta_{\mathrm{dir}}(t)$ to the results of the time independent indirect MSD $\langle R^2 \rangle_{\mathrm{ind}}$ Heisenberg's uncertainty principle is used (see Eq. (3.13)):

$$\Delta_{\mathrm{dir}}(\tau_{\mathrm{FWHM}}) \approx \langle R^2 \rangle_{\mathrm{ind}}\,, \text{ with} \tag{6.18}$$

$$\tau_{\mathrm{FWHM}} = \frac{\hbar}{\mathrm{FWHM}}. \tag{6.19}$$

The corresponding times for each instrument are summarised in Table 6.2. In Figure 6.7 the results of the MSD of the three different fitting models GA, Yi and PK (coloured) and the direct calculated MSD (black) are compared for all four simulated samples at the three instrumental resolutions. The dry powder is shown on the left hand side and the hydrated powder on the right. They are ordered from 280K to 300K and A-L_{dep} next to A-L_{ca}. For the dry protein and for all resolutions, the directly calculated MSD is clearly larger than the MSD calculated from the models, but the quantitative behaviour between the simulations is the same. Assuming that the direct calculation represents a result as close as possible to the true MSD, the difference between the black curve and the others could indicate the order of magnitude of the error introduced by using models. As anticipated, the MSD increases with increasing temperature but the effect seems to be larger on smaller timescales, i.e. lower instrumental resolutions (IN13 and OSIRIS). The A-L_{ca} simulations also have a slightly higher MSD for all instrumental resolutions. When comparing the different models, the GA evaluates to a higher MSD than the Yi and PK model for IN13 and OSIRIS. For SPHERES this behaviour is inverted.

For the hydrated protein the effects are similar to the dry protein but there are three main differences. First, the MSD of the models is much closer to the direct MSD but still smaller. Secondly, the difference between the MSDs at 280 K and 300 K is much larger. Thirdly, for SPHERES the MSD for A-L_{ca} at 300 K is slightly lower than for A-L_{dep} which is the case for all four methods.

An analogous plot is shown in Figure 6.8 for the standard deviation of the MSD. The evaluated standard deviation of the MSD is obtained by the direct method $\sigma_{\mathrm{dir}}(\tau_{\mathrm{FWHM}})$ and with the PK and Yi model σ_{ind}. Similar tendencies as for the MSD are visible, however, the modelled σs do not always follow the same behaviour for all samples as it was the case for the MSD. For example, for the dry protein on SPHERES, the σ_{dir} is smaller for

Table 6.2. - Instrument resolution FWHM vs time. Relation of the FWHM of Gaussian instrument resolution in energy space to the time τ in time space (see Eq. (6.19)).

Instrument	FWHM [μeV]	τ_{FWHM} [ps]
OSIRIS	24.8	26
IN13	10.8	60
SPHERES	0.62	1060

A-L_{dep} at 300 K than for A-L_{ca} at 280 K. In contrast, the PK model suggests a higher value. Another important feature is that the PK model has always a larger σ than the Yi model.

The distribution of the MSD in the protein can also be determined from the simulations. This is calculated following the method used by [9] which enables an evaluation of the main contributions to the heterogeneity and of how many populations with different motions exist in the sample. Figure 6.9 shows the distribution at $t = 30$ ps ($\approx$ OSIRIS), $t = 60$ ps ($\approx$ IN13) and $t = 1$ ns ($\approx$ SPHERES) for all simulations. The line for each time was obtained by binning the MSD values in steps of 0.025 Å^2 together and normalised by the total number of H atoms. The individual MSD values were obtained by the average value of the four independent slices of 5 ns for each simulation in the same way as for the direct MSD evaluation. In all simulations and for all three times t, one large peak at around 0.25 Å^2 is visible. Only for the distribution at 1 ns (green) a small second peak around 2.75 Å^2 is visible. For the hydrated samples the first peak is shifting slightly to higher MSD values and its peak is significantly smaller than for 30 and 60ps. This effect is emphasised at 300 K. No significant variation is observed between the A-L_{dep} and A-L_{ca} samples.

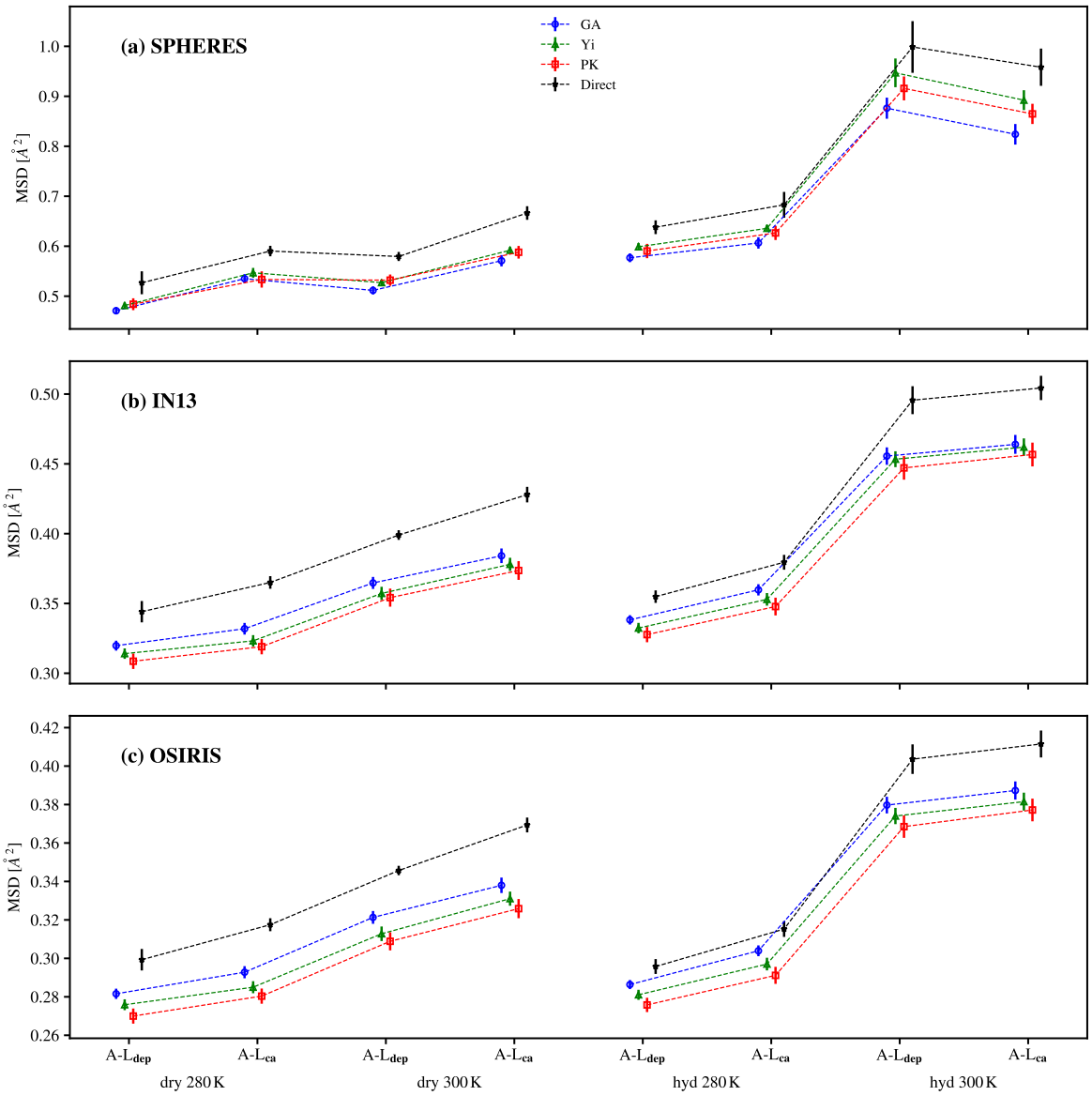


Figure 6.7. - MSD in MD simulations: Indirect vs. Direct calculations. The figure summarises the results of the MSD obtained by the direct and indirect calculations of the MSD ($\Delta_{\mathrm{dir}}(\tau_{\mathrm{FWHM}})$, $\langle R^2 \rangle_{\mathrm{ind}}$). $\langle R^2 \rangle_{\mathrm{ind}}$ is calculated via the GA, Yi and PK model. Figure a) shows the results for the Gaussian resolution function of SPHERES, b) for IN13 and c) for OSIRIS. On the left side the A-L with at dry hydration level of 0.05h is shown and on the right side the hydrated protein for 0.4h. As indicated, each side is ordered in the same way by increasing temperature (280 K and 300 K) and A-L_{dep} next to A-L_{ca}.

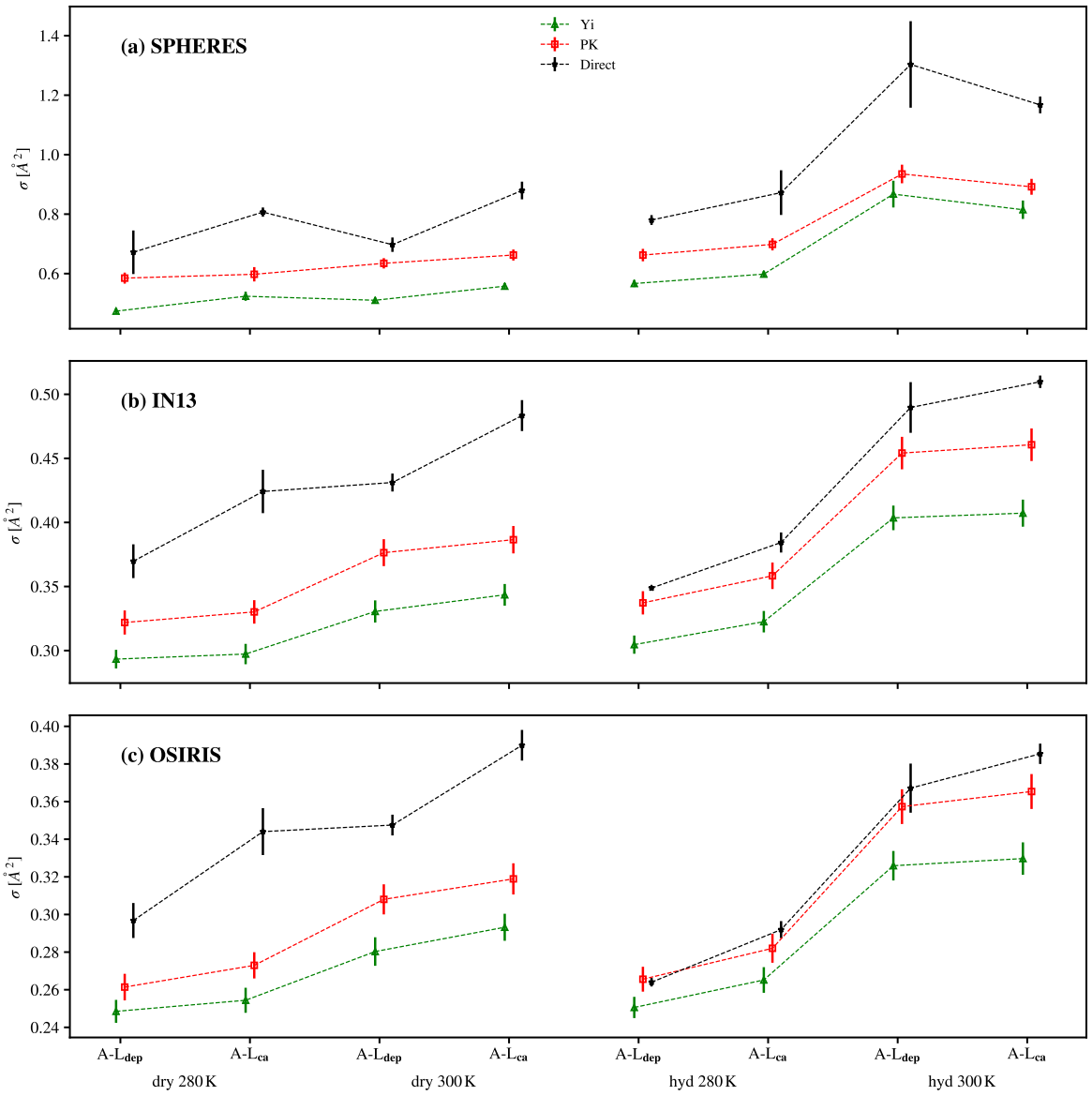


Figure 6.8. - Standard deviation σ of MSD in MD simulations: Indirect vs. Direct calculations. The figure summarises the results of the standard deviation of the MSD obtained by the direct and indirect calculations of the MSD ($\sigma_{\mathrm{dir}}(\tau_{\mathrm{FWHM}})$, σ_{ind}). σ_{ind} is calculated via the Yi and PK model. Figure a) shows the results for the Gaussian resolution function of SPHERES, b) for IN13 and c) for OSIRIS. On the left side the A-L with at dry hydration level of 0.05h is shown and on the right side the hydrated protein for 0.4h. As indicated, each side is ordered in the same way by increasing temperature (280 K and 300 K) and A-L_{dep} next to A-L_{ca}.

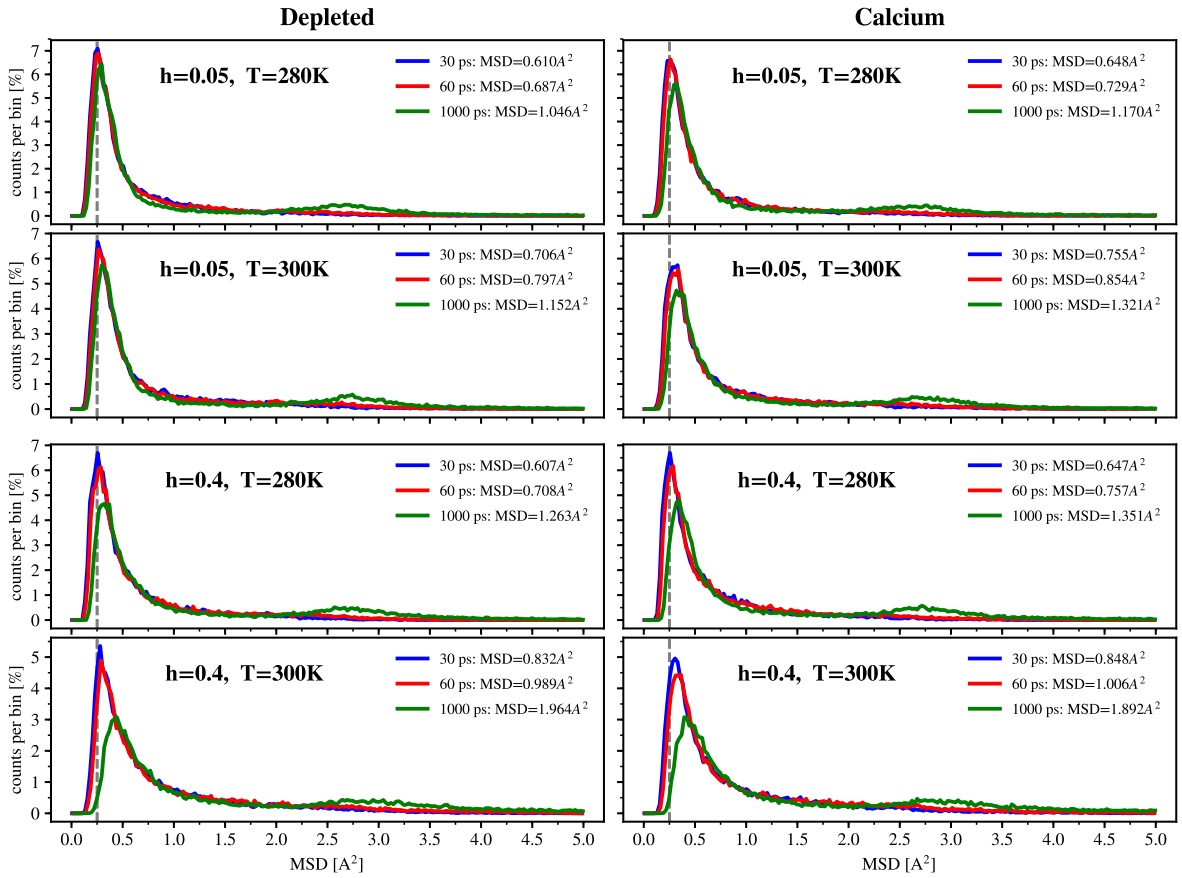


Figure 6.9. - Distribution of MSD. Comparison of distributions of the MSD at 30ps (≈ OSIRIS), 60ps (≈ IN13) and 1000ps (≈ SPHERES) for A-L_{ca} (right) and A-L_{dep} (left). The upper two figures are the dry samples at 280 and 300 K, the lower two figures are the hydrated samples at 280 and 300 K. The lines were obtained by binning MSD values in steps of 0.025 Å^2 together and normalised by the total number of H atoms. The raw MSD values were obtained by the average value of the four independent slices of 5 ns for each simulation. The MSD in the legend shows the mean MSD which is the same as $\bar{\mu}$ in section 6.2.2.

6.4. Comparison to experimental data

In order to compare the MD simulations with the experimental data the results of the fitting models from the previous section are plotted together with the results of the experiments (see chapter 5).

The experimental data has a limited amount of temperature values (every $5 - 10\,\mathrm{K}$) which are not always in coincidence with the simulation temperatures. In addition, the experimentally evaluated MSD sometimes shows statistical outlayers at some temperature values. To reduce the effects on the results, the experimental MSD values were averaged over three temperature values (smoothing average). To achieve the MSD at the same temperature as the simulations, the MSD between two smoothed temperature data points have been linearly interpolated. The same was done for the standard deviation of the MSD.

As can be seen in Figure 6.10, the MSD values of the simulations and experiments are not always in quantitative agreement but the trends are similar. For the dry protein the experimental values are smaller in the case of SPHERES and OSIRIS and in good agreement for IN13. For the hydrated protein, the experimental values are, in general, larger than the simulated values. For SPHERES they agree quite well at 300 K. In all cases the directly calculated MSDs are larger than those extracted through the models, so the absolute values of the MSD shown in the figure cannot be absolutely trusted. However, the experimental MSD of the depleted hydrated sample seem systematically higher than those of the A-L_{ca} sample, which is hardly visible within the statistics in the simulation results at 280 K and below 1 ns. It might indicate slightly enhanced dynamics for the depleted sample in such conditions, which could be expected as calcium has a stabilising effect [51]. The higher mobility becomes visible only in the simulations at higher temperatures and longer time scales, as the variations in the sample are certainly small. An interesting point is the difference between the models. For the simulations the Yi and PK model mainly evaluate to smaller MSD values whereas for the experiments they were larger. But it also has to be taken into account that here the fitted Q-range plays also an important role and influences this behaviour. Also the differences between the models is larger for the experimental data reflecting the worse statistics.

In Figure 6.11 the standard deviation of the MSD is shown for the Yi and PK models. Here the experimental values fluctuate much more than in the simulations. For OSIRIS the experimental points have a high statistical error due to the underlying data which is not good enough to fit a second independent parameter with good precision. The values for OSIRIS are shown for completeness but cannot be used for a valuable comparison.

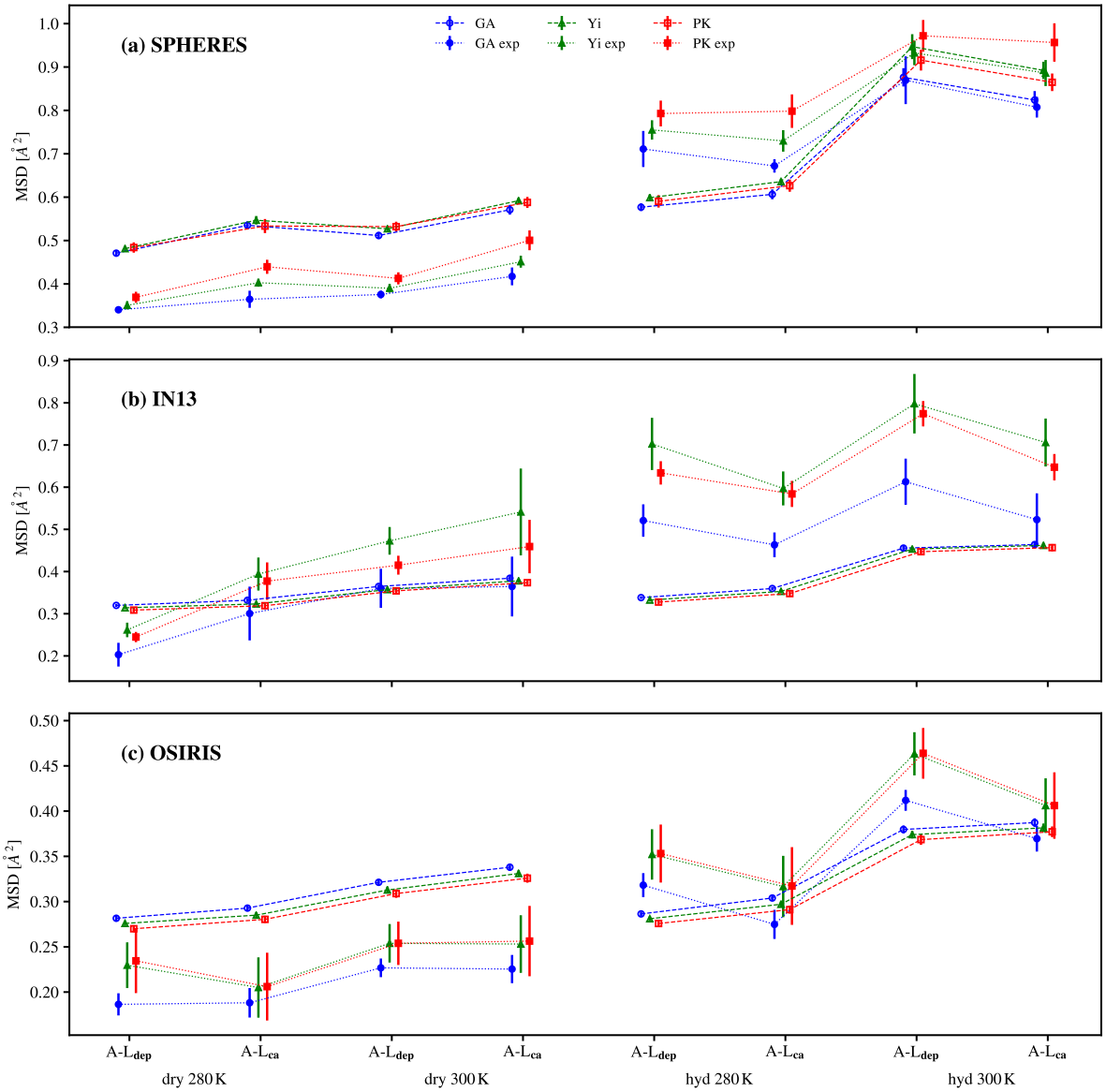


Figure 6.10. - MSD: Experimental data vs. MD simulations. Comparison between the three models obtained by fitting the experimental data (open symbols, dashed lines) and the MD simulations (filled symbols, dotted lines).

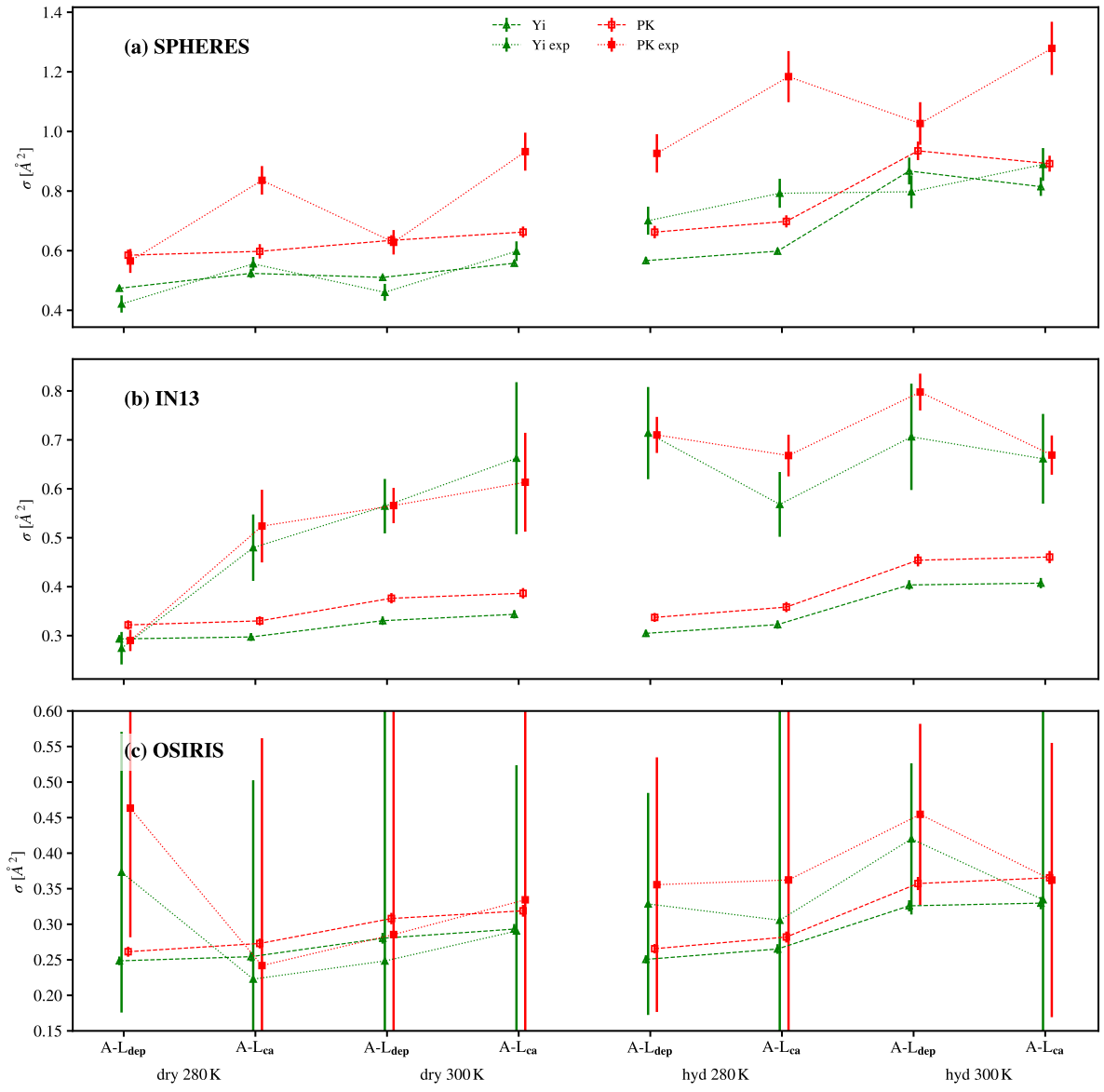


Figure 6.11. - Standard deviation σ of MSD: Experimental data vs. MD simulations. Comparison between the σ of the Yi and PK models obtained by fitting the experimental data (open symbols, dashed lines) and the MD simulations (filled symbols, dotted lines).

6.5. Conclusions and Discussion

This chapter compared simulated to measured data. Molecular dynamics simulations are indeed a very powerful tool to understand in more detail the dynamics of individual atoms in a sample measured in a neutron scattering experiment as both techniques give access to comparable temporal and spatial scales. Usually, simulations are run in solution, but since EINS experiments are frequently done with hydrated powders scientists, albeit few, have developed approaches to simulate hydrated powders [9, 135] by adapting the setup accordingly. The work in this thesis has benefited from the collaboration with simulation expertise from the group of Prof. Liang Hong (Shanghai Jiao Tong University, China) which provided the simulated trajectories in such an environment.

A direct comparison of neutron data and simulated signals is not always trivial as the absolute values depend significantly on data corrections and normalisation, the accuracy of force fields and starting structures. It is also common to find that simulations cannot reproduce results extracted from neutron scattering data quantitatively (see [136] or Figure 6.10). Hence, the decision to compare MSDs by extracting them in a very similar way from both experiment and simulation.

The order of magnitude of the experimental MSD are well reproduced by the simulations with MSD of hydrated samples being above those of the dry samples. The results indicate that the models describing the simulated EISF (obtained from the DISF) underestimate the simulated directly calculated MSD (see Figure 6.7). One might therefore speculate that no model is able to take the whole dynamics into account and that effects due to the limited space and time windows are not negligible. For the hydrated protein the differences are not as large as for the dry protein. In addition, the difference between the models is not negligible, but the trends are always the same and in agreement with the direct MSD, i.e. all curves obtained through the different models are simply parallel. Interestingly, the hierarchy between the models is always the same for an identical instrument resolution (with the exception of SPHERES in a dry environment). One can therefore conclude that the GA gives equally good results as the other models, since the absolute values of the MSD are unknown.

In comparison to the experimental data, the simulation cannot provide reliable quantitative results (see Figure 6.10). In general, the experimental MSDs of the Yi and PK models are higher than those from the GA model, whereas for the simulations this trend is inverted in most cases. Here, it has to be stressed again, that this behaviour is highly dependent on the chosen Q-range and thus no definitive trend can be concluded. One main difference between the Yi and PK model is the additional information obtained from the standard deviation of the MSD. Figure 6.8 indicates that the models give some meaningful estimation of the standard deviation but similar to the MSD, the values are quantitatively not the same in most cases. However, Figure 6.10 indicates that all simulated curves lie very closely together and it is not possible to disentangle the quality of the models. In contrast, the experimental curves show larger differences and in particular the GA model gives MSD which are more strikingly different from the MSD obtained through the Yi and PK models. Nevertheless, none of these results favour any one model over another, as the statistics are probably not good enough to discriminate small effects, eventually due to

the different Q-ranges used.

As shown by Figure 6.9 the distribution of the MSD can be mainly described by two different peaks which are independent of hydration. The second peak is most visible above 1 ns, whereas below 60 ps it is not well distinguished. It is mainly the H-atoms of the methyl groups (not shown here) that are contributing to this peak, which is in accordance to the findings of Yi et al. [9]. Like introduced in section 5.1, methyl group rotations contribute to the elastic neutron spectra and the findings here support that they are a major contributor to heterogeneity originating from these motions, which becomes more visible at longer time scales. Yi et al. [9] simulated the camphor-bound cytochrome P450 at h=0.4 in a way comparable to the simulations here. They also showed that this peak is more dominant at higher temperatures. Furthermore, the second peak at larger amplitudes is also more pronounced at 1 ns. At 100 ps it is closer to the first peak and much broader. Tokuhisa et al. [12] simulated Staphylococcal nuclease (SNase) in a water box at 300 K and also found two distinguishable peaks. The time was not documented but the evaluated simulation time was 1 ns, indicating that the investigated time window was likely smaller than 100 ps.

Overall this leads to the conclusion that the three models give reasonable results in comparison to the direct MSD from the MD simulations. For a precise data set, the differences between the models are not significant concerning the trends, but the quantitative values are, depending on the evaluated Q-range. The Yi and PK model give further insight into the standard deviation of the MSD, but with respect to the MSD they do not give more accurate results. Furthermore, it is also important to state again that in contrast to the experimental data, the simulated EISF was not normalised to the lowest temperature data due to the lack of such simulation data, which could also partly explain the quantitative differences. Doing that, one would more consistently treat experimental and simulated data and eliminate more uncertainties which might arise.

7. Conclusion and outlook

7.1. English version

For decades, neutron scatterers have used only few models to analyse scattering data. Concerning elastic incoherent neutron scattering, which can be applied to biological samples in solution or in hydrated powder forms, the most commonly used model is the Gaussian approximation (GA), introduced by Rahman et al. in 1962 [7]. However, the limits of its applicability have never been systematically studied or clearly defined so that one can still find many approaches and justifications for the data treatment in the literature. Although a few authors have proposed several theoretical models [9, 19, 23, 26, 27] to better describe the data, they only have been tested on a very small and limited amount of experimental data. Therefore, a more complete and systematic study to asses what is the best way to analyse EINS data was lacking, in particular a complete data set taken on different spectrometers but on the same samples. This thesis has addressed this gap. One has to remember that underlying the scattering data is a unique sample with all its characteristics whose comprehensive study requires the combination of many experiments. The results of such experiments have to complement each other and if contradictions exist, they have to be investigated systematically. This will only be achieved by combining data taken on several instruments in order to obtain as complete a view as possible of the system.

In 2013, Peters and Kneller [10] developed a model that takes into account heterogeneity of harmonic motions, in order to describe data beyond the Gaussian approximation and thus beyond the small Q-range for which it is theoretically valid. They found that the new model described the data of human acetylcholinesterase (hAChE) very well over the whole available Q-range (up to $4.5\,\text{Å}^{-1}$ on IN13), but that at longer time scales (IN16, $\approx 1\,\text{ns}$) the MSD showed a maximum around 250 K which decreased at higher temperatures (see [10] and Figure 7.1a, red curve). Such behaviour had never been observed before and its origin was not clear. This gave room for speculation about effects arising at higher Q-values which were neglected within the Gaussian approximation. For instance, confinement effects which occur at low Q values are no longer visible at high Q values. With this in mind, this work wanted to address such effects in more detail including the influence of the time scale by measuring the same samples on different spectrometers.

At first, the investigations were meant to illustrate the importance of measuring data at the high Q values and the possibility of using this data to obtain a more accurate description of the MSD. As discussed in section 5.5 the momentum transfer range at high Q may indeed hold some additional information, but the work in this thesis has shown that, in fact, the data at very low Q values may be even more important. Thus, when looking for qualitative differences between sample conditions or types, the GA model seems to be sufficient in most cases, but the models describing a larger Q range can provide

further insights. Each case has to be carefully evaluated as the addition of one extra fitting parameter may lead to better results or become detrimental due to poor statistics, especially for instruments covering a narrow Q range. The data in this thesis has shown that, in most cases, a bi-modal fit seems to improve the results, which is in agreement with recent publications. Vural et al. [28] suggest a new heterogeneity model for EINS data including two distinct Gaussian distributions. Doster [25] introduces a bi-modal distribution function where one contribution is due to local translations of the sample and one due to rotations from methyl groups. These two publications show the topicality and timeliness of the work presented and the need for more complete data sets to confirm or reject such models.

The work in this thesis has also shown that the measured elastic intensity at zero momentum transfer EI(0) is one of the key parameters to compare samples in a more accurate way and to obtain more quantitative MSD. Before this work, this issue had

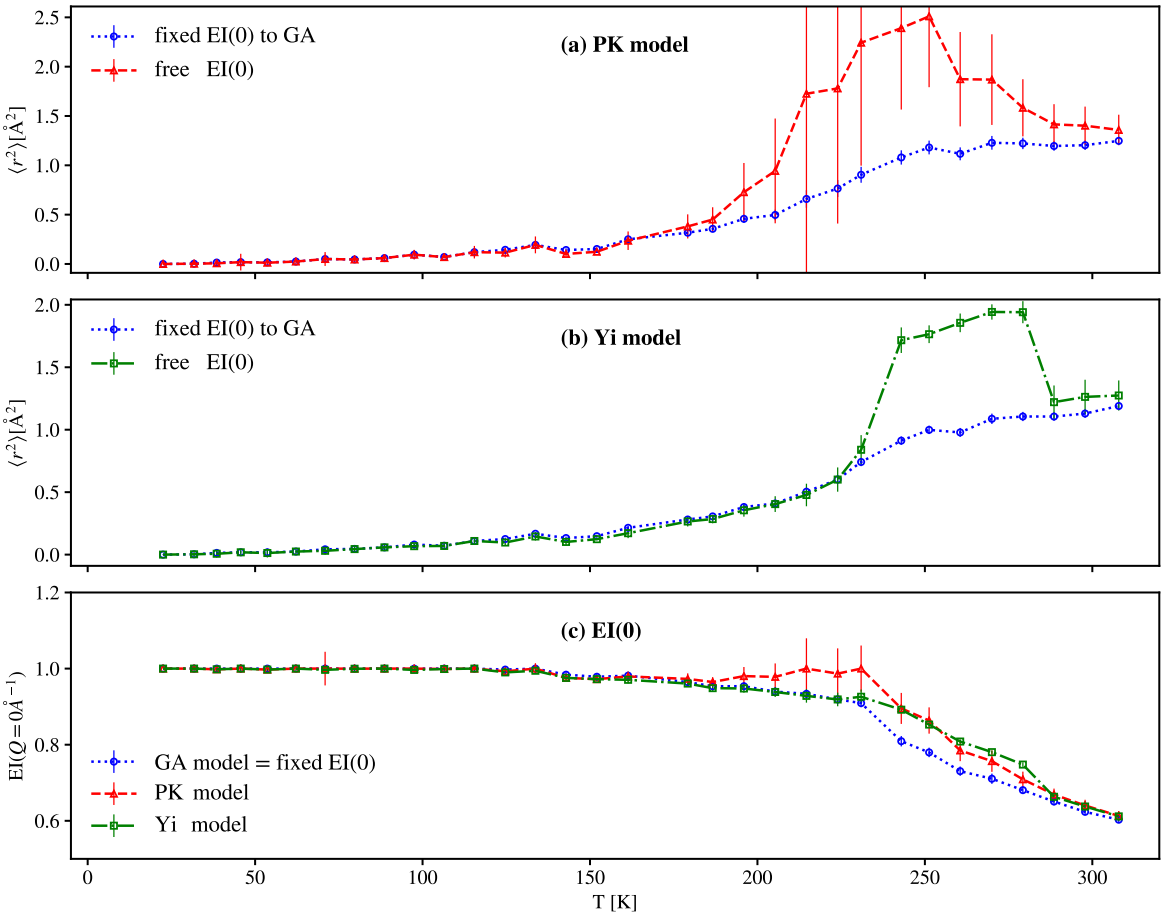


Figure 7.1. - The influence of EI(Q=0) for hAChE, IN16. The image depicts the difference of the MSD for using the PK (a) and Yi (b) models with fixed or unfixed value of EI(Q=0). In the unfixed case the PK (red curve) and Yi (green curve) models have a peak in the MPSF/MSD around 250 K. When they are fixed to the EI(Q=0) (c) evaluated with the GA model (blue curve), they do not show this peak any more and the MSD is increasing linearly.

not been addressed in detail, specifically the strong influence on the resulting values of dynamical parameters such as the much quoted MSD of proteins. Typically, the value is not mentioned or data normalised to one at $Q = 0$, without stating the method used for interpolation. As has been shown here, the value of EI(0) is often not close to the theoretical value of one and changes with sample composition, temperature and instrument. The three most likely reasons for this are multiple scattering effects, coherent scattering contributions and neglecting the QENS contributions to the EINS (see section 5.5). These have not been investigated in detail in this work, but the need to understand and quantify them better was clearly demonstrated. To be consistent in the analysis, the value of EI(0) was fixed to the value obtained from the GA model which is strictly valid at low Q values. This can have dramatic effects on the MSD which is depicted in Figure 7.1. Respecting this rule, the data used by [10] was fit successfully displaying a linearly increasing MSD, even at high temperatures.

Analysing data from MD simulations of A-L protein powder, it was demonstrated that the extracted EISF fitted with three different models gave similar results for the MSD (indirect method) as the direct calculation from the trajectories (direct method). Compared to the MSD for the time windows of the specific instruments, the indirect method always evaluated to smaller values indicating a consistent undervaluation of the mean square displacement. The qualitative results of the MSD (direct vs. indirect method) between the different simulations were the same and, again, it was shown that the GA model gives an accurate estimation of the MSD, even if only a small Q range was used. In contrast, the width of the MSD distribution was not as clearly evaluated with the Yi and PK model. In the comparison with the experimental results the quantitative values of the MSD were in the same order of magnitude but not in good agreement for most cases. For the hydrated samples the simulated MSD were larger, for the dry sample, they were smaller. In addition, qualitatively they were not always consistent. This indicates that there are not large, but still significant differences between the MD simulations of powders and the experimental results. This could be due to inaccuracies with the MD simulations or experimental issues, or both. The way experimental data are corrected and normalised is certainly crucial for a better quantitative agreement and the force fields of the simulations need also more and more input from experiments to be able to describe the samples in the most realistic way. A definite statement cannot be made since only two temperatures have been evaluated and for a more precise comparison a complete simulated data set including an entire temperature scan would be needed.

QENS allows us to measure molecular dynamics much more distinctly, but such measurements are more time-consuming. For instance, complete temperature scans cannot be performed in a reasonable amount of time. Another aspect to consider is that, in order to improve modelling of neutron scattering data, progress in the domain of QENS analysis is needed and some publications are now going in this direction. Recently, even the basics of incoherent neutron scattering analysis, the Van Hove correlation functions [137] in the classical limit, have been questioned by Frauenfelder et al. [138] who challenge the neglection of quantum effects in the classical limit that has been used for over 50 years and is the foundation of most modelling efforts. Also Kneller continues to work on the improvement of the used classical limit of the Van Hove theory and states that his new

theory "provides [...] a physical interpretation for the complex quantum version of the Van Hove correlation functions and shows in particular that there exists a physical meaningful and intuitively understandable relation between the quantum and the classical version of the Van Hove functions" [139]. His newly proposed approach contains an interpretation based on scattering-induced transitions between quantum states of the target. Such suggestions are exciting and should be tested against experimental data to promote a better understanding of scattering events. Furthermore, a better description of QENS data will also improve the understanding of EINS and the ability to separate the two contributions to incoherent scattering.

Another important aspect is the development of new software to analyse the data. Especially new users of neutron facilities have to be able to analyse data and understand the advantages and disadvantages of different methods. It is important to provide different models and consistent data evaluation software. Thus, new models have to implemented in the facilities' software such as Lamp [68] or Mantid [69], which makes support easier and also data evaluation more transparent. Moreover, openly accessible software allows users to have insights into the data treatment and contribute to eliminate eventual errors in the coding. Such software packages should ideally take into account all possible effects on the evaluations like heterogeneity, anharmonicity, coherent scattering, multiple scattering, QENS, etc. as described in this thesis. In addition, they should provide good documentation to be able to understand the underlying theory and coding.

Lastly, new instrumentation or facilities with higher neutron fluxes, like the European Spallation Source ESS, can help to develop new techniques enabling the validation of neutron scattering modelling. In the long run, this will help us to close the gap between simulations and experiments, resulting in a deeper understanding of dynamics in biomolecules.

7.2. Version française

C'est un fait établi que depuis des décennies, les utilisateurs de neutrons n'utilisent que peu de modèles pour analyser les données de diffusion. En ce qui concerne la diffusion neutronique élastique incohérente, applicable aux échantillons biologiques en solution ou sous forme de poudre hydratée, le modèle le plus courant et le plus simple est l'approximation gaussienne, introduite par Rahman et al. [7]. Cependant, les limites de son applicabilité n'ont jamais été systématiquement étudiées ni clairement définies, de sorte que l'on peut encore trouver de nombreuses approches et justifications dans la littérature. Bien que certains auteurs ont proposé plusieurs modèles théoriques [9, 19, 23, 26, 27] pour mieux décrire les données, ils ont seulement été testés avec un petit nombre limité de données expérimentales. Par conséquent, il manquait une étude plus complète sur la meilleure façon d'analyser les données EINS, en particulier avec un ensemble complet de données prélevées sur différents spectromètres, mais sur les mêmes échantillons. C'est ce qui vient d'être présenté dans ce travail de thèse. Il ne faut pas oublier que derrière les données de diffusion, il existe un échantillon unique avec toutes ses caractéristiques, et pour les déterminer de manière complète, il est nécessaire de combiner de nombreuses expériences. Cependant, les résultats doivent se compléter entre eux, et si des contradictions apparaissent, elles doivent être systématiquement examinées. Cet objectif est uniquement atteint en combinant des données provenant de plusieurs instruments afin d'obtenir une vision aussi complète que possible du système biologique.

En 2013, Peters et Kneller ont développé un modèle prenant en compte l'hétérogénéité des mouvements harmoniques, afin de pouvoir décrire des données au-delà de l'approximation gaussienne et donc au-delà de la petite gamme de valeurs Q pour laquelle elle est valable. Ils ont constaté que le nouveau modèle décrivait très bien les données de la protéine *human acetylcholinesterase* (hAChE) sur l'ensemble de la gamme Q disponible (jusqu'à $4{,}5\,\text{Å}^{-1}$), mais qu'à des échelles de temps plus longues (IN16, $\approx 1\,\text{ns}$), les MSD montraient un maximum à 250 K pour diminuer pour des températures plus élevées (voir [10] et Figure 7.1). Un tel comportement n'avait jamais été observé auparavant et son origine n'était pas claire. Cela laissait place à des spéculations sur des effets apparaissant à des valeurs de Q élevées et négligés dans l'approximation gaussienne, comme par exemple des effets de confinement à des valeurs de Q faibles qui ne seraient plus visibles aux valeurs de Q élevées. Dans le cadre de cette recherche, nous avons voulu étudier ces possibles effets plus en détail, y compris l'influence de l'échelle de temps, i.e. en mesurant les mêmes échantillons sur différents spectromètres.

L'étude initiale avait pour objectif de montrer l'importance des valeurs de Q élevées et de déterminer si elles pouvaient être utilisées pour décrire plus précisément les MSD. Comme indiqué dans la section 5.5, ces valeurs pourraient contenir des informations supplémentaires, mais les résultats de cette thèse ont montré qu'en réalité, la gamme en Q des faibles moments transférés s'avèrerait même plus importante. Ainsi, pour des tendances qualitatives, le modèle de l'AG semble être correct dans la plupart des cas, mais les modèles décrivant une gamme en Q plus large peuvent donner quelques indications supplémentaires. Chaque cas doit être soigneusement évalué, comme l'ajout d'un paramètre d'ajustement supplémentaire peut conduire à de meilleurs résultats, voire à de moins bons résultats en

raison de mauvaises statistiques, en particulier pour les instruments avec une gamme en Q étroite. Nous avons montré que dans la plupart des cas, un ajustement bimodal semble améliorer les résultats, ce qui est en accord avec des publications récentes. Vural et al. [28] suggèrent un nouveau modèle d'hétérogénéité pour les données EINS comprenant deux distributions gaussiennes distinctes. Doster [25] introduit une fonction de distribution bimodale dans laquelle une des contributions correspond à des mouvements diffusifs locaux dans l'échantillon, et l'autre à la rotation des groupes méthyles. Ces deux publications prouvent que cette étude est plus que jamais d'actualité et qu'il est nécessaire de disposer de jeux de données plus complets pour valider ou infirmer de tels modèles.

Au cours des travaux sur le projet, il est également apparu que l'intensité élastique au moment transféré zéro, EI(0), était un des paramètres clés pour comparer les échantillons de manière plus précise et obtenir des MSD plus quantitatifs. Avant cette recherche, personne n'avait montré de manière aussi détaillée l'influence de ce point et celui-ci était souvent négligé ou les données normalisées à $Q = 0$, sans préciser la méthode utilisée pour l'interpolation. Comme nous l'avons montré, cette valeur n'est souvent pas proche de la valeur théorique de un et varie en fonction de la composition de l'échantillon, de la température et de l'instrument. Il existe probablement trois raisons pour cela : les effets de diffusion multiple, la contribution de la diffusion cohérente, et les contributions négligées du QENS à l'EINS (voir la section 5.5). Elles n'ont pas été étudiées en détail dans ce travail, mais la nécessité de mieux les comprendre et de les quantifier a été démontrée. Pour être cohérent dans notre analyse, nous avons décidé de fixer la valeur de EI(0) à la valeur obtenue par le modèle de l'AG qui est strictement valide pour les valeurs de Q faibles. Ceci peut avoir des effets considérable sur les MSD, comme nous le montrons dans la Figure 7.1. En respectant cette règle, nous avons pu ajuster les données utilisées par [10] et obtenir un comportement de croissance linéaire des MSD, même à des températures élevées.

Avec les simulations MD de la poudre de protéine A-L, nous avons démontré que les MSD extraits avec trois modèles différents de l'EISF (méthode indirecte) donnaient des résultats similaires au calcul direct du MSD à partir des trajectoires (méthode directe). Comparée aux MSD pour des fenêtres temporelles d'instruments spécifiques, la méthode indirecte déterminait toujours des valeurs plus petites, indiquant une sous-évaluation des MSD. Les résultats qualitatifs des MSD (méthode directe vs indirecte) entre les différentes simulations se sont avérés très similaires et il a de nouveau été montré que l'AG donnait une estimation précise des MSD, même si seulement une petite gamme en Q était utilisée. En revanche, la largeur de la distribution des MSD n'a pas été aussi bien évaluée avec les modèles de Yi et PK. Lorsqu'ils étaient comparés aux résultats expérimentaux, les valeurs quantitatives des MSD étaient du même ordre de grandeur, mais n'étaient pas en parfait accord pour la plupart des cas. Pour les échantillons hydratés, les MSD simulés étaient plus grands, alors que pour les échantillons secs, ils étaient plus petits. Également, de manière qualitative, ils n'étaient pas toujours cohérents. Cela indique qu'il existe encore des différences significatives entre les simulations MD de poudres et les résultats expérimentaux. Les raisons pourraient être les simulations MD elles-mêmes ou des problèmes expérimentaux, ou les deux. La manière dont les données sont corrigées et normalisées est certainement cruciale pour un meilleur accord quantitatif, et les champs

de force des simulations ont encore besoin d'être adaptés à des données expérimentales pour pouvoir décrire les échantillons de la manière la plus réaliste possible. Un bilan définitif ne peut pas être donné car seules deux températures ont été évaluées ici, et pour une comparaison plus précise, il serait nécessaire de calculer un jeu de données complet, comprenant un balayage complet de la température.

Le QENS permet d'approcher les mouvements moléculaires de manière beaucoup plus précise, mais ces mesures sont beaucoup plus longues et ne permettent pas, par exemple, un balayage complet de la température. Cependant, pour améliorer la modélisation des données de diffusion neutronique, des progrès dans le domaine de l'analyse QENS sont nécessaires, et certaines publications vont maintenant dans ce sens.

Récemment, même les bases de l'analyse de la diffusion incohérente de neutrons, les fonctions de corrélation de van Hove dans la limite classique, ont été remises en question : Frauenfelder et al. contestent la négligence des effets quantiques dans la limite classique utilisée depuis plus de 50 ans, qui constitue le fondement de la plupart des efforts de modélisation. Kneller travaille également sur l'amélioration de la limite classique utilisée dans la théorie de Van Hove et affirme que sa nouvelle théorie "fournit [...] une interprétation physique de la version quantique complexe des fonctions de corrélation de Van Hove, et montre notamment qu'il existe une relation physique significative et intuitivement compréhensible entre les versions quantique et classique des fonctions de Van Hove" [139]. Son approche récemment proposée contient une interprétation fondée sur des transitions entre les états quantiques de la cible induites par la diffusion. Ces suggestions sont passionnantes et doivent être testées à l'aide de données expérimentales afin de mieux comprendre les phénomènes de diffusion. En outre, une meilleure description des données QENS améliorera également la compréhension de l'EINS, et permettra de séparer les deux contributions à la diffusion incohérente.

Un autre aspect important est le développement de nouveaux logiciels pour analyser les données. Les nouveaux utilisateurs des grands instruments de neutrons, en particulier, doivent être en mesure d'analyser les données et de comprendre les avantages et les inconvénients des différentes méthodes. Il est important de fournir différents modèles et un logiciel d'évaluation des données cohérent. Par conséquent, de nouveaux modèles doivent être implémentés dans des logiciels tels que Lamp [68] ou Mantid [69], ce qui simplifiera la prise en charge et permettra une évaluation des données plus transparente. De plus, des logiciels librement accessibles permettraient aux utilisateurs d'avoir un aperçu du traitement des données et de contribuer à éliminer les éventuelles erreurs du code. Ces logiciels devraient idéalement prendre en compte tous les effets possibles sur les évaluations telles que l'hétérogénéité, l'anharmonicité, la diffusion cohérente, la diffusion multiple, QENS, etc., comme décrit dans cette thèse. En outre, ils devraient fournir une bonne documentation pour pouvoir comprendre la théorie et le code sous-jacent.

Enfin, une nouvelle instrumentation ou de nouvelles installations à flux neutronique plus élevé, telles que l'European Spallation Source ESS, peuvent contribuer au développement de nouvelles techniques permettant de valider la modélisation de la diffusion des neutrons. À long terme, cela nous aidera à réduire l'écart entre simulations et expériences, ce qui permettra de mieux comprendre la dynamique des biomolécules.

Appendices

A. Additional information

Here the detailed amino-acid composition of Alpha-Lactalbumin is shown. Further, two tables with the measured weights of the prepared A-Lsamples, Batches 1 and 2, are presented.

Table A.1. - Amino Acids - Name, Symbols, chem. formulas. Values are taken from [140]. The amino acid residues sequence for A-L is [1-letter symbols]: EQLTKCEVFR ELKDLKGYGG VSLPEWVCTT FHTSGYDTQA IVQNNDSTEY GLFQINNKIW CKDDQNPHSS NICNISCDKF LDDDLTDDIM CVKKILDKVG INYWLAHKAL CSEKLDQWLC EKL.

Long Name	3-Letter Symbol	1-Letter Symbol	chem. formula
Aliphatic Amino Acids with Hydrophobic Side Chain			
Alanine	ALA	A	$C_3H_7N_1O_2$
Isoleucine	ILE	I	$C_6H_{13}N_1O_2$
Leucine	LEU	L	$C_6H_{13}N_1O_2$
Valine	VAL	V	$C_5H_{11}N_1O_2$
Aromatic Amino Acids with Hydrophobic Side Chain			
Phenylalanine	PHE	F	$C_9H_{11}N_1O_2$
Tryptophan	TRP	W	$C_{11}H_{12}N_2O_2$
Tyrosine	TYR	Y	$C_9H_{11}N_1O_3$
Amino Acids with Neutral Side Chain			
Asparagine	ASN	N	$C_4H_8N_2O_3$
Cysteine	CYS	C	$C_3H_7N_1O_2S_1$
Glutamine	GLN	Q	$C_5H_{10}N_2O_3$
Methionine	MET	M	$C_5H_{11}N_1O_2S_1$
Serine	SER	S	$C_3H_7N_1O_3$
Threonine	THR	T	$C_4H_9N_1O_3$
Amino Acids with Positive Charged Side Chain			
Arginine	ARG	R	$C_6H_{14}N_4O_2$
Histidine	HIS	H	$C_6H_9N_3O_2$
Lysine	LYS	K	$C_6H_{14}N_2O_2$
Amino Acids with Negative Charged Side Chain			
Aspartic Acid	ASP	D	$C_4H_7N_1O_4$
Glutamic Acid	GLU	E	$C_5H_9N_1O_4$
Unique Amino Acids			
Glycine	GLY	G	$C_2H_5N_1O_2$
Proline	PRO	P	$C_5H_9N_1O_2$

Table A.2. - Weights for A-L, Batch 1. Measured and calculated values for the weight in [g] of the alpha-lactalbumin samples without calcium (A-L_{dep}) for Batch 1 prepared in November 2016.

Batch 1 – Final values

			Prepared in November 2016
calculated values in yellow	all numbers are weights in gram		
sample holder thickness = (3 - A) mm	A=2.3		A=2.3
CA depleted samples	dry	0.40000	0.80000
weight of sample holder + indium wire	22.4395	21.7895	21.4844
weight of dry sample before hydration	22.5373	21.8920	21.5859
amount of dry sample	**0.0978**	**0.1025**	**0.1015**
weight after hydration before closing	22.5373	21.9330	21.6718
Total amount of sample	0.0978	0.1435	0.1874
amount of D2O	0.0000	0.0410	0.0859
ratio g D2O / g Protein	**0.0000**	**0.4000**	**0.8463**
total weight after closing	37.9936	37.4143	38.5682

Table A.3. - Weights for A-L, Batch 2. Measured and calculated values for the weight in [g] of the alpha-lactalbumin samples with and without calcium (A-$L_{ca}$and A-L_{dep}) for Batch 2 prepared in January 2017.

Batch 2 – Final values

Prepared in January 2017

calculated values in yellow	all numbers are weights in gram		
sample holder thickness = (3 - A) mm	A=2.3	A=2.4	A=2.4
CA saturated samples	dry	0.4	0.8
weight of sample holder + indium wire	24.77000	24.85720	22.08916
weight of dry sample before hydration	24.84845	24.95200	22.19490
amount of dry sample	**0.07845**	**0.09480**	**0.10574**
weight after hydration before closing	24.84845	24.98986	22.27990
Total amount of sample	0.07845	0.13266	0.19074
amount of D2O	0.00000	0.03786	0.08500
ratio g D2O / g Protein	**0.00000**	**0.39937**	**0.80386**
total weight after closing	40.88985	40.43023	40.11920

sample holder thickness = (3 - A) mm	A=2.3	A=2.4	A=2.4
CA depleted samples	dry	0.40000	0.80000
weight of sample holder + indium wire	22.45261	22.10866	21.81679
weight of dry sample before hydration	22.56280	22.22460	21.92870
amount of dry sample	**0.11019**	**0.11594**	**0.11191**
weight after hydration before closing	22.56280	22.27040	22.01860
Total amount of sample	0.11019	0.16174	0.20181
amount of D2O	0.00000	0.04580	0.08990
ratio g D2O / g Protein	**0.00000**	**0.39503**	**0.80332**
total weight after closing	38.35207	39.85973	37.9879

B. Publications

In the following, four refereed publications to which I contributed during my research work are listed.

B.1. Analysis of incoherent neutron scattering data beyond the Gaussian approximation

To the following publication of Zeller et al. [33] my contribution was

- preparation of the samples,
- the design and execution of all experiments together with co-authors,
- the data reduction, correction and evaluation,
- the preparation of all figures in the publication,
- the writing of the manuscript.

THE JOURNAL OF CHEMICAL PHYSICS **149**, 234908 (2018)

Analysis of elastic incoherent neutron scattering data beyond the Gaussian approximation

D. Zeller,[1,2,3] M. T. F. Telling,[3] M. Zamponi,[4] V. García Sakai,[3] and J. Peters[1,2,a)]
[1]*Université Grenoble Alpes, CNRS, LiPhy, 140 av. de la Physique, 38000 Grenoble, France*
[2]*Institut Laue-Langevin, 71 Avenue des Martyrs, 38042 Grenoble Cedex 9, France*
[3]*ISIS Pulsed Neutron and Muon Source, STFC Rutherford Appleton Laboratory, Chilton, Oxfordshire OX11 0QX, United Kingdom*
[4]*Forschungszentrum Jülich GmbH, Jülich Centre for Neutron Science (JCNS) at Heinz Maier-Leibnitz Zentrum (MLZ), Lichtenbergstr. 1, 85748 Garching, Germany*

(Received 26 July 2018; accepted 27 November 2018; published online 21 December 2018)

This work addresses the use of the Gaussian approximation as a common tool to extract atomic motions in proteins from elastic incoherent neutron scattering and whether improvements in data analysis and additional information can be obtained when going beyond that. We measured alpha-lactalbumin with different levels of hydration on three neutron backscattering spectrometers, to be able to resolve a wide temporal and spatial range for dynamics. We demonstrate that the Gaussian approximation gives qualitatively similar results to models that include heterogeneity, if one respects a certain procedure to treat the intercept of the elastic intensities with the momentum transfer axis. However, the inclusion of motional heterogeneity provides better fits to the data. Our analysis suggests an approach of limited heterogeneity, where including only two kinds of motions appears sufficient to obtain more quantitative results for the mean square displacement. Finally, we note that traditional backscattering spectrometers pose a limit on the lowest accessible momentum transfer. We therefore suggest that complementary information about the spatial evolution of the elastic intensity close to zero momentum transfer can be obtained using other neutron methods, in particular, neutron spin-echo together with polarization analysis. *Published by AIP Publishing.* https://doi.org/10.1063/1.5049938

I. INTRODUCTION

Neutron scattering techniques have been successfully employed to study biological systems for about 50 years.[1–3] Neutron scattering from any sample gives rise to two contributions, a coherent and an incoherent part.[4] Whereas coherent scattering requires a constant phase relation to be maintained between the incident and scattered neutron, giving access to structural or collective motional information, incoherent scattering is a probe of the average motions of individual particles within the sample and thus molecular dynamics.[4] In the present work, we concentrate on the incoherent part assuming that most of the signal comes from the dominant incoherent scattering from the sizable number of protons that are distributed throughout the sample. The incoherent scattering signal can be further subdivided into an elastic part, corresponding to those atoms in the sample that move with a characteristic time scale that falls outside the temporal resolution of the neutron instrument used and thus are not seen to exchange energy with the neutron upon scattering, called the Elastic Incoherent Neutron Scattering (EINS), and those whose motions lead to an energy exchange with the neutron. Traditionally, one distinguishes here quasi-elastic neutron scattering (QENS), where only small amounts of energy are exchanged (typically <2 m eV) and which manifests itself as a broadening of the zero-energy transfer centered elastic peak, from inelastic neutron scattering (INS), which appears as satellite peaks well separated from the elastic part and correspond to specific modes or excitations within the sample.[4] For the analysis of QENS data, the approach suggested by Van Hove[5] based on pair correlation functions is most commonly applied and is nearly the exclusive model used so far. Only recently, alternative approaches were suggested by Frauenfelder *et al.*[6] and Kneller.[7] However, new formalisms have to be tested and validated against experimental data, which is done in the present study for EINS.

EINS is generally used to compare the dynamical behavior of biomolecules, especially proteins.[3,8,9] For this, mean square displacements (MSDs) of the atoms in the sample are extracted over a wide temperature range. The MSDs are considered as a measure of the sample flexibility at a given temperature.[10] Most biomolecules undergo transitions, which are also visible in the MSD by a change of the slope. For instance, proteins show a dynamical transition at around 220 K which characterizes the crossover from harmonic to anharmonic motions.[2] Zaccai introduced an empirical parameter, known as the effective force constant or resilience,[11] and it corresponds to the slope of the MSD before and after the dynamical transition, which is used to quantify the protein's stability in a given state. Similarly, lipids undergo structural phase transitions which can also be detected by a variation in the dynamics through a change in the MSD.[12] Thus, EINS is a standard mode of measurement on neutron backscattering

a)Electronic mail: jpeters@ill.fr

spectrometers. Despite such a common use, scientists often compare the MSD obtained from a set of samples measured under the same conditions and on the same instrument, only. This is because the obtained absolute values of the MSDs are not always suited for a quantitative comparison with results from other techniques. Moreover, if one searches for specific motions whose amplitudes are known, as for instance to differentiate 2- or 3-site jump motions,[13] precise spatial information would be highly desirable. The aim of the present study is thus to improve the data evaluation and to obtain more certainty about the quantitative accuracy of MSD analysis and results. Such results could, in addition, be used to validate molecular dynamics simulations and force fields, as both approaches give access to the same time scales as neutron scattering.

To date, there is no complete model taking into account all possible dynamic contributions to the EINS. As a standard, the Gaussian Approximation (GA) is used to extract the MSD of the elastic data.[14] The limits of the applicability of the models are not always clear and respected such that the way in which they are implemented may seem sometimes arbitrary with full details lacking in many publications. As long as different samples are analyzed in exactly the same way and compared to each other, it may appear to be consistent, but it makes it difficult to compare the results from different samples, spectrometers, and experimental groups.[15] In particular, using different neutron spectrometers is important for two main reasons. First, each instrument has a specific energy resolution providing access to different dynamical time scales, and second, each instrument has a specific range of accessible scattering angles, which define the spatial information probed. For the first effect, Doster *et al.* proposed an analysis via elastic resolution spectroscopy in 2001[16] and connected it to temperature scans via a scaling function in 2013.[8]

While the different temporal regimes' information accessed by different neutron spectrometers (different energy resolutions) makes them complementary, the full spatial information accessible on the different instruments due to the range of scattering angles (transformed into momentum transfer Q in reciprocal space) is often redundant, mainly due to theoretical applicability of the GA. Indeed, the GA is in general only valid in a restricted region of Q, specifically at low Q values,[17] since it is neglecting any effects from anharmonicity, heterogeneity, and anisotropy.[18] To date, limited efforts have been devoted to developing approaches that model EINS data over a wider momentum transfer window, explore the effects of these three aspects, and question whether extra information is contained in the higher Q regime. Anharmonicity of motions was introduced in a model by Doster *et al.*[2] and described in terms of a double-well potential. The introduction of heterogeneity of motions found to be the main reason for the non-Gaussian behavior[19] has been approached in two ways. In the publications of Becker and Smith[20] and more recently Yi *et al.*,[21] a second term of a series expansion is included to describe the standard deviation of the MSD, whereas the GA uses only the first term of a cumulant expansion. Similar expansions were already proposed earlier by Rahman[22] and Sköld *et al.*[23] although they did not relate them explicitly to dynamical heterogeneity. The alternative approach consists in describing the heterogeneity by assuming different forms of the distribution. Nakagawa *et al.* compared a bimodal, exponential, and Gaussian distribution.[24] Meinhold *et al.* used a Weibull distribution,[25] and Peters and Kneller used a gamma distribution[26] based on the work of Kneller and Hinsen.[27] Most recently, a bimodal Gaussian distribution was investigated by Vural *et al.*[28] Doster also addressed the issue of protein dynamical heterogeneity by suggesting three different approaches to it.[29] His main conclusion was to say that data can be fitted as successful when heterogeneity is reduced to two kinds of molecular processes, one corresponding to translational and the other to rotational (coming mainly from methyl groups) motions, the latter being non-Gaussian scattering processes. The former models assume, by contrast, a distribution of purely Gaussian motions, which according to Refs. 19 and 30 is also a good approximation when the motions are anharmonic. All the aforementioned models can describe EINS data over an extended Q-range to a greater or lesser extent. In the present study, we will therefore investigate four different models to analyze the same set of EINS data to not only substantiate limits of applicability but also arrive at their advantages and disadvantages by considering data from different instruments and under different experimental conditions. The models to be assessed are GA, Yi *et al.* (Yi), Peters *et al.* (PK), and Doster *et al.* (Do), and a subset is chosen to represent the different approaches.

As the experimental dataset, we choose to measure the MSD of a simple protein powder at different levels of hydration on various instruments. We selected a small commercially available protein: Alpha-lactalbumin (A-L). It is a globular protein (molecular weight of 14.2 kDa), which regulates lactose synthase and can be found in all mammalian milk. It can strongly bind calcium Ca^{2+}, but also other metal ions like magnesium. The hydration conditions chosen are ≈0 (dry), 0.4, and 0.8 g D_2O/g protein. Data were collected using three neutron backscattering spectrometers covering a wide temporal range, IN13 at the Institut Laue Langevin (ILL) in Grenoble, France, SPHERES operated by JCNS at the Heinz Maier-Leibnitz Zentrum (MLZ) in Garching, Germany, and OSIRIS at the STFC Rutherford Appleton Laboratory ISIS in Oxfordshire, UK. They correspond to the most common type of neutron backscattering instruments used for EINS experiments. SPHERES and OSIRIS use crystal analyzers that reflect cold neutrons (λ of 6.27 and 6.66 Å, respectively) and therefore access a limited Q-range (up to 1.8 $Å^{-1}$), whereas IN13 uses a thermal neutron crystal analyzer (λ of 2.23 Å), resulting in the largest available Q-range for a neutron backscattering instrument (up to 4.9 $Å^{-1}$). In addition, the instruments have different energy resolutions allowing access to motions from a few picoseconds to a few nanoseconds.

The aim of this study is to test the applicability and determine the limitations of four distinct EINS models with respect to different samples and neutron instruments. As we will show, in some cases, clear conclusions can be made, and in others, we simply point out the limitations and let the reader choose the best data analysis method according to priorities imposed by the experiment.

II. MATERIALS AND METHODS

A. Sample preparation

The sample, Ca^{2+} depleted A-L, for which all calcium atoms have been purged from A-L, was bought from Sigma-Aldrich as a lyophilized powder and was used without further purification. According to the supplier, the powder might contain traces of ammonium sulfate or sodium phosphate. Such buffers can contribute to small spurious effects at low temperatures,[8] but such investigation was beyond the scope of the study. Here, we compared three different levels of hydration, i.e., h ≈ 0 (dry), h = 0.4, and h = 0.8, with h designating g D_2O/g of dry powder. The dry sample represents the case where only harmonic motions are present up to room temperature. 0.4h corresponds to about one or two layers of water on the protein surface,[31] sufficient to allow for anharmonic motions above the dynamical transition temperature. 0.8h represents a gel close to full hydration. In neutron experiments, one often uses D_2O instead of H_2O as the solvent since the scattering contribution will be small compared to the very large incoherent scattering contribution from the hydrogens within the protein.[32] To prepare the hydrated samples, the lyophilized dry protein powder was placed in an open flat aluminum sample holder and in a desiccator over D_2O and the final hydration level was checked by weighing the sample periodically. When the desired uptake was reached, the sample holder was vacuum sealed with an indium wire. Aluminum is the preferential material for this type of neutron experiments since it is mainly transparent to neutrons, and thus, its contribution to the scattering signal will be negligible. The amount of protein in each sample was close to 100 mg. The sample transmissions, as measured on the IN13 spectrometer, were between 92% and 96%, and therefore, no correction for multiple scattering was applied.

B. Elastic incoherent neutron scattering experiment

The measurements were carried out on IN13[33] (data are available in Ref. 34), SPHERES,[35,36] and OSIRIS[37] with the characteristic energy resolution ΔE_{res} and maximal used momentum transfer, Q, ranges being given in Table I. SPHERES is a reactor-based neutron backscattering spectrometer which uses Si111 analyzer crystals to achieve a sub-μeV resolution, similar to IN16B at the ILL and HFBS at the NIST Center for Neutron Research in Gaithersburg, MD, USA. Due to the instrument design constraints, these spectrometers suffer from slightly degraded energy resolution at the low angles and sometimes lower flux in the first detectors. OSIRIS is a time-of-flight near-backscattering neutron spectrometer at a spallation source. It utilizes PG002 analyzer crystals, and while it affords lower energy resolution than SPHERES, it has a larger neutron flux at the sample position and constant ΔE across the detector bank. Finally, IN13 is a reactor-based neutron backscattering spectrometer which uses thermal neutrons and a $CaF_2(422)$ analyzer. This analyzer type results in a much larger Q range, compared to the OSIRIS and SPHERES instruments, but, consequently, it does suffer worse Q resolution across the detector bank.

All samples were initially cooled to cryogenic temperatures (<20 K), and then measurements are taken on warming. At IN13 and SPHERES, the data were collected while applying a continuous heating gradient. At OSIRIS, heating was done in steps of 10 K or 20 K and the measurement was then performed at constant temperature once the sample temperature had equilibrated. The data were reduced with LAMP,[38] Mantid[39] (v3.11.0), and Slaw (http://apps.jcns.fz-juelich.de/man/slaw.html). Usual data corrections implemented in these packages were applied, including a correction for detector efficiency and subtraction of the empty sample holder. The resulting intensities were normalized to the intensity at the lowest available temperature to correct for any geometrical effects.

C. Theoretical background and used models

The atomic motions in a given sample, as measured by neutron spectroscopy, are described by the so-called incoherent dynamic scattering function (DSF) $S_{inc}(\mathbf{Q}, \omega)$,

$$S_{inc}(\mathbf{Q}, \omega) = \frac{1}{2\pi} \int_{\infty}^{-\infty} I_{inc}(\mathbf{Q}, t) \exp(-i\omega t) \mathrm{d}t, \quad (1)$$

where $\mathbf{Q}$ is the momentum transfer, ω is the energy transfer from the neutron to the sample in units of $\hbar$, and $I(\mathbf{Q}, t)$ is the incoherent intermediate scattering function for one type of atom in the classical limit. It is defined as[4]

$$I_{inc}(\mathbf{Q}, t) = \frac{1}{N} \sum_{j=1}^{N} \left\langle \exp\left\{ i\mathbf{Q} \cdot \mathbf{r}_j(t) \right\} \exp\left\{ -i\mathbf{Q} \cdot \mathbf{r}_j(0) \right\} \right\rangle, \quad (2)$$

where N is the number of atoms in the sample, $\mathbf{r}_j$ is the position of each individual atom α, and $\langle\rangle$ depicts the statistical average of the ensemble. In this study, only H atoms are considered since they have by far the largest incoherent cross section in a protein.[32]

In the case of confined motion, as is the case in low hydrated powders, the $I_{inc}(\mathbf{Q}, t)$ tends to a plateau value for $t \to \infty$,

TABLE I. Characteristics and sample temperatures measured on each instrument. The time window Δt_{res} is calculated using Heisenberg's uncertainty principle $\Delta t_{res} \cdot \Delta \omega_{res} \geq \hbar$.

Instrument	Max. used Q range (Å^{-1})	$\Delta\omega_{res}$ (μeV)	Δt_{res}	T range (K)	Heating
OSIRIS (ISIS)	0.30–1.5	25	≈25 ps	10–310	Steps
IN13 (ILL)	0.49–4.5	8	≈0.1 ns	20–315	Gradient
SPHERES (MLZ)	0.27–1.9	0.7	≈1 ns	3–310	Gradient

$$\lim_{t\to\infty} I_{\mathrm{inc}}(\mathbf{Q},t) = I_{\mathrm{inc}}(\mathbf{Q},\infty) = \frac{1}{N}\sum_{j=1}^{N}\left\langle \exp\left\{-i\mathbf{Q}\cdot\mathbf{r}_j\right\}\right\rangle\left\langle \exp\left\{i\mathbf{Q}\cdot\mathbf{r}_j\right\}\right\rangle. \quad (3)$$

Here 2 things were assumed: (1) At $t = \infty$, $\mathbf{r}_j(\infty)$ and $\mathbf{r}_j(0)$ are statistically independent and (2) $\mathbf{r}_j(t)$ is translationally invariant in time such that $\mathbf{r}_j(\infty) = \mathbf{r}_j(0) = \mathbf{r}_j$. Equation (3) can be rewritten using a cumulant expansion of $\left\langle \exp\left\{\pm i\mathbf{Q}\cdot\mathbf{r}_j\right\}\right\rangle$,

$$I_{\mathrm{inc}}(\mathbf{Q},\infty) = \frac{1}{N}\sum_{j=1}^{N}\exp\left\{2\left[-Q^2\rho_2(\mathbf{n_q}) + Q^4\rho_4(\mathbf{n_q}) - Q^6\rho_6(\mathbf{n_q}) + \cdots\right]\right\}, \quad (4)$$

where ρ_i are the moments

$$\rho_2(\mathbf{n_q}) = \frac{1}{2!}\left\langle(\mathbf{n_q}\cdot\mathbf{r}_j)^2\right\rangle, \quad (5)$$

$$\rho_4(\mathbf{n_q}) = \frac{1}{4!}\left\{\left\langle(\mathbf{n_q}\cdot\mathbf{r}_j)^4\right\rangle - 3\left\langle(\mathbf{n_q}\cdot\mathbf{r}_j)^2\right\rangle^2\right\}, \quad (6)$$

$\ldots$,

$Q = |\mathbf{Q}|$, and $\mathbf{n_q}$ is the unit vector along $\mathbf{Q}$. In energy space, this gives rise to the elastic incoherent structure factor (EISF) $S_{\mathrm{inc}}(\mathbf{Q}, \omega = 0)$,

$$S_{\mathrm{inc}}(\mathbf{Q},\omega) = \int_{-\infty}^{\infty} \mathrm{d}t\,\exp(i\omega t) I(\mathbf{Q},t) R(t), \quad (7)$$

$$S_{\mathrm{inc}}(\mathbf{Q},\omega\approx 0,\Delta\omega_{\mathrm{res}} = 1/\tau_{\mathrm{res}}) = \int_{-\infty}^{\infty} \mathrm{d}t I(\mathbf{Q},t)R(t) = A I_{\mathrm{inc}}(\mathbf{Q},\tau_{\mathrm{res}}), \quad (8)$$

$$S^{\mathrm{n}}_{\mathrm{inc}}(\mathbf{Q},\omega\approx 0) \equiv \frac{S_{\mathrm{inc}}(\mathbf{Q},\omega\approx 0,\Delta\omega_{\mathrm{res}})}{S_{\mathrm{inc}}(\mathbf{Q}=0,\omega\approx 0,\Delta\omega_{\mathrm{res}})} = I_{\mathrm{inc}}(\mathbf{Q},\tau_{\mathrm{res}}) \approx I_{\mathrm{inc}}(\mathbf{Q},\infty), \quad (9)$$

where $R(t)$ is the resolution function of the instrument with a width of $\tau_{\mathrm{res}} = \hbar/\omega_{\mathrm{res}}$ due to which the integration over $I(Q, t)$ is truncated.[8,16] A is a convenient normalization factor to take into account the integration over the instrumental resolution. Here it is important to mention that the plateau value of $t \to \infty$ is only reached for an ideal instrument with an energy resolution of $\Delta\omega_{\mathrm{res}} \to 0$. Since a value of zero is impossible to reach, experimentally the value of $S_{\mathrm{inc}}(\mathbf{Q}, \omega \approx 0)$ is dependent on the resolution of the instrument, and therefore to observe different dynamics, one needs to combine experiments performed using different spectrometers.[40,41] In the following, we suppress the explicit mention of the instrumental resolution, as done in Eq. (8). Additionally, the QENS signal at zero frequency ($\omega = 0$) was neglected.

The Gaussian Approximation (GA) uses only the first non-zero cumulant such that

$$S^{\mathrm{n}}_{\mathrm{inc}}(Q,\omega\approx 0) \approx \exp\left(-\frac{1}{3}Q^2\left\langle r^2\right\rangle_{\mathrm{GA}}\right), \quad (10)$$

where $\left\langle r^2\right\rangle$ is the atomic mean square displacement (MSD), which corresponds to the average of the amplitudes of motion of all atoms within the sample. The GA always holds in the case of $Q^2\left\langle r^2\right\rangle_{\mathrm{GA}} \ll 1$; since then, the higher cumulant terms evaluate to zero, and thus, the GA is exact for some specific cases, i.e., an harmonic solid or an ideal gas. It is by far the most commonly used model in publications about protein dynamics investigated with EINS[3] and implemented by fitting $\ln[S(Q)]$ vs Q^2. However, the choice of the Q-range used to fit the data and the value of $\ln[S(Q)]$ at $Q^2 = 0$ Å^{-2} is not always specified nor discussed in publications, neither is the limit of $\left\langle r^2\right\rangle_{\mathrm{GA}} Q^2_{\mathrm{max}} \leq 1$ justified.[42] Furthermore, sometimes this limit is surpassed and the usage of a larger Q^2_{max} is justified if there is an extended linear behavior of the data points over wider Q-ranges. However a clear rule is not well documented.

It is important to note that the factor of 1/3 in front of the MSD is due to the fact that elastic scattering is time independent and a plateau value for the ISF is reached (which is dependent on the instrument resolution, as mentioned above). For time resolved measurements [MSD(t)], the factor evaluates to 1/6.

The assumption of the Gaussian approximation for an individual atom j is a good approximation beyond low Q out to $Q \gtrsim 4$ Å^{-1} for almost all atoms according to Tokuhisa *et al.*[19] and Vural *et al.*,[30] as the higher order terms are negligible compared to the Q^2 term. Moreover, Gaussian heterogeneity is able to describe approximately non-Gaussian (for instance rotational) motions, whose importance is well documented by Liu *et al.*[43] The two main reasons, why the GA might not be a valid approximation for an individual atom, are large anharmonic and anisotropic effects. By contrast, Kneller and Chevrot[18] claim that models could be improved by accounting for anisotropy, but due to the lack of precision of today's available data (neither experimental nor with simulations), additional parameters cannot be fit unambiguously. In the following, we therefore use two heterogeneity models to verify these assumptions.

The first heterogeneity model will be referred to as the Yi model. Applying similar expansions as suggested earlier by Rahman[22] and Sköld *et al.*,[23] Becker and Smith[20] introduced a model beyond the Gaussian approximation within the context of dynamical heterogeneity. Yi *et al.*[21] recently proposed it as a "simple correction" of the GA which can be written as

$$S^{\mathrm{n}}_{\mathrm{inc}}(Q,\omega\approx 0) \approx \exp\left(-\frac{1}{3}Q^2\left\langle r^2\right\rangle_{\mathrm{Yi}}\right)\cdot\left(1+\frac{Q^4}{18}\sigma^2_{\mathrm{Yi}}\right), \quad (11)$$

where σ_{Yi} describes the standard deviation of the MSD $\left\langle r^2\right\rangle_{\mathrm{Yi}}$. It is similar to the expansion of the fourth cumulant term [Eq. (6), second non-zero term] and the same for $\frac{Q^4}{18}\sigma^2_{\mathrm{Yi}} \ll 1$. The introduction of Q^4 accounts for the heterogeneity of the motions, and the spread of the MSD is described via σ_{Yi}, without assuming any functional form for the distribution. Since it does not contain higher order terms further than Q^4, its applicability is limited in Q.

The second heterogeneity model that we will consider is that introduced by Peters and Kneller[26] (referred to as the PK model from here on) which uses a gamma function to describe the distribution of mean square position fluctuations (MSPFs). It is mathematically written as

$$S^{\mathrm{n}}_{\mathrm{inc}}(Q,\omega\approx 0)\approx \frac{1}{\left(1+\frac{Q^2\langle r^2\rangle_{\mathrm{MSPF}}}{3\beta}\right)^{\beta}}, \tag{12}$$

$$\sigma_{\mathrm{MSPF}}=\frac{\langle r^2\rangle_{\mathrm{MSPF}}}{\sqrt{\beta}}, \tag{13}$$

where $\langle r^2\rangle_{\mathrm{MSPF}}$ describes the mean of the individual MSPF and σ_{MSPF} is its standard deviation. The parameter β describes the distribution of the MSPF with a gamma function. In the case of $\beta\rightarrow\infty$, the GA is retrieved. For a constant $\langle r^2\rangle_{\mathrm{MSPF}}$, the heterogeneity increases with decreasing β. This model poses no Q-limitation since it includes all higher order terms of the cumulant expansion as long as the individual atoms can be described by the Gaussian term alone and under the assumption that heterogeneity is the only reason for the higher order terms.

As can be seen, the standard deviations introduced for the Yi and PK models are strongly correlated to the respective definitions of the MSD and MSPF, and although a comparison, at least qualitatively, is certainly possible, a detailed study of the additional information they provide is beyond the scope of this paper, where we concentrate on the MSD itself.

The last model used in this study is that introduced by Doster *et al.*[2] In the same way as the PK model, it can be used to describe the entire Q range and is based on a double-well potential model to describe the anharmonicity of atomic motions. In this model, each hydrogen atom can be found in one of the two different harmonic wells, separated by a distance d and by a free energy difference ΔG. The model will be referred to as the Do model from here on and is mathematically expressed as

$$S^{\mathrm{n}}_{\mathrm{inc}}(Q,\omega\approx 0)\approx \exp\left(-\frac{1}{3}Q^2\left\langle r^2\right\rangle_{\mathrm{Do},G}\right)\times(1-2p_{12}(1-\mathrm{sinc}(Qd))), \tag{14}$$

where the first term with $\left\langle r^2\right\rangle_{\mathrm{Do},G}$ describes the Gaussian contribution to the MSD and the second term describes the two state model. p_{12} is the product of p_1 and p_2 which denote the probability of finding an atom in the ground or excited state, respectively, with $p_2/p_1\propto\exp(-\Delta G/RT)$. The total MSD is defined as

$$\frac{\left\langle r^2\right\rangle_{\mathrm{Do,tot}}}{3}=-\left(\frac{d\ln[S(Q,\omega=0)]}{d(Q^2)}\right)_{Q=0}=\frac{\left\langle r^2\right\rangle_{\mathrm{Do},G}}{3}+\frac{p_{12}d^2}{3}. \tag{15}$$

In the case of only one well, $p_{12}=0$ or $d=0$ such that the GA is retrieved.

An alternative way of comparing EINS data is through the evaluation of the elastic neutron intensities summed over all (or a range of) accessible Q-values I_{sum}. In this way, a comparison can be made without using a model. The summed intensities are much less affected by errors, so they give a more accurate overview over the dynamics taking place within the time scales probed by the spectrometer. Nevertheless, in the limit of the GA, it is possible to relate the I_{sum} to the inverse of the square of the MSD,

$$I_{\mathrm{sum}}\overset{\mathrm{GA}}{=}\int_{Q'_{\mathrm{max}}}^{Q'_{\mathrm{min}}}\exp\left(-\frac{1}{3}\left\langle r^2\right\rangle' Q'^2\right)\mathrm{d}Q' \tag{16}$$

$$\Rightarrow\left\langle r^2\right\rangle'\overset{\mathrm{GA}}{=}\frac{1}{I^2_{\mathrm{sum}}}\cdot C^2,$$

with

$$C=\frac{\sqrt{3\pi}}{2}\left[\mathrm{erf}\left(Q'_{\mathrm{max}}\right)-\mathrm{erf}\left(Q'_{\mathrm{min}}\right)\right], \tag{17}$$

where $Q'=Q/l$ and $\left\langle r^2\right\rangle'=\left\langle r^2\right\rangle/l^2$ are the dimensionless quantities and l is chosen as a typical length scale, i.e., 1 Å. erf is the error function.[44] In the limit of experimental precision, the I_{sum} is discrete and can be evaluated for each measured temperature T through

$$I_{\mathrm{sum}}(T)=\sum_{i=Q_{\mathrm{min}}}^{Q_{\mathrm{max}}}I_i(T), \tag{18}$$

$$\left\langle r^2\right\rangle_{\mathrm{SumI}}(T)\overset{\mathrm{GA}}{\propto}\frac{1}{I^2_{\mathrm{sum}}(T)}\ \text{Å}^2 \tag{19}$$

where Q_{min} and Q_{max} are defined by the validity of the GA or given by the instrument.

III. METHODOLOGY

In order to compare the models to the experimental data, the resolution broadened DSF $S_{\mathrm{inc}}(Q,\omega\approx 0)$ is normalized by the measured DSF $S_{\mathrm{inc}}(Q,\omega\approx 0,T=\mathrm{T_{low}})$ at the lowest measured temperature $\mathrm{T_{low}}$ where the mobility is assumed to be very close to zero,

$$\frac{S_{\mathrm{inc}}(Q,\omega\approx 0,T)}{S_{\mathrm{inc}}(Q,\omega\approx 0,T=\mathrm{T_{low}})}=S_{\mathrm{norm}}(Q,\omega\approx 0,T)\equiv\mathrm{EI}(Q,T). \tag{20}$$

The models are applied to this normalized DSF which for simplicity is labeled as EI(Q) from now on.

All models were implemented in Python with the help of the LMFIT-package.[45] A least-squares fitting procedure was used taking into account the error of counting statistics $\epsilon=\sqrt{\text{no of counts}}$. This fitting method minimizes the following formula:

$$\chi^2=\sum_{i=1}^{N}\frac{\left(y_i^{\mathrm{data}}-y_i^{\mathrm{model}}\right)^2}{\epsilon_i^2}, \tag{21}$$

where N is the total number of data points, y_i^{data} are the experimental data, and y_i^{model} are the values obtained by the fitting model. The reduced value of $\chi^2_{red}=(\chi^2/\text{no of free parameters})$ can then be used to quantify the quality of the fit.

As described in Sec. II C, all models used in the comparison have different Q-limitations. The PK and Do models can be used to fit the entire available Q-range for all instruments. The PK model will be applied here to all instruments, whereas the Do model is only shown for IN13 as it has the broadest Q-range. The Yi model can also fit a broad Q-range, but in cases of fast decay of the EISF with increasing momentum transfer Q, the expansion of the GA to $\mathrm{Q^4}$ is limited as it neglects higher

order terms. Therefore, we introduce a cutoff to the Yi model, similar to the GA, after which the EISF at high Q-values will not be fitted anymore. To use a consistent Q-range between the three different hydrated A-L samples, the following procedure is used:

First, a lowest Q-value, Q_{min}, is chosen to be the same for fitting all models to a dataset from a given instrument. Then the maximum Q-value Q_{max} is evaluated for the most hydrated sample (0.8 A-L) at the highest temperature T_{max} since the decay of the EISF with increasing Q is the largest. For the GA, Q_{max} is determined as the last Q-value where the fit ln[EI(Q)] vs Q^2 is still linear. The Yi model can describe the entire available Q-range for the OSIRIS and SPHERES instruments. However, for IN13 (up to 4.5 Å^{-1}), the model reaches its limit as mentioned before and a cut-off value at $Q_{max,Yi}$ = 2.5 Å^{-1} is introduced. It provides the best compromise between including the widest range of Q-values and yet still describing the experimental data satisfactorily for the 0.8h sample. The Q_{max} evaluated for each model and each instrument was then the same for all samples and temperatures. The consequences due to the limited number of data points in the available Q-range are elaborated in Sec. III B.

Another important point is the analysis of different Q-ranges which is covered in Sec. III C. A summary of all different Q-ranges evaluated with the GA, Yi, PK, and Do model is found in Table II.

TABLE II. Different Q-ranges used for the various models and instruments.

	Q-range (Å^{-1})			
Model	GA	Yi	PK	Do
IN13	0.5 – 1.7	0.5–2.5	0.5–4.5	0.5–4.5
	1.7 – 4.5	...		...
SPHERES	0.34–0.6	0.34–1.8		...
	0.60–1.2	0.60–1.8		...
	0.96–1.8	...		...
OSIRIS	0.29-1.5			...

A. The question of the intercept EI(Q = 0 Å^{-1})

All models should start with the same value of the EI(Q) at zero momentum transfer. As discussed in Sec. III A, the theoretical value should be 1 for the normalized DSF EI(Q). Due to instrumental and experimental effects, like multiple scattering or coherent effects, the starting value is often lower than 1, especially at higher temperatures. For this reason, the value of EI(Q = 0 Å^{-1}) is introduced as a fitting parameter ≤1 and its consequences are evaluated here. In most publications about EINS data, the value at EI(Q = 0 Å^{-1}) is not clearly defined. On the contrary, in the case of the normally used linear fit of ln[EI(Q)] vs Q^2, only the slope (∝MSD) is reported and not the intercept with the y-axis (= $\ln\left[\text{EI}(Q = 0\,\text{Å}^{-1})\right]$). This is surprising as will be apparent later in this paper since this value has a strong impact on the resulting MSD value (see Sec. III B). If the GA is used, this value is unique since a linear fit has a global minimum and therefore has only one solution, but in the case of more complex models, a change in the value at EI(Q = 0 Å^{-1}) results in a significant variation in MSD. Furthermore, due to the limited experimental information at Q

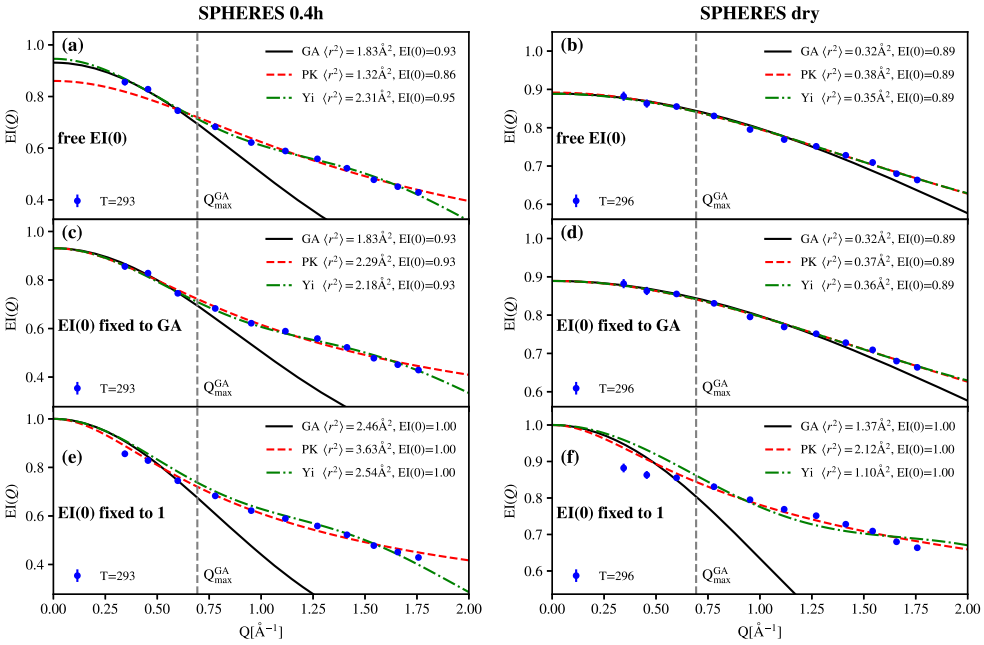


FIG. 1. Example of the effect of EI(Q = 0 Å^{-1}) for the SPHERES instrument. The left graphs [(a), (c), and (e)] show an example for A-L with 0.4h at 293 K, and the right graphs [(b), (d), and (f)] show an example for dry A-L at 296 K. In graphs (a) and (b), EI(Q = 0 Å^{-1}) is a free fit parameter. In graphs (c) and (d), EI(Q = 0 Å^{-1}) is fixed to the value evaluated with the GA. In graphs (e) and (f), EI(Q = 0 Å^{-1}) is fixed to 1 for all models.

values close to zero given by instrument geometry constraints, and the statistical error of experimental data, it is often not possible to get a unique solution for EI($Q = 0$ Å^{-1}). Since it is also not possible to fix it to 1 in all cases, we fixed this value, after several trials (see Sec. III A), to be the same for all models and to that obtained by the GA.

To illustrate the importance of the axis intercept EI($Q = 0$ Å^{-1}), two representative examples are shown in the following. Both datasets are measured on SPHERES at around 295 K. The first one is A-L at 0.4h [Figs. 1(a), 1(c), and 1(e)], and the second one is A-L dry [Figs. 1(b), 1(d), and 1(f)]. In order to qualify the differences between two fits on the same dataset, they are compared in terms of the least-squares error, evaluated by χ^2_{red} [see Eq. (21)]. The evaluated MSPF/MSD and χ^2_{red} for the different cases are shown in Table III.

Figure 1(a) shows a visual under-evaluation of the EISF EI($Q = 0$ Å^{-1}) for the PK model in comparison to the GA if one does not constrain the fit. The gray dashed line indicates the maximum Q-value used for the fit of the GA. The PK and Yi models consider all available Q-values from 0.3 to 1.8 Å^{-1}. The smallest Q-values, 0.3 and 0.46 Å^{-1}, are not well described by the fit of the PK model because EI($Q = 0$ Å^{-1}) = 0.86 ± 0.03 is much smaller than the GA EI($Q = 0$ Å^{-1}) = 0.93 ± 0.02 which fits these points well. By contrast, the Yi model has a slightly higher value of EI($Q = 0$ Å^{-1}) = 0.95 ± 0.01 than the GA, leading to a difference of almost a factor of two in the MSPF/MSD between the PK and Yi model. If the EI($Q = 0$ Å^{-1}) of these two models is fixed to the value obtained by the GA [see Fig. 1(b)], the lowest two Q-values are now well described by both models, as are the higher Q-values. The differences in the resulting MSD and χ^2_{red} for the fixed and free case are shown in Table III. For the PK model, the $\chi^2_{\text{red, fixed}} = 12$ for the fit where the intercept was fixed is ≈20% larger than that of the free fit result $\chi^2_{\text{PK, free}} = 10$. For the Yi model, the $\chi^2_{\text{red, fixed}} = 5.1$ for the fixed fit is also ≈20% larger than for the free fit $\chi^2_{\text{red, free}} = 4.0$. In addition, there are big differences in χ^2_{red} between the PK model and the Yi model even if the fits are visually quite similar for the fixed case [see Fig. 1(b)]. The reason for these is the very small error bars of the counting statistics by which the χ^2 is weighted [see Eq. (21)]. The main differences are visible in the Q-range 1.0–1.8 Å^{-1} where the Yi model follows the data better [see Fig. 1(b)]. However, more important is the effect on the evaluated MSPF of the PK model. In the free case, it is 1.3 ± 0.3 Å^2, while in the fixed case, it is almost a factor of 2 larger with 2.3 ± 0.2 Å^2. The changes of the MSD in the Yi model are not as large (2.3 vs 2.2 Å^2) since the value EI($Q = 0$ Å^{-1}) only changed by a small amount. This example is shown to illustrate that even though the least squares chi statistical value may be better for a free fit, the small Q-values can be under-evaluated for two reasons: (1) The statistical error is smaller for the higher Q-values leading to a higher weight on them and (2) the larger Q-values can be better described by the models with a lower value at EI($Q = 0$ Å^{-1}). This under-evaluation of the first Q-values is not only a problem for the PK model. It is not shown here, but it happens for all models describing higher Q-ranges.

Figure 1(c) shows what happens if the EI($Q = 0$ Å^{-1}) is fixed to 1. Visually, the GA does not describe the first data point well, the PK model is describing the range better, and the Yi model is worse. The resulting MSD/MSPF values are much larger than in the two cases before (see Table III). To emphasis that EI($Q = 0$ Å^{-1}) = 1 is not only problematic for hydrated samples, an example for the A-L dry sample at the same temperature is shown in Fig. 1(f). Here the fits show that no model is able to describe the data in the low Q-range. By contrast, they perform well if a free and a fixed value of EI($Q = 0$ Å^{-1}) to the GA is chosen [see Figs. 1(b) and 1(d)]. It also shows that in this case, fixing the offset is unnecessary since all models evaluate to the same value and that the GA is also closer to the larger Q-values than for the case of EI($Q = 0$ Å^{-1}) = 1.

The two examples here show the following: (1) In general, a higher EI($Q = 0$ Å^{-1}) leads to a higher MSD and (2) the differences between the models can be large if they are allowed to have different values of EI($Q = 0$ Å^{-1}) and should be kept the same for all models. (3) Fixing EI($Q = 0$ Å^{-1}) = 1 is not always possible; therefore when comparing models, the value EI($Q = 0$ Å^{-1}) should be fixed to the same value, which can be the value evaluated by fitting the GA in the low Q-regime where $\langle r^2 \rangle Q^2 \leq 1$. It may be the case that many experimenters use this method to fix the value in their publications, specifically where the authors state that they normalized to $Q = 0$, but many do not explain how they achieved that without the knowledge of EI($Q = 0$ Å^{-1}).

B. Differences between considered Q-ranges within the GA and the influence of statistics

A second consideration when fitting models to EINS data to extract MSD is the definition of the Q-range to be fitted.

TABLE III. Values for χ^2_{red} and MSPF/MSD $\langle r^2 \rangle$ for the PK and Yi model with the value of EI($Q = 0$ Å^{-1}) as a free parameter, fixed to the value obtained by the GA and 1.

	Example	A-L 0.4h 294 K			A-L dry 296 K		
Model	EI($Q = 0$ Å^{-1})	Free	Fixed to GA	Fixed to 1	Free	Fixed to GA	Fixed to 1
PK	$\langle r^2 \rangle$[Å^2]	1.3 ± 0.3	2.3 ± 0.2	3.6 ± 0.5	0.38 ± 0.05	0.37 ± 0.02	2.1 ± 0.4
	χ^2_{red}	10	12	17	1.1	0.98	9.7
Yi	$\langle r^2 \rangle$[Å^2]	2.3 ± 0.1	2.2 ± 0.1	2.6 ± 0.2	0.35 ± 0.03	0.36 ± 0.02	1.1 ± 0.2
	χ^2_{red}	4.0	5.1	28	1.2	1.06	29

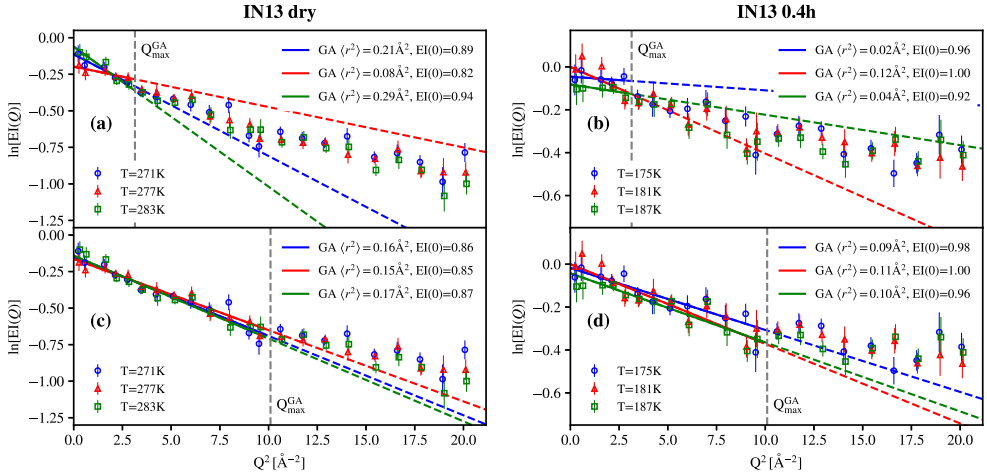


FIG. 2. Effect of statistics on the value of the EI($Q = 0$ Å^{-1}) and the MSD evaluated with the GA model. For the dry A-L, an example is shown between 271 K and 283 K taken on IN13: (a) If we consider only the low Q-range (gray line), the results differ from 0.08 to 0.29 Å^2 which explains the drop of the MSD at 277 K shown in Fig. 5(a). (c) Considering Q-values within the validity of the GA model up to $Q^2_{max} = 10$ Å^{-2} gives similar results for the MSD since the low Q-values are less weighted. [(b) and (d)] Similar example for the 0.4h A-L between 175 K and 187 K: (b) low Q-range for GA (d) $Q^2_{max} = 10$ Å^{-2} (gray line).

Typically this is vague and deserves further consideration. Many authors in publications cite the criterion defining the upper value Q_{max} and some discuss arguments why it is reasonable to surpass this limit.[46] This becomes important because even the value of EI($Q = 0$ Å^{-1}) for the GA, which will be taken in the following as the starting value for all models, depends on the chosen Q-range and the statistical error of each measured intensity data point. This can be best shown through data collected on the backscattering spectrometer IN13 at the ILL. In order to compare accurately the differences of the MSD of similar proteins, usually the same Q-ranges are chosen for the fit over all temperatures.

Figure 2 illustrates two problems if a limited Q-range is chosen. First Figs. 2(a) and 2(b) show fits to the GA which include data points up to $Q^2_{max} = 2.8$ Å^{-2}, evaluated as the largest Q-range for the A-L sample from the highest temperature (in total, 5 distinct Q-values with a statistical error of around 5%–10%). This region is similar to the Q-range of the other two spectrometers SPHERES and OSIRIS. By fitting only the very first few points, where the GA is strictly valid, large differences result in the MSD even over a temperature range of just 10 K; see Figs. 2(a) and 2(b) for the dry and 0.4h A-L, respectively. Within the statistics, the intensities EI(Q) at a given momentum transfer are almost the same for all temperatures, but due to the variation in absolute height of the first 5 Q-values, the resulting MSDs can vary by a factor of 3 (from 0.08 to 0.29 Å^2) [see Fig. 2(a)] or even 5 [from 0.02 to 0.12 Å^2] [see Fig. 2(b)]. On the contrary, if we extend the Q-range and assume the validity of the GA until $Q^2_{max} = 10$ Å^{-2} at all temperatures, a region that still reasonably corresponds to a linear fit region and is just slightly larger than the theoretically proposed limit $\langle r^2 \rangle_{GA} Q^2_{max} \leq 1$, then the resulting MSD is almost the same for all models within error bars: Fig. 2(c), 0.15–0.17 Å^2, and Fig. 2(d), 0.09–0.11 Å^2. A summary of the obtained $\langle r^2 \rangle$ and χ_{red} is shown in Table IV.

This example shows two important aspects of using the largest available Q-range: (1) A larger Q-range results in more precise and consistent results since we can include more data points and (2) it also leads, in general, to different EI($Q = 0$ Å^{-1}) and MSD values, even when both should give the same result since they are still in the limit of the GA. Therefore, it is important to include the highest possible Q-range to be as precise as possible, but also to stick with the same Q ranges to compare data accurately.

TABLE IV. Values for χ^2_{red} and MSD $\langle r^2 \rangle$ for the GA model with $Q^2_{max} = 2.8$ Å^{-2} and $Q^2_{max} = 10$ Å^{-2}.

	Example	A-L dry			A-L 0.4h		
Q^2_{max}[Å^{-2}]	T (K)	271	277	283	175	181	187
2.8	$\langle r^2 \rangle$[Å^2]	0.21 ± 0.04	0.08 ± 0.03	0.29 ± 0.06	0.02 ± 0.04	0.12 ± 0.04	0.04 ± 0.04
	χ^2_{red}	0.38	0.21	0.92	0.30	1.3	0.24
10	$\langle r^2 \rangle$[Å^2]	0.16 ± 0.02	0.15 ± 0.01	0.17 ± 0.02	0.09 ± 0.02	0.11 ± 0.01	0.10 ± 0.02
	χ^2_{red}	0.98	0.48	1.1	0.75	0.96	0.75

C. Two regimes—high Q-range

As already mentioned, the GA model is strictly valid for $Q \to 0 Å^{-1}$, but some experimental and instrumental issues arise specifically at low Q values. We can have multiple scattering[47] or non-negligible coherent contributions which result in EI(0) < 1, detectors of different resolutions for low momentum transfers as is the case for SPHERES[48] and the very similar instruments IN16B and HFBS. By contrast, IN13 and OSIRIS do not suffer from detector resolution effects. But in general, the counting statistics are also worse at lower Q-values.

To illustrate this, an example for each instrument and hydration between 305 and 310 K is shown in Fig. 3 as plots of ln [EI(Q)] vs Q^2. As can be seen, the data appear to have different linear regimes. For IN13 and SPHERES, a second linear regime at high Q is clearly visible, whereas OSIRIS shows only one. The reason is the different dynamical processes which are visible at different time scales (see Sec. II B). On SPHERES, we also note that the first two Q-values used at 0.34 and 0.45 $Å^{-1}$ are clearly higher for the hydrated samples, which is probably due to a slightly reduced resolution of these two detectors.[48] Therefore, two different low Q regimes are fitted for SPHERES, the first three detectors, including the two Q detectors with lower resolution (0.34–0.60 $Å^{-1}$) and then the next four detectors after these two detectors (0.60–1.2 $Å^{-1}$). To factor out effects at low Q and to evaluate the information that can be obtained in the second linear regime, the GA is also fitted to high Q values, even though it strictly falls out of $Q^2\langle r^2\rangle_{GA} \ll 1$. A similar approach was taken in previous publications of IN13 data (e.g., in Ref. 49 or Ref. 50). Figure 3 illustrates using dashed lines all the different linear regimes fitted in this paper (for Q-ranges, see Table II).

In addition to the second linear regime at high Q-values, the ordering between the different instruments, and therefore resolution, can be clearly observed; the break between regions of linearity is moving to smaller momentum transfers with increasing resolution.

IV. RESULTS AND DISCUSSION

In the following, we discuss the validity of the aforementioned models in more detail comparing to datasets from the three spectrometers IN13, SPHERES, and OSIRIS, in turn. In addition, we show and discuss the summed intensities versus MSPF for the three instruments and hydration levels.

A. The instrument IN13

The Q-ranges used for IN13 data evaluation are given in Table II. Representative fits are shown at three different temperatures for the A-L at 0.4h and 0.8h in Figs. 4(a) and 4(b), respectively. All models describe the data points well within their specific Q-ranges. At high temperatures for the 0.8h sample, the Yi model follows the general behavior of the experimental data but does not fit as well as the other models. For these data, a lower Q^{Yi}_{max} would be needed to obtain better agreement. However, a smaller Q-range would not include much more Q information than that already considered using the GA model. For reasons already explained since we want to compare all samples within the same Q-range, we stick to $Q^{Yi}_{max} = 2.5 Å^{-1}$ which is the best compromise between including the largest Q-values possible and a good description of the data with a given model.

The MSD results of the fits to the elastic, normalized DSF EI(Q) are shown in Fig. 5. Figure 5(a) shows the difference between the GA, PK, and Yi model for the dry protein. The differences between the models are very small, and all models show a similar behavior. The same plot is shown for 0.4 hydration in Fig. 5(b). Here, the differences between the models are also small, but overall, the PK and Yi models give

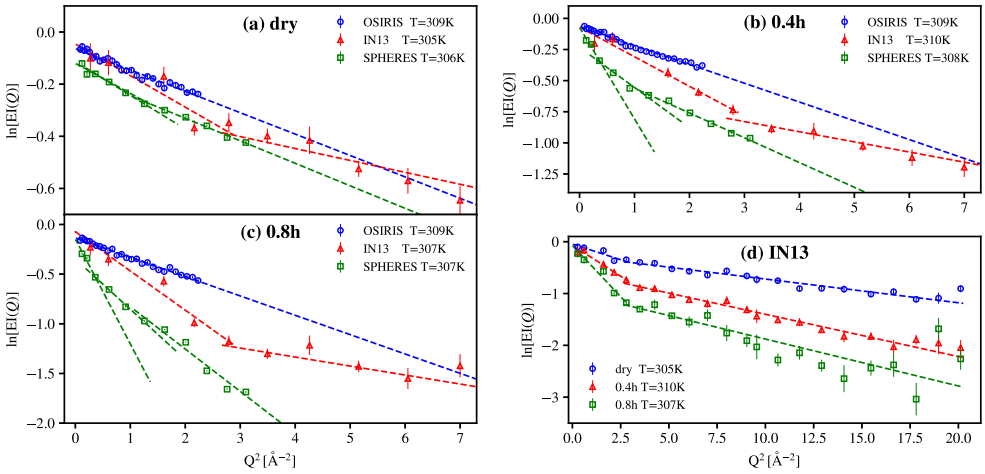


FIG. 3. Normalized raw data ln[EI(Q)] vs Q^2 for the three different hydration levels [(a)–(c)] and the three different instruments, OSIRIS (blue), IN13 (red), and SPHERES (green). The dashed lines indicate linear fits to the data in their respective range. For OSIRIS, at all hydration levels, mainly one linear regime is visible. For IN13, clearly two regimes are visible, separated around 3 $Å^{-2}$. For SPHERES, two low Q regions and one high Q region can be identified. (d) shows the entire Q-range of IN13 and the linearity of the second regime up to 20 $Å^{-2}$.

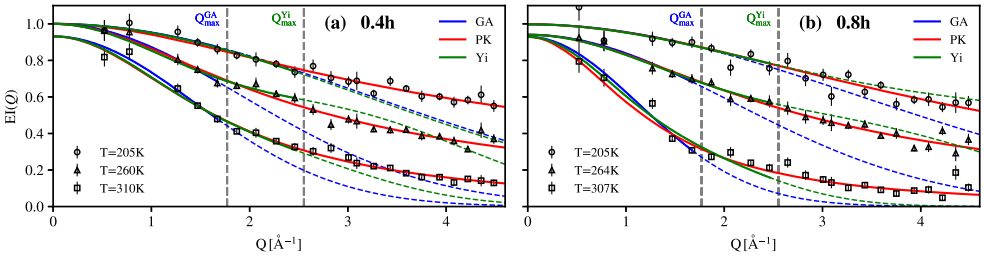


FIG. 4. Representative fits for the three models at three different temperatures for IN13 data. (a) corresponds to 0.4h and (b) corresponds to 0.8h. The vertical gray lines indicate the Q_{max} used for each model: 0.5–1.7 Å^{-1} for GA and 0.5–2.5 Å^{-1} for Yi.

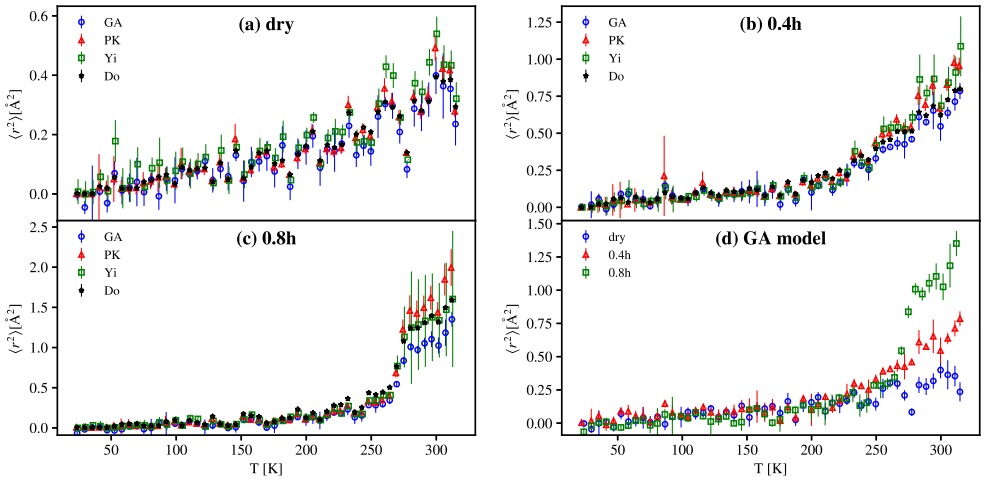


FIG. 5. MSD values extracted from the GA, PK, Yi, and Do models for dry, 0.4, and 0.8 hydrated A-L depleted samples [(a)–(c)]; IN13 data. (d) shows the MSD of all three hydration levels evaluated with the GA model.

rise to slightly higher MSPF/MSD values due to the inclusion of higher Q values. At 0.8 hydration [see Fig. 5(c)], the MSPF/MSD values are clearly higher at high temperatures for both models. Both hydrated samples show an increase in the MSD at around 230 K compared to the dry protein as expected at the dynamical transition temperature,[2] commonly observed in hydrated proteins. For the 0.8h sample, around 270–280 K, a very steep increase in the MSD is visible that can be attributed to the melting point of heavy water at 278 K. Probably for 0.8 hydration, some free water exists that can freeze and therefore inhibit the motion of the protein. Therefore, the MSD shows an abrupt rise at the melting point. After 280 K, the increase in the MSD for all models is similar to the 0.4h sample, but the absolute values are higher. The MSD for the 0.8h sample is higher than for the 0.4h sample since the higher hydration decreases the crowding in the sample and thus allows more motions. For both hydrated samples, the error bars of the GA are smaller than for the other models. This gives the illusion of a higher accuracy, which is only due to the cut-off effect and to the exclusion of certain amplitudes. In the case of the Yi model at 270–320 K, very large errors are evaluated since the fit is not describing the data as well as for the other cases [e.g., see Fig. 4(b) at 307 K]. This can be confirmed by the higher χ^2_{red} shown in Fig. 6.

Finally, we compare the performance of the models that include heterogeneity to the double well potential model of Doster *et al.* that takes into account anharmonicity. Since IN13 has by far the largest available Q-range, the Doster model is only evaluated on this instrument to show that it also gives good results. Figures 5(a)–5(c) show the results of $\langle r^2 \rangle_{Do,tot}$ defined

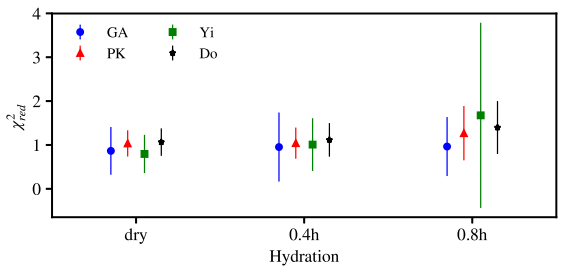


FIG. 6. Reduced χ^2_{red} for the fits, averaged over all temperatures. It is important to note that the reduced χ^2_{red} values are calculated with the respective Q-values used for the fit. Therefore, the GA and Yi models take less values into account. All four models have a χ^2_{red} value around 1, but the value for the Yi model at 0.8 hydration is around 2 and has a much bigger standard deviation. The reason for this was already mentioned in the results of the evaluated MSD. The fit of the Yi model is indeed worse at high temperatures and therefore also its χ^2_{red}.

in Eq. (15). They compare rather well with the results from the other three models. The evaluated values for the enthalpy $\Delta G = 6.1$ kJ mol^{-1} and d = 1.7 Å (0.4h) are similar to values found for myoglobin at 0.38h ($\Delta G = 12$ kJ mol^{-1}, d = 1.5 Å).[2] In Fig. 6, the reduced χ^2_{red} for the different fits averaged over the entire temperature range are shown.

Overall the results on IN13 lead to the conclusion that for the dry A-L, all models are equal since the system is still close to a harmonic system (no dynamical transition visible) and the heterogeneity does not influence the evaluated MSD. When the hydration increases, the anharmonicity grows and quantitative differences can be seen between the models. Especially at 0.8 hydration, the three models diverge at high temperatures, showing the influence of the heterogeneity to the MSD. The quality of the fits is similar for all three models in their respective Q-range. Only in the case of the Yi model, it is worse at high temperatures for the A-L at 0.8h.

1. IN13—only high Q-range

Taking advantage of the fact that IN13 covers a wide Q-range, we follow on from the discussion in Sec. III C to now consider fitting the high Q-range alone using the GA model, to evaluate the effect on the dynamical transition temperature and the MSD. Specifically, we fit the Q-range 1.7–4.5 Å^{-1}. The result is shown in Fig. 7 and can be compared to the GA at low Q-values in Fig. 5(d). The absolute value of the MSD from the high Q-range is lower by up to a factor of 5 for the 0.8h sample. This is a huge effect, but might be reasonable when separating motions of large and small amplitudes, for instance, atoms of the side chains of the amino acids and fluctuations within the backbone of the amino acids. The relative change in the MSD with increasing temperature between the different hydrations is however similar. All hydrations follow the same trend until 220 K and then deviate from each other. In the high Q-range, the 0.4h and 0.8h curves are then superposed until the second splitting due to the melting of heavy water, whereas for the small Q-range, the MSD of the higher hydrated sample lies even below the 0.4h sample.

In the paper of Combet and Zanotti,[51] the authors study a protonated protein hydrated in D_2O and the same protein in a per-deuterated form hydrated in H_2O on two different instruments, IN13 and the spectrometer MIBEMOL with a resolution of ≈140 μeV. The corresponding short time window of ≈10 ps s reveals a weak dynamical transition, observed for both the protein and its hydration water. By contrast, the larger time window of IN13 permits a separation of the experimental data into large and local motions with a crossover at around 1.2 Å^{-1}; clear differences in the motions of water and protein molecules are being visible since larger amplitude motions can be probed at low Q. Their approach of using H/D-contrast allows direct evidence of two populations of motion. Our findings without contrast variation confirm that a division in two population is indeed reasonable.

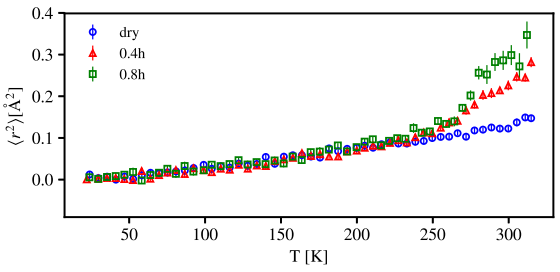


FIG. 7. MSD for all three hydrations evaluated with the GA model at a high Q range (1.7–4.5 Å^{-1}); IN13 data.

Another approach in this same line of thought involves fitting the whole Q-range using a bimodal fitting model, where $\text{EI}(Q) \approx p_1 \exp(-\langle r^2\rangle_{\text{large}} Q^2) + p_2 \exp(-\langle r^2\rangle_{\text{local}} Q^2)$, with $p_1 + p_2 = 1$. It assumes the presence of two well-separated MSDs defined within the GA $\langle r^2\rangle_{\text{large}} > \langle r^2\rangle_{\text{local}}$ in the sample, as proposed by Nakagawa *et al.*[24] This approach works well for the protein staphylococcal nuclease with an instrumental resolution of 1 m eV, obtaining $\langle r^2\rangle_{\text{large}} \approx 0.7$ Å^2 and $\langle r^2\rangle_{\text{local}} \approx 0.15$ Å^2 at 300 K.

In conclusion, our results suggest that by analyzing the high Q range only on an instrument like IN13, one can access local motions in the proteins, which give rise to small amplitudes of motion, but which follow quite closely the temperature and hydration behavior of the large amplitudes in the case of hydrated powders.

B. The instrument SPHERES

For the SPHERES spectrometer, the used Q-range is 0.34–1.8 Å^{-1}. The first two detectors at $Q = 0.34$ and $Q = 0.45$ Å^{-1} have a lower energy resolution.[48] To see the influence of this resolution effect, two different low Q-ranges were fitted in which the GA is still valid, i.e., ln[EI(Q)] vs Q^2 linear for all temperatures. The first low Q-range only includes the first three available Q-values 0.34–0.6 Å^{-1} (see Sec. IV B 1), whereas the second low Q-range excludes the first two Q-values with lower resolution, corresponding to a Q-range of 0.6–1.2 Å^{-1} (see Sec. IV B 2). In both cases, the PK and Yi models are fitted to the same Q-values as the respective fit of the GA and using in addition the larger available Q-values until 1.8 Å^{-1}. A third Q region is fitted to only high Q-values, as described in Sec. III C. There only the GA was used in the Q-range 0.95–1.8 Å^{-1}, neglecting the first four Q-values.

1. Low Q-range (GA, 0.34–0.6 Å^{-1})

In Figs. 8(a) and 8(b), representative fits for three different temperatures are shown for the 0.4h and 0.8h A-L samples. The fits show that it is not possible to include more Q-values for the GA if the small Q-values should still be well described by the fit. A comparison to a larger Q-range which does not describe the lowest two Q-values is shown in Sec. IV B 2. The SPHERES spectrometer has ten times better resolution than IN13 such that larger motions are included and the resulting MSD becomes larger, leading to a lower Q-range accessible by the GA. These larger motions are possibly the movements of the side chains of A-L. Also small differences are visible between the samples in comparison to IN13.

The results of the fits of the EI(Q) are shown in Fig. 9 in the same way as for IN13. For all samples, a linear increase of the MSD/MSPF is visible until 200 K where the dynamical transition sets in.

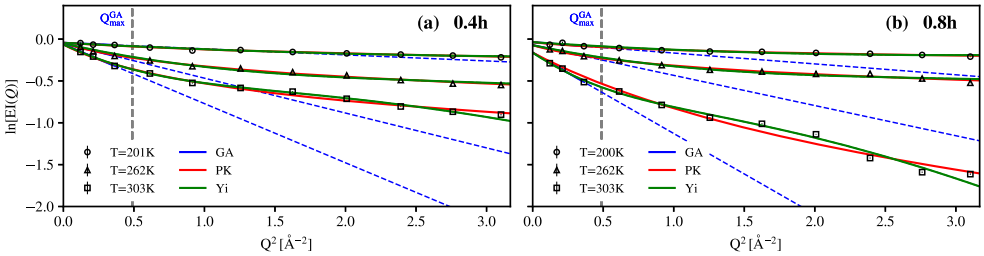


FIG. 8. Representative fits for the three models at three different temperatures for SPHERES data. (a) corresponds to 0.4h, and (b) corresponds to 0.8h. The vertical gray line indicates Q_{max} = 0.6 Å^{-1} used for the GA.

For the dry sample, the evaluated MSD is very noisy which can be explained by the low statistics implied in using only the first three Q-values. Higher Q-values could be included, but in order to compare the same Q-ranges between the different hydrations, we stick to the Q-range evaluated for the highest hydration (for a larger Q-range, see Sec. IV B 2). The problem with using the first three data points only is emphasized by the small decay of the EI(Q) for the dry sample. The Yi and PK models fit the data well and with a similar goodness, resulting in similar values of MSD/MSPF, and with a much better accuracy than the GA since they are including all experimental data points.

At 0.4 hydration, the dynamical transition is visible around 200 K. The MSPFs/MSDs of the PK and Yi model are higher than the MSDs of the GA model following the trends seen on IN13. At 0.8 hydration, two changes in the slope are visible at ~200 K and ~270 K. The first change of the slope is again attributed to the dynamical transition and the second is attributed to the melting of free heavy water which enhances the movements of the protein. The highest MSD/MSPF values are around 3.5 Å^2. The PK model has the same MSPF as the MSD of the GA, whereas the Yi model has higher MSD at temperatures above 270 K. This coincides again with the melting point of D_2O. The MSD from 270 to 280 K increases in the Yi model by almost 1.5 Å^2. This transition is smoother for the two other models. On the other hand, after the jump, the increase of the MSD in the Yi model is much slower in comparison to the other models. Therefore, at 310 K, the 3 models reach a similar value in MSD/MSPF. The jump of the Yi model can be explained by the counter-balancing between the Q^2 and Q^4 term [see Eq (11)], which results likely into a mathematical, but not a physical solution. Also the reduced χ^2_{red} statistics indicate that the fitting of the PK and Yi model for 0.8 hydrated samples is much worse than for the dry and 0.4 hydration samples (data not shown).

2. Low Q-range (GA, 0.6–1.2 Å^{-1})

In Sec. IV B 1, the GA was only fitted to the first three Q-values (0.35–0.6 Å^{-1}) to take into account the limit of validity given in Eq. (10). This leads to a good description of the data in the low Q-range, but also to a high statistical error. In addition, as stated before, the resolution is larger for the first two low Q detectors in comparison to the other detectors. If the GA is fitted in the range of 0.6–1.2 Å^{-1}, it leads to a more consistent linear fit with a smaller error. Examples for the fits are shown in Figs. 10(a) and 10(b) as before. The resulting MSDs are shown in Fig. 11, together with the Yi and PK models using the Q-range 0.34–1.8 Å^{-1}.

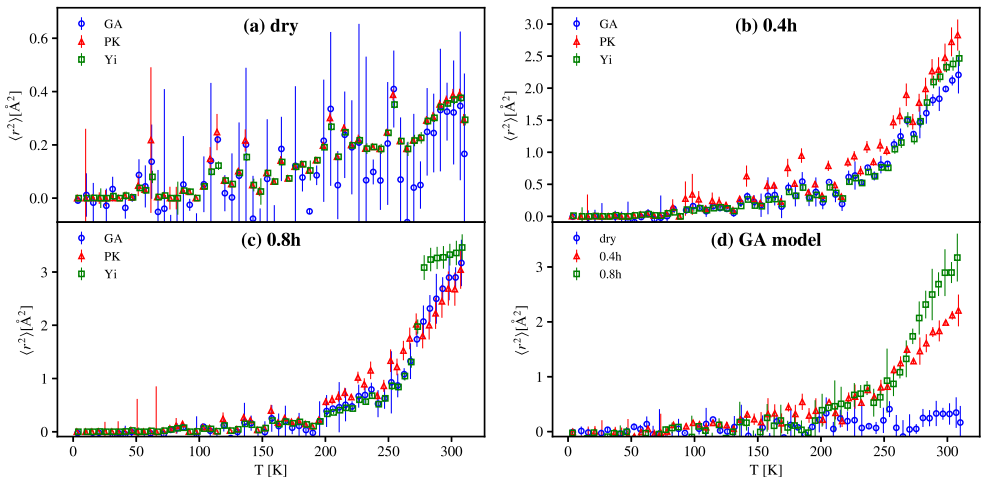


FIG. 9. MSD values extracted from the GA, PK, and Yi models for dry, 0.4 and 0.8 hydrated A-L depleted samples [(a)–(c)]; SPHERES data. (d) shows the MSD of all three hydration levels evaluated with the GA model. For the GA, only the first three Q-values (0.34–0.6 Å^{-1}) were used as shown in Fig. 8.

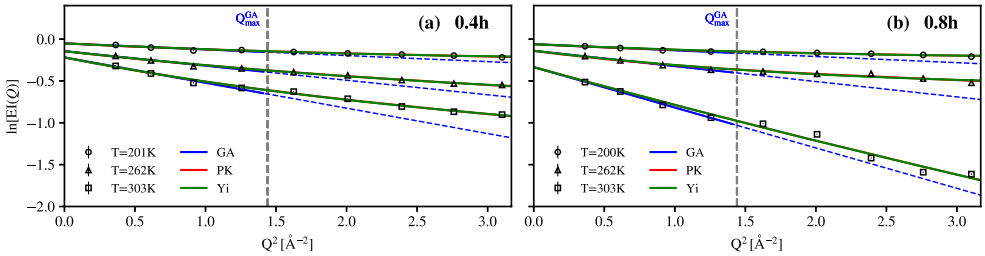


FIG. 10. Representative fits for the three models at three different temperatures for SPHERES data. (a) corresponds to 0.4h, and (b) corresponds to 0.8h. The vertical gray line indicates Q_{max} = 1.2 Å^{-1} used for the GA. Here, the lowest two Q-values are neglected in contrast to Fig. 8.

Figures 11(a)–11(c) summarize the three different hydrations of A-L, as shown before. For all three hydrations, the GA, PK, and Yi models evaluate to similar MSPF/MSD values. The values of the MSD for the dry protein are very similar between the two different Q-ranges of the GA. For the 0.4h and 0.8h samples, the MSD is lower by a factor of ≈2 compared to fits including the lowest Q-value. Figure 11(d) shows the results of the GA for all three hydrations. A clear difference in the increase of the MSD is visible between the dry and the two hydrated samples at around 220 K. The two hydrated samples are then following the same pattern until around 270 K. At higher temperatures, the 0.8h sample has a much larger MSD than the 0.4h sample. As explained before, the reason is the melting of frozen D_2O. Nevertheless, it is interesting that both curves have the same MSD until 270 K which could be due to the non-frozen water shell around the protein being the same in both hydrations.

The comparison between the two different Q-ranges at low Q for the GA [see Figs. 9(a)–9(d) and 11(a)–11(d)] shows that the quantitative value of the MSD is different depending on which range is chosen. The difference of a factor of 2 in the MSD for hydrated samples originates mainly in the difference of the value of the EI(Q = 0 Å^{-1}). This is especially visible in the cases of the PK and Yi models. They both use the entire available Q-range at high Q, but fixing EI(Q = 0 Å^{-1}) to the respective value obtained by the GA changes their quantitative results dramatically.

3. High Q-range for GA: 0.96–1.8 Å^{-1}

As shown in Sec. III C, it is possible to describe only the high Q range of SPHERES using the GA. The fitted Q-range in which ln[EI(Q)] vs Q^2 is linear is 0.96–1.8 Å^{-1}. The resulting MSDs are shown in Fig. 12. Comparing to the low Q fits without the small angle detectors [Fig. 11(d)], the MSDs are smaller but only up to a factor of 2 in the case of the 0.4h sample. More interesting are the changes in the behavior between the different samples. First, it appears that the dry powder shows a slightly larger MSD in the temperature range 150–240 K compared to the hydrated samples which have the same MSD in this temperature range. Such an observation has been reported by Nickels *et al.*[52] for green fluorescent protein (GFP) [0.4 D_2O hydration vs. dry, Fig. 2(a) in their publication], which was measured on the very similar instrument, HFBS. They suggest that the frozen hydration shell reduces the MSD at low T by suppressing fast ps fluctuations. A similar behavior was also found in molecular dynamics simulations of GFP by Hong *et al.*[53] Second, the MSD of the 0.8h sample is the same as that for the dry sample between 240 and 270 K

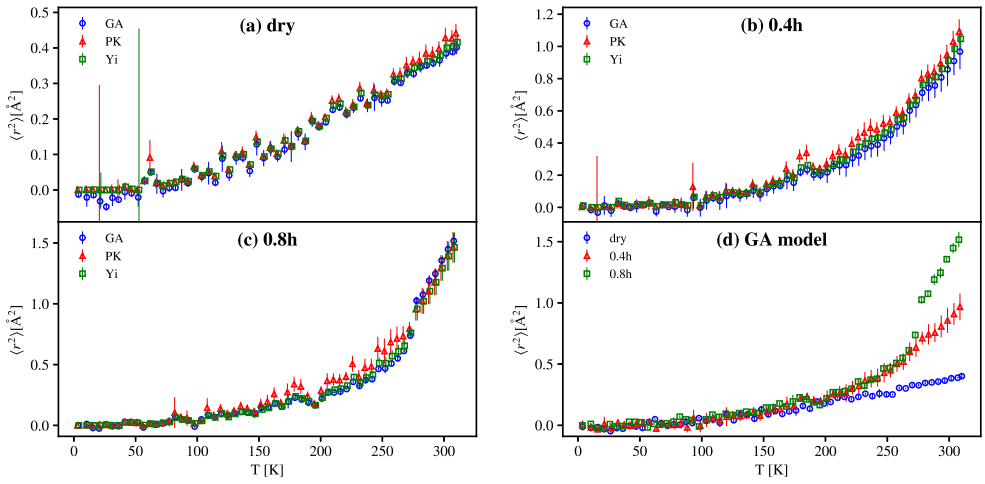


FIG. 11. MSD values extracted from the GA, PK, and Yi models for dry, 0.4, and 0.8 hydrated A-L depleted samples [(a)–(c)]; SPHERES data. (d) shows the MSDs of all three hydration levels evaluated with the GA model. For the GA, the Q-range 0.6–1.2 Å^{-1} is evaluated, and for the PK and Yi models, the Q-range is 0.6–1.8 Å^{-1}.

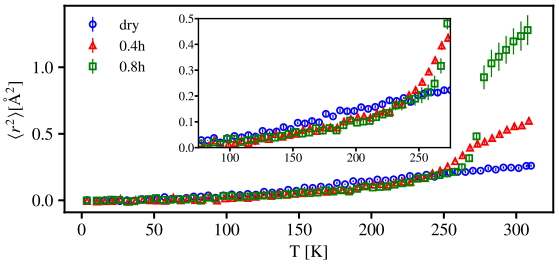


FIG. 12. MSD for all three hydrations evaluated with the GA model at a high Q range (0.96–1.8 Å^{-1}); SPHERES data. The inset shows a zoomed in view of the low T region.

and lower than for the 0.4h sample. It seems that the MSD is inhibited by the frozen D_2O for the 0.8h sample. Above 270 K, the MSD of the 0.8h sample increases steeply. This again coincides with the melting point of heavy water at 278 K. Finally, in the inset of Fig. 12, one can see that the dynamical transition still takes place around 200 K, but since the dry sample has a higher MSD at lower T, the crossing between the hydrated and dry sample takes place at 250 K.

C. The instrument OSIRIS

For the OSIRIS spectrometer, the chosen Q-range is 0.29–1.5 Å^{-1}. Note that OSIRIS has more detectors that cover this Q-range as compared to the other two spectrometers, allowing for better Q-resolution. The instrumental time resolution is three times larger than on IN13 such that faster motions up to around 25 ps can be probed. This may explain why it is possible to use all three models until 1.5 Å^{-1}. Mainly small localized movements are observed and not larger side-chain motions. The resulting MSDs are thus small in comparison to IN13 and SPHERES. Representative fits of the EI(Q) for the two hydrated samples are shown in Fig. 13.

The MSDs/MSPFs for the different models are illustrated in the same way as for IN13 and SPHERES in Fig. 14. For each hydration, the GA, PK, and Yi models yield almost the same MSD/MSPF values. At 310 K, the PK and Yi models evaluate slightly higher MSPF/MSD values for the dry and 0.4 hydration samples and for the 0.8 hydration sample above 280 K. This behavior confirms that all models give similar results if they use the same Q-range on this instrument. The dry sample shows a linear increase of the MSD/MSPF with increasing temperature. For the 0.4 and 0.8 hydration, the dynamical transition is visible and starts at around 250 K.

D. Comparison between summed intensities and MSPF

We would like to note that the MSDs/MSPFs are not always the best way to look at data when small differences are expected between similar samples. In such cases, the

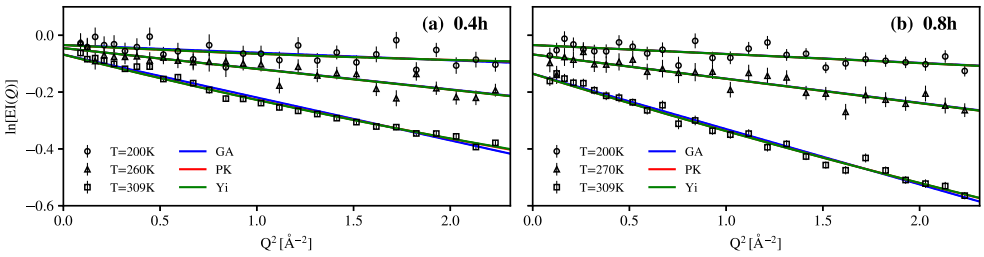


FIG. 13. Representative fits for the three models at three different temperatures for OSIRIS data. (a) corresponds to 0.4h, and (b) corresponds to 0.8h.

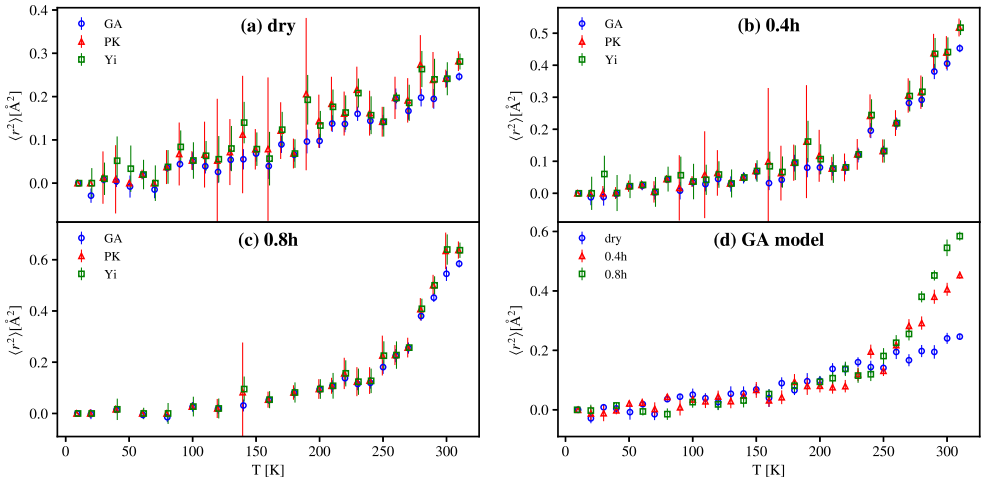


FIG. 14. MSD values extracted from the GA, PK, and Yi models for dry, 0.4, and 0.8 hydrated A-L depleted samples [(a)–(c)]; OSIRIS data. d) shows the MSD of all three hydration levels evaluated with the GA model. All models use the same Q-values (0.29–1.5 Å^{-1}).

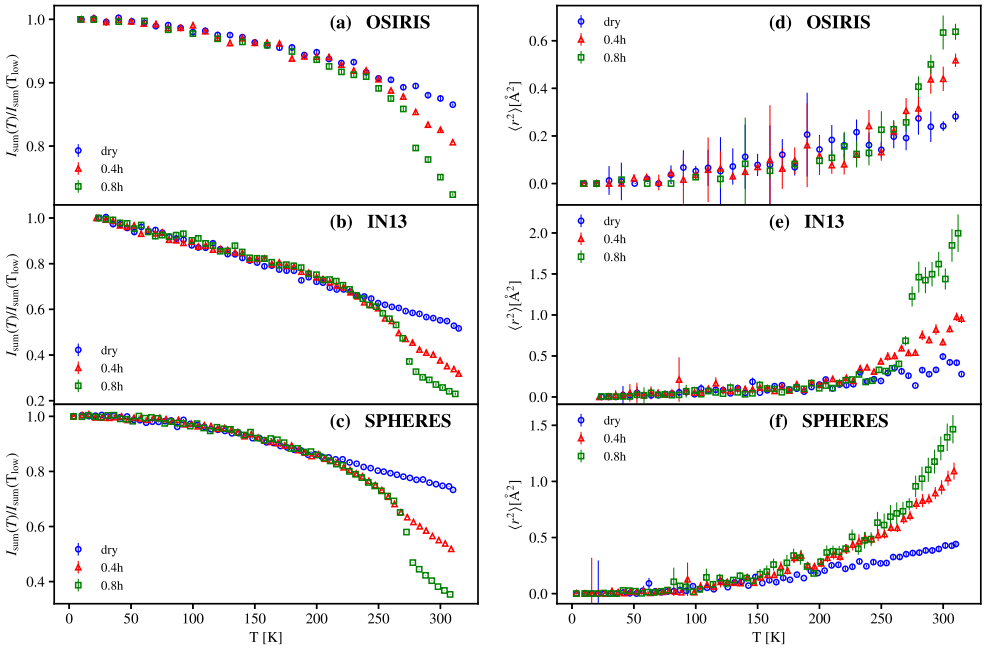


FIG. 15. [(a)–(c)] Summed intensities over the entire available Q-range of each instrument. [(d)–(f)] Respective evaluated MSPF within the PK model over the same Q-range.

comparison of the intensities summed over all (or a range of) available scattering angles, I_{sum}, can be much more insightful. In the limit of the GA, they can moreover be related to the MSD [see Eq. (19)]. We show here the I_{sum} for the three different instruments and hydration levels in Fig. 15 and compare them to the MSPF evaluated by the PK model using the same Q-range. The results demonstrate the inverse hierarchy for the I_{sum} and the MSPF between the samples. As the error bars of the summed intensities are much smaller, it allows us to better separate the curves. The differences between the hydration levels become more visible in the summed intensities. For instance, the kink for the 0.8h sample at ≈275 K clearly indicates the melting of the ice. Moreover, one distinguishes on IN13 that in the case of I_{sum} [see Fig. 15(b)], the curve corresponding to the dry sample lies below the curves of the hydrated samples in the temperature domain from 150 to 210 K. We discussed this effect earlier in Sec. IV B 3. This behavior cannot be observed in the MSPF [see Fig. 15(e)] where this information is lost due to data fitting. However, for the MSPF evaluated on SPHERES, a small bump at 180 K can be seen for the hydrated A-L samples [see Fig. 15(f)], which is not visible in the I_{sum} [see Fig. 15(c)] and which shows up also when applying the GA [see Fig. 11(d)]. It might be a real effect, eventually due to an ice phase transition,[54] which appears only through the evaluation of the slopes. Both data analyses might therefore furnish complementary information.

V. SUMMARY AND CONCLUSION

We have evaluated a number of models and the resulting MSD/MSPF of A-L for different instruments and hydration levels, and we used various Q-ranges for data analysis. We show that the intercept of EI(Q = 0 Å^{-1}) is very important in fitting the data to any given model. Fixing it to the theoretical value of EI(Q = 0 Å^{-1}) = 1 is often not possible for the GA, as shown in Sec. III A, nor for the models that take into account a larger Q-range. Since all models should converge in the limit of Q → 0 to the same value, but can present significant variations when leaving it as a free parameter, we advise to fix EI(Q = 0 Å^{-1}) to a common value for the same dataset. In our case, this value was obtained by the fit of the GA. With such settings, the qualitative results between the models are similar, concerning the dynamical transition and dynamical changes as a function of the hydration level, which leads to the conclusion that the most basic and mainly used model, the GA, is indeed a good standard, if a consistent protocol is followed, i.e., respecting a linear behavior of ln[EI(Q)] vs Q^2 and comparing the same Q-ranges.

The inclusion of the smallest Q-values in traditional reactor-based neutron backscattering spectrometers can change the quantitative MSD dramatically as shown on SPHERES, i.e., resulting in a difference by a factor of two for the MSD. As stated before, the first detectors on SPHERES have a lower resolution than the other detectors, but this is unlikely the only reason for the large differences in MSD. As the temporal range of SPHERES allows us to probe larger amplitudes of motion, it could also arise from the onset of movements of molecular subgroups. Furthermore, if the same Q-range is respected when comparing hydration levels or different samples, the GA is an adequate model and the inclusion of non-Gaussian terms is not needed in order to have a good estimation on the dynamics. By contrast, we show that for a quantitatively accurate analysis, the MSD can be quite

different between models, especially at a higher hydration level. In fact, even the PK and Yi models do not provide much higher precision, mainly because they are too dependent on the value of EI(Q = 0 Å^{-1}). If the experimenter is expecting small differences in dynamics between samples, then the extended models could give more accurate values for the MSD since they include information over a larger Q-range. However, the data will be subject to larger errors than the GA since they also have more free parameters to fit, as can be seen from the MSD/MSPF on OSIRIS (see Sec. IV C). The comparison of the summed intensities I_{sum} may be more helpful to determine small differences, and they can deviate from the behavior of the MSPF/MSD even when both use the same raw data and Q-ranges (see Sec. IV D).

As EINS is often applied to a class of samples, which have only minor differences (enzymes in the presence or not of an inhibitor,[55] a wild-type biomolecule against a mutant,[56] proteins at different concentrations,[57] with various co-solutes,[58] etc.), the comparison of such a dataset requires highest consistency concerning:

- use of the same hydration level,
- use of the same instrumental resolutions,
- use of the same Q-values for extraction of MSD/MSPF or summed intensities,
- set EI(Q = 0 Å^{-1}) to the same value if comparing different models for the same dataset.

In conclusion, our investigation evidences that, despite many efforts to improve the quantitative results for the MSD, significant questions remain and we were not able to establish a reliable method on how to treat the data to get results with highest precision. However, we were able to show that a prevailing although often ignored point is the correct treatment of the point at EI(Q = 0 Å^{-1}). Interestingly, our study suggests that in some cases, a bi-modal approach might be sufficient and helpful to distinguish smaller and larger motions as already suggested by, e.g., Nakagawa *et al.*,[24] Doster and Settles,[59] or Combet and Zanotti[51] in the past. On the one hand, such treatment could help to distinguish large motions, which are sometimes associated with movements of hydration water if H_2O is used or with local translational diffusive displacements, whereas smaller motions account for localized dynamics or vibrations within the biomolecules. Doster very recently also showed that a bi-modal treatment of data taken on myoglobin clearly permits to identify at least two molecular processes which might be sufficient to describe the neutron scattering spectra of proteins.[29] On the other hand, as discussed in the present work, if a clear separation of linear regimes is no longer appropriate, the models using a continuous distribution of individual MSD describe the dynamics very well without further assumptions. For a reliable conclusion, the exact knowledge of EI(Q = 0 Å^{-1}) is of utmost importance, but most instruments do not permit a direct measurement of this value. We suggest further studies by either spin-echo spectroscopy giving access to Q-values as small as 10^{-2} Å^{-1} (IN15[60] at the ILL) or polarisation analysis in combination with neutron techniques (diffraction and/or spectroscopy[61,62]) to separate the coherent and incoherent scattering signals. When addressing motions at low Q's, it is important to note that multiple scattering becomes increasingly important and has to be corrected carefully.[29] This approach would also help us to determine the exact reasons for a deviation from EI(Q = 0 Å^{-1}) = 1 and how it can be estimated. In addition, it may enable more quantitative precision in determining the values of MSD/MSPF in the future.

ACKNOWLEDGMENTS

D.Z. was supported by a Ph.D. scholarship co-funded by the Communauté Université Grenoble Alpes, the STFC Rutherford Appleton Laboratory, and the Institut Laue Langevin. J.P. and V.G.S. gratefully acknowledge the support by M. Johnson to partly acquire the Ph.D. grant and fruitful discussions. The authors gratefully acknowledge the financial support provided by JCNS to perform the neutron scattering measurements at the Heinz Maier-Leibnitz Zentrum (MLZ), Garching, Germany. Experiments at the ISIS Neutron and Muon Source were supported by a beamtime allocation from the Science and Technology Facilities Council. We thank the ILL for beamtime allocation.

[1] B. P. Schoenborn, Nature **224**, 143 (1969).
[2] W. Doster, S. Cusack, and W. Petry, Nature **337**, 754 (1989).
[3] L. Rusevich, V. García Sakai, B. Franzetti, M. Johnson, F. Natali, E. Pellegrini, J. Peters, J. Pieper, M. Weik, and G. Zaccai, Eur. Phys. J. E **36**, 80 (2013).
[4] M. Bée, *Quasielastic Neutron Scattering: Principles and Applications in Solid State Chemistry, Biology and Materials Science* (Adam Hilger, Philadelphia, 1988).
[5] L. Van Hove, Phys. Rev. **95**, 249 (1954).
[6] H. Frauenfelder, R. D. Young, and P. W. Fenimore, Proc. Natl. Acad. Sci. U. S. A. **114**, 5130 (2017), http://www.pnas.org/content/114/20/5130.full.pdf.
[7] G. R. Kneller, Proc. Natl. Acad. Sci. U. S. A. **115**, 9450 (2018), http://www.pnas.org/content/early/2018/08/29/1718720115.full.pdf.
[8] W. Doster, H. Nakagawa, and M. Appavou, J. Chem. Phys. **139**, 045105 (2013).
[9] F. Gabel, B. Bicout, U. Lehnert, M. Tehei, M. Weik, and G. Zaccai, Q. Rev. Biophys. **35**, 327 (2002).
[10] G. Zaccai, J. Non-Cryst. Solids **357**, 615 (2011), 6th International Discussion Meeting on Relaxation in Complex Systems.
[11] G. Zaccai, Science **288**, 1604 (2000).
[12] M. Trapp, J. Marion, M. Tehei, B. Deme, T. Gutberlet, and J. Peters, Phys. Chem. Chem. Phys. **15**, 20951 (2013).
[13] J. Roh, V. Novikov, R. Gregory, J. Curtis, Z. Chowdhuri, and A. Sokolov, Phys. Rev. Lett. **95**, 038101 (2005).
[14] G. Squires, *Introduction to the Theory of Thermal Neutron Scatering* (Dover Publications, 1978).
[15] M. T. Telling, S. Howells, J. Combet, L. A. Clifton, and V. García Sakai, Chem. Phys. **424**, 32 (2013), Neutron Scattering Highlights on Water and Biological Systems.
[16] W. Doster, M. Diehl, W. Petry, and M. Ferrand, Physica B **301**, 65 (2001).
[17] A. Rahman, K. S. Singwi, and A. Sjolander, Phys. Rev. **126**, 986 (1962).
[18] G. R. Kneller and G. Chevrot, J. Chem. Phys. **137**, 225101 (2012).
[19] A. Tokuhisa, Y. Joti, H. Nakagawa, A. Kitao, and M. Kataoka, Phys. Rev. E **75**, 041912 (2007).
[20] T. Becker and J. C. Smith, Phys. Rev. E **67**, 021904 (2003).
[21] Z. Yi, Y. Miao, J. Baudry, N. Jain, and J. C. Smith, J. Phys. Chem. B **116**, 5028 (2012).
[22] A. Rahman, Phys. Rev. **136**, A405 (1964).
[23] K. Sköld, J. M. Rowe, G. Ostrowski, and P. D. Randolph, Phys. Rev. A **6**, 1107 (1972).
[24] H. Nakagawa, H. Kamikubo, I. Tsukushi, T. Kanaya, and M. Kataoka, J. Phys. Soc. Jpn. **73**, 491 (2004).
[25] L. Meinhold, D. Clement, M. Tehei, R. Daniel, J. L. Finney, and J. C. Smith, Biophys. J. **94**, 4812 (2008).
[26] J. Peters and G. R. Kneller, J. Chem. Phys. **139**, 165102 (2013).
[27] G. R. Kneller and K. Hinsen, J. Chem. Phys. **131**, 045104 (2009).

[28]D. Vural, J. C. Smith, and H. R. Glyde, Biophys. J. **114**, 2397 (2018).
[29]W. Doster, Int. J. Mol. Theor. Phys. **2**, 1 (2018).
[30]D. Vural, L. Hong, J. C. Smith, and H. R. Glyde, Phys. Rev. E **91**, 052705 (2015).
[31]S. Perticaroli, G. Ehlers, C. B. Stanley, E. Mamontov, H. O'Neill, Q. Zhang, X. Cheng, D. A. Myles, J. Katsaras, and J. D. Nickels, J. Am. Chem. Soc. **139**, 1098 (2017).
[32]V. F. Sears, Neutron News **3**, 26 (1992).
[33]F. Natali, J. Peters, D. Russo, S. Barbieri, C. Chiapponi, A. Cupane, A. Deriu, M. T. Di Bari, E. Farhi, Y. Gerelli, P. Mariani, A. Paciaroni, C. Rivasseau, G. Schiró, and F. Sonvico, Neutron News **19**, 14 (2008).
[34]D. Zeller and J. Peters, Internal time on IN13 (dataset), Inst. Laue-Langevin (ILL) (2016).
[35]J. Wuttke, A. Budwig, M. Drochner, H. Kammerling, F. J. Kayser, H. Kleines, V. Ossovyi, L. C. Pardo, M. Prager, D. Richter, G. J. Schneider, H. Schneider, and S. Staringer, Rev. Sci. Instrum. **83**, 075109 (2012).
[36]Heinz Maier-Leibnitz Zentrum, J. Large-Scale Res. Facil. **1**, A30 (2015).
[37]M. T. F. Telling and K. H. Andersen, Phys. Chem. Chem. Phys. **7**, 1255 (2005).
[38]D. Richard, M. Ferrand, and G. J. Kearley, J. Neutron Res. **4**, 33 (1996).
[39]O. Arnold, J. Bilheux, J. Borreguero, A. Buts, S. Campbell, L. Chapon, M. Doucet, N. Draper, R. F. Leal, M. Gigg, V. Lynch, A. Markvardsen, D. Mikkelson, R. Mikkelson, R. Miller, K. Palmen, P. Parker, G. Passos, T. Perring, P. Peterson, S. Ren, M. Reuter, A. Savici, J. Taylor, R. Taylor, R. Tolchenov, W. Zhou, and J. Zikovsky, Nucl. Instrum. Methods Phys. Res., Sect. A **764**, 156 (2014).
[40]M. Jasnin, L. van Eijck, M. M. Koza, J. Peters, C. Laguri, H. Lortat-Jacob, and G. Zaccai, Phys. Chem. Chem. Phys. **12**, 3360 (2010).
[41]M. Trapp, M. Trovaslet, F. Nachon, M. M. Koza, L. van Eijck, F. Hill, M. Weik, P. Masson, M. Tehei, and J. Peters, J. Phys. Chem. B **116**, 14744 (2012).
[42]V. Réat, G. Zaccai, C. Ferrand, and C. Pfister, *Biological macromolecular dynamcis, Proceedings of a Workshop on Inelastic and Quasielastic Neutron Scattering in Biology* (Adenine Press, 1996), pp. 117–122.
[43]Z. Liu, J. Huang, M. Tyagi, H. O'Neill, Q. Zhang, E. Mamontov, N. Jain, Y. Wang, J. Zhang, J. C. Smith, and L. Hong, Phys. Rev. Lett. **119**, 048101 (2017).
[44]*DLMF: NIST Digital Library of Mathematical Functions*, http://dlmf.nist.gov/, Release 1.0.18 of 2018-03-27, edited by f. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller, and B. V. Saunders (NIST, 2018).
[45]M. Newville, T. Stensitzki, D. B. Allen, and A. Ingargiola, LMFIT: Non-Linear Least-Square Minimization and Curve-Fitting for Python (2014), available at https://doi.org/10.5281/zenodo.11813.
[46]M. Tehei, R. Daniel, and G. Zaccai, Eur. Biophys. J. **35**, 551 (2006).
[47]J. Wuttke, Phys. Rev. E **62**, 6531 (2000).
[48]J. Wuttke and M. Zamponi, Rev. Sci. Instrum. **84**, 115108 (2013).
[49]U. Lehnert, V. Reat, M. Weik, G. Zaccai, and C. Pfister, Biophys. J. **75**, 1945 (1998).
[50]J. M. Zanotti, M. C. Bellissent Funel, and S. Chen, Europhys. Lett. **71**, 91 (2005).
[51]S. Combet and J. M. Zanotti, Phys. Chem. Chem. Phys. **14**, 4927 (2012).
[52]J. D. Nickels, H. O'Neill, L. Hong, M. Tyagi, G. Ehlers, K. L. Weiss, Q. Zhang, Z. Yi, E. Mamontov, J. C. Smith, and A. P. Sokolov, Biophys. J. **103**, 1566 (2012).
[53]L. Hong, D. C. Glass, J. D. Nickels, S. Perticaroli, Z. Yi, M. Tyagi, H. O'Neill, Q. Zhang, A. P. Sokolov, and J. C. Smith, Phys. Rev. Lett. **110**, 028104 (2013).
[54]C. D. Andersson, N. Martinez, D. Zeller, A. Allgardsson, M. M. Koza, B. Frick, F. Ekström, J. Peters, and A. Linusson, J. Phys. Chem. B **122**, 8516 (2018).
[55]J. Peters, N. Martinez, M. Trovaslet, K. Scannapieco, M. M. Koza, P. Masson, and F. Nachon, Phys. Chem. Chem. Phys. **18**, 12992 (2016).
[56]S. Sacquin-Mora, P. Sebban, V. Derrien, B. Frick, R. Lavery, and C. Alba-Simionesco, Biochemistry **46**, 14960 (2007).
[57]M. Erlkamp, J. Marion, N. Martinez, C. Czeslik, J. Peters, and R. Winter, J. Phys. Chem. B **119**, 4842 (2015).
[58]A. Paciaroni, E. Cornicchi, A. De Francesco, M. Marconi, and G. Onori, Eur. Biophys. J. **35**, 591 (2006).
[59]W. Doster and M. Settles, Biochim. Biophys. Acta, Proteins Proteomics **1749**, 173 (2005).
[60]See https://www.ill.eu/fr/users-en/instruments/instruments-list/in15/characteristics/ for information about the IN15 instrument characteristics; accessed 25 July 2018.
[61]A. M. Gaspar, S. Busch, M.-S. Appavou, W. Haeussler, R. Georgii, Y. Su, and W. Doster, Biochim. Biophys. Acta, Proteins Proteomics **1804**, 76 (2010), Includes Special Section: Protein-Water Interactions.
[62]G. M. Paternò, J. R. Stewart, A. Wildes, F. Cacialli, and V. García Sakai, Polymer **105**, 407 (2016) .

B.2. Changes in dynamics of α-chymotrypsin due to covalent inhibitors investigated by elastic incoherent neutron scattering

To the following publication of Andersson et al. [141] my contribution was

- partly the neutron data reduction and corrections,
- the evaluation of the neutron data to obtain the MSD [and MSPF, finally not shown] and the summed intensities in collaboration with J. Peters (basis for Figure 2 in the publication).

PCCP

PAPER

View Article Online
View Journal | View Issue

Cite this: *Phys. Chem. Chem. Phys.*, 2017, **19**, 25369

Changes in dynamics of α-chymotrypsin due to covalent inhibitors investigated by elastic incoherent neutron scattering†

C. D. Andersson, [a] N. Martinez,[bc] D. Zeller,[bc] S. H. Rondahl, [d] M. M. Koza, [b] B. Frick, [b] F. Ekström, [d] J. Peters *[bc] and A. Linusson *[a]

Received 15th June 2017,
Accepted 1st September 2017

DOI: 10.1039/c7cp04041e

rsc.li/pccp

An essential role of enzymes is to catalyze various chemical reactions in the human body and inhibition of the enzymatic activity by small molecules is the mechanism of action of many drugs or tool compounds used to study biological processes. Here, we investigate the effect on the dynamics of the serine protease α-chymotrypsin when in complex with two different covalently bound inhibitors using elastic incoherent neutron scattering. The results show that the inhibited enzyme displays enhanced dynamics compared to the free form. The difference was prominent at higher temperatures (240–310 K) and the type of motions that differ include both small amplitude motions, such as hydrogen atom rotations around a methyl group, and large amplitude motions, such as amino acid side chain movements. The measurements were analyzed with multivariate methods in addition to the standard univariate methods, allowing for a more in-depth analysis of the types of motions that differ between the two forms. The binding strength of an inhibitor is linked to the changes in dynamics occurring during the inhibitor-enzyme binding event and thus these results may aid in the deconvolution of this fundamental event and in the design of new inhibitors.

1. Introduction

Enzymes are dynamical by nature and catalyze various chemical reactions in biological systems, *in vivo* or in biochemical applications *in vitro*. The motions of enzymes are functionally important and especially in the events leading up to catalysis, where conformational changes in the enzymes may facilitate or impede this process.[1–3] Inhibition of the catalytic ability by reversible (non-covalent) or irreversible (covalent) inhibitors is an invaluable method to control enzymatic activity and to study enzymatic function. Understanding these mechanisms is of fundamental importance in many research fields, including drug discovery,[4] enzyme engineering,[5] and enzyme function studies, since it may aid in the design of more potent inhibitors and functionally modified enzymes. The effects of inhibitor binding on local and overall enzyme dynamics have been investigated with various experimental techniques including nuclear magnetic resonance (NMR)[6] and incoherent neutron scattering.[7] It has become evident that these effects are not easily predicted and that enzyme inhibition can cause increased,[8–10] decreased,[11,12] or no change in dynamics.[13,14] For instance, we have recently shown with neutron scattering experiments that the covalently bound nerve agent Soman causes a stiffening of the enzyme acetylcholinesterase,[15] and does not affect the dynamics of butyrylcholinesterase at comparable temperature ranges.[16] Here, we have investigated the dynamical effect that covalent inhibition causes in the serine protease α-chymotrypsin.

Classic serine proteases such as chymotrypsin, subtilisin, carboxypeptidase Y, and Clp protease, are involved in biological processes including digestion, reproduction, immune response, apoptosis and hemostasis,[17] where they degrade proteins and peptides by cleaving peptide bonds. Serine proteases bearing the chymotrypsin fold (*e.g.* chymotrypsin, trypsin, thrombin, and elastase) belong to the clan PA, and the cleavage reaction these enzymes perform is catalyzed by a clan-specific catalytic triad consisting of amino acids His–Asp–Ser present in the enzymes' active site.[18] This clan of enzymes has been extensively studied and includes targets for drug research programs,[19] for example, lowering of blood pressure (thrombin and coagulation factor Xa), and pancreatitis (trypsin-like). These enzymes are also of interest in biotechnological applications making use of their catalytic ability in, for example, prodrug design,[20–22] enzyme engineering,[23,24] and functional nanomaterials.[25] The structure of α-chymotrypsin was revealed in 1967[26] and the enzyme consists of two longer and one short 13 amino acid chains (Fig. 1). The catalytic site is

[a] *Department of Chemistry, Umeå University, SE-90187 Umeå, Sweden. E-mail: anna.linusson@umu.se*
[b] *Institut Laue Langevin, F-38042 Grenoble Cedex 9, France. E-mail: peters@ill.fr*
[c] *Univ. Grenoble Alpes, IBS and LiPhy, F-38044 Grenoble, France*
[d] *CBRN Defence and Security, Swedish Defence Research Agency, SE-90621 Umeå, Sweden*

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c7cp04041e

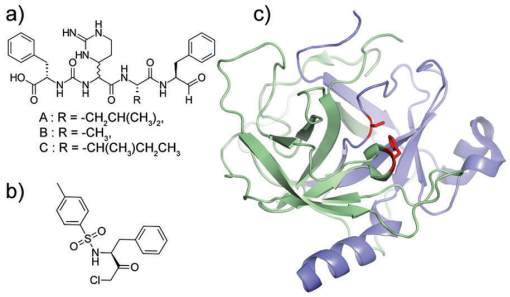


Fig. 1 Molecular structure of (a) chymostatin, a mixture of molecules A, B and C with different side chains indicated by R, (b) TPCK, and (c) α-chymotrypsin (PDB code 1YPH)[33] where the three protein chains are colored in white, green and blue, respectively, and the amino acids that participate in catalysis and that the inhibitors bind to are indicated in red.

positioned between the two long chains in a shallow gorge, with a deeper part where inhibitors or parts of inhibitors may bind.[27]

Numerous investigations have dealt with the binding mechanism of inhibitors of α-chymotrypsin[28,29] and the inhibitor-enzyme complex that are formed (see for instance ref. 27 and 30–32) but not much is known about the effect of inhibitor-binding on the enzyme dynamics at sub-nanosecond timescales. In this study, the motions of α-chymotrypsin were investigated in its free form and when inhibited with two different covalent inhibitors, 1-chloro-3-tosylamido-4-phenyl-2-butanone (TPCK) and chymostatin (Fig. 1a and b). These two inhibitors bind in the enzyme catalytic site; TPCK to His57[28] and chymostatin to Ser[195] (Fig. 1c).[29] The two inhibitors were selected based on that they bind to different amino acids and differ in size and, thus, they supposedly perturb the structure of α-chymotrypsin differently.

The flexibility of α-chymotrypsin was investigated with neutron spectroscopy, which is a powerful method to probe molecular motions on the atomic level. The incoherent neutron scattering cross section of the hydrogen atom is much higher than that of all other types of atoms present in biological systems,[34] and thus incoherent neutron scattering experiments reveal mainly the motions of hydrogen atoms and of the molecular groups to which they are bound. Enzymatic activities occur on micro- to millisecond time scale, but a recent study supported a correlation between activity and molecular dynamics at a nanosecond time scale within the family of human cholinesterases.[35] To evaluate the results of the neutron scattering measurements we used the classical univariate analysis methods such as summed intensities and atomic mean-square displacements (MSD).[36] In addition, we used multivariate methods including principal component analysis (PCA)[37,38] and orthogonal partial least squares discriminant analysis (OPLS-DA),[39,40] methods previously not used in this context, to elucidate more detailed effects of the inhibitors on the dynamics of α-chymotrypsin on the pico- to nanosecond time scale. In this way, we explored the data using two substantially different methods, the first one averaging over limited range of scattering angles, the second one exploiting all accessible angles or scattering vectors in a statistical meaningful way. This is an alternative to the self-distribution function procedure,[41] which permits to separate different molecular motions in various time windows from elastic incoherent neutron scattering data. Another successful method was suggested by L. Hong *et al.*[42] to identify global and internal protein motions from spin-echo and simulation data through wave vector dependent diffusion coefficients. Here we wanted to investigate how far an analysis distinguishing all wave vectors could permit to shed light on different motions and contributions within a rather limited space-time window defined through the resolutions of the three spectrometers used.

2. Experimental section

2.1 Sample preparation

Chymotrypsin is readily available as powder and very stable as a solid and in solution (pH 2–9) at temperatures up to 54.3 °C,[43] making the enzyme suitable for neutron scattering experimental conditions. 500 mg of α-chymotrypsin from Bovine pancreas and treated with 1-chloro-3-tosylamido-7-amino-2-heptanone to inhibit residual trypsin activity (Worthington Biochemical Corporation) was dissolved in 24 mL of a 25 mM ammonium acetate/deuterated water (D_2O) buffer. After 10 min of stirring on ice, the sample flash frozen in liquid nitrogen and freeze dried for 24 h. The dry protein powder was stored in −20 °C. Four samples were prepared from this dry protein powder: two equivalent samples including the inhibitor-free α-chymotrypsin, hereafter called CT1 and CT2, one sample with chymotrypsin inhibited by 1-chloro-3-tosylamido-4-phenyl-2-butanone (TPCK, CAS 402-71-1) and one with chymotrypsin inhibited by chymostatin (CAS 9076-44-2), giving samples CT/TP and CT/CS, respectively (Fig. 1). The average molecular weight (MW) of the three compounds in chymostatin (A, MW 607.7, 82%; B, MW 593.7, 11%; and C, MW 607.7, 7%) was estimated to 606.16 g mol^{-1}. Inhibitor stock solutions were prepared as follows. 6.06 mg (0.01 mmol) chymostatin was dissolved in 300 μL dimethylsulfoxide (DMSO) and then dissolved in ammonium acetate/D_2O buffer. 3.52 mg (0.01 mmol) TPCK was dissolved in 267 μL DMSO/EtOH 1 : 1 and then suspended in 12 mL ammonium acetate/D_2O buffer (resulting in a turbid solution). The inhibited chymotrypsin samples were prepared by addition of two (mole) equivalents of inhibitor to chymotrypsin in solution; dry chymotrypsin was dissolved in the inhibitor solution whereby all solutions became clear. Samples were freeze dried according to the protocol above. Protein samples were transferred to flat neutron scattering sample holders (3 × 4 cm^2) of 1 mm thickness and put under vacuum until completely dry (monitoring sample weight) followed by rehydration with D_2O to a final amount of D_2O/sample ratio (g g^{-1}) of 0.40, corresponding to at least one hydration layer around the protein. Final sample weights in the sample holder were 126.5 mg (CT1), 163.4 mg (CT2), 185.2 mg (CT/CS), and 147.9 mg (CT/TP).

2.2 Elastic incoherent neutron scattering experiments

We measured our enzyme samples with elastic incoherent neutron scattering (EINS) on three instruments, IN6,[44] IN13,[45] and IN16B[46] at the Institut Laue Langevin (ILL, France). The three

experiments were done consecutively using the same samples. The highest temperature reached was 310 K, corresponding to the human body temperature; therefore, no enzyme denaturation was expected. Both inhibited α-chymotrypsin samples (CT/TP and CT/CS) and one "free" α-chymotrypsin sample (CT2) were measured on all three spectrometers. The additional α-chymotrypsin sample (CT1) was included in the IN6 experiment to analyze inter-sample variation. The three spectrometers were chosen for their distinctly different time and spatial scale coverage. All three instruments are sensitive to local motions, but they give different information regarding the time window in which motions of different amplitudes occur. IN6 is a cold neutron time-of-flight spectrometer permitting to determine the incident wavelength and, thus, the instrumental energy resolution by a monochromator. In the present experiment, we opted for an incident wavelength of 5.1 Å corresponding to an energy resolution- and time window of 90 μeV and 7.5 ps (calculated according to Magazù *et al.*,[47]), respectively. Motions at this time scale correspond to very fast local motions of atoms. The elastic scattering was obtained on this time-of-flight spectrometer by integrating the peak in the range from −0.12 to 0.12 meV. A figure presenting IN6's resolution function and the integration limits is shown in the ESI† (Fig. S1). Each instrument gives access to a different Q range, Q being the momentum transfer between the incident and the scattered neutron in units of $\hbar$. Q is inversely related to space dimensions or in our case to the amplitudes of motion, which can be calculated through Bragg's law $L \propto 2\pi/Q$.[48] IN6 permits to measure intensities at Q values of the detector in between 0.4 and 2.0 $Å^{-1}$, corresponding to length scales between 16 and 3 Å. IN13 is a thermal backscattering spectrometer using the wavelength of 2.23 Å. The time window is about 100 ps and it covers a spatial scale of 1–30 Å. Such a time range typically covers local elastic vibrations and rotations of amino acid side chains at the surface.[49] IN16B has the highest energy resolution of about 1 μeV full width at half maximum and gives access to motions up to 1 ns, with length scales comparable to IN6 (3–16 Å). These motions can be associated to torsional vibrations of buried molecular sub-groups.[49] Following the evolution of the dynamics over time permits the detection of any correlations between different time scales, and measurements at the different Q values inform about the extension of the motions. However, as motions corresponding to the same Q value are observed on different energy resolutions and thus time scales, they are nonetheless not directly comparable and have to be considered and commented individually.

The samples were hydrated with D_2O at 0.4 g g^{-1} to highlight the motions of the hydrogens in the sample and not in the surrounding water. This prevented whole enzyme rotation and translation in the samples, so that we could focus on internal protein dynamics. The EINS data collected on the three instruments were analyzed individually, however following the same procedure. It is possible to show that the summed intensities are, up to a Q-range dependent multiplier, inversely proportional to the root of the MSD, but are less affected by statistical errors than the MSD.[15] They were calculated first by summing over the whole Q-range of each instrument. Furthermore within the so called Gaussian approximation,[50] which assumes that the distribution of the atoms around their average position follows a Gaussian distribution, the scattered elastic incoherent intensity (S_{el}) is given by the dynamic structure factor at zero energy exchange (S_0)

$$S_{el}(Q, \omega = 0 \pm \Delta E) \approx S_0 \exp\left(-\frac{1}{3}\langle u^2\rangle Q^2\right) \quad (1)$$

where ω and Q are the exchanged energy and momentum in units of $\hbar$, respectively, $\langle u^2\rangle$ is the MSD, and ΔE is the half width at half maximum of the instrumental energy resolution. The approximation is strictly valid for $Q \to 0$ and it holds up to $\langle u^2\rangle\, Q^2 \approx 1$. We determined the Gaussian approximation to be valid for the Q-ranges of 0.49–1.24 $Å^{-1}$ on IN6, 0.5–1.67 $Å^{-1}$ on IN13 and 0.7–1.48 $Å^{-1}$ on IN16B. The MSD can be obtained for each temperature by the slope of the semi-logarithmic plot of the incoherent scattering function through

$$\langle u^2\rangle \approx -3\frac{\mathrm{d}\ln S_{el}(Q, \omega = 0 \pm \Delta E)}{\mathrm{d}Q^2} \quad (2)$$

The MSD can be a measure for the flexibility of the biological system at a given temperature.[51] To obtain the intensities scattered by the sample only, scattering from the empty sample holder was subtracted, and the data were normalized to the lowest measured temperature data (IN6: 80 K, IN13: 20 K, IN16B: 40 K). At such low temperature, all motions are frozen and the neutron intensities thus reflect the relative detector efficiency and the instrumental resolution. The lowest temperature was different on the various spectrometers for technical reasons, but the differences in the data are negligible within the range of cryo-temperatures. The acquisition was continuous on IN6 and IN16B with ramps of 1 K min^{-1} below 80 K and 0.5 K min^{-1} above 80 K and on discrete points on IN13 (with steps of 10 or 20 K) with 1 h counting time below 280 K and 2 h above 280 K. Absorption correction was based on the correction formula of Paalman–Pings.[52] The complete data reduction was carried out using the LAMP software available at ILL.[53]

2.3 Multivariate data analysis

PCA[37,38] is a projection method that visualizes the main variation in multivariate data by transforming a multidimensional set of correlated variables to a smaller set of uncorrelated new variables. These are called principal components (PCs) and the PCs are eigenvectors of the original data. Here, the data matrix (**X**) consists of intensities and the matrix row elements are the measurements in sequential order at the different temperatures (T). Column elements (variables) are the Q-values at the detector (Q:s), yielding $\mathbf{X}(Q_i, T_j)$, where $i = 1,\ldots, M$ dimensions of Q space and $j = 1,\ldots, N$ number of measurements. The intensities were normalized to the intensity at the lowest measured temperature and were mean-centered prior to modeling. The first PC (PC1) explain (per definition) the main variation in the data. After calculation of PC1, each measured data point received a new value, a score value s to build up a score vector s_1, which is the eigenvector that has the largest eigenvalue. The contribution of the original variables, Q:s, to

the new score vector (the position of the samples in the score vector) is described by the values in the loading vector **l**. The second PC (PC2) is placed orthogonally to PC1, so that it captures the remaining main variation (second largest eigenvalue), the third PC (PC3) is placed orthogonally to PC2, and so on. These PCs give rise to a new decomposition matrix (**SL**′) according to

$$\mathbf{X} = \bar{\mathbf{X}} + \mathbf{SL}' + \mathbf{E} \quad (3)$$

where $\bar{\mathbf{X}}$ is the **X** matrix average, **S** is the score matrix, **L**′ is the loading matrix, and **E** is the residual. The *l*- and *s*-values are visualized in two separate plots; the loading plot shows the *Q*:*s* and their weights on the PCs and the score plot shows how the EINS measurements differ in terms of their intensities in the *Q*-range.

OPLS-DA[39,40] makes use of a predefined matrix (**Y**) containing class assignment based on the samples. This information is used to decomposed the **X** matrix containing the measured *Q* intensities in such way to that it describes large variation in **X** and correlate with **Y**. **X** is decomposed into three parts according to

$$\mathbf{X} = \mathbf{S}_\mathrm{P}\mathbf{L}_\mathrm{P}' + \mathbf{S}_\mathrm{O}\mathbf{L}_\mathrm{O}' + \mathbf{E} \quad (4)$$

where $\mathbf{S}_\mathrm{P}$ is the predictive score matrix, $\mathbf{L}_\mathrm{P}'$ the predictive loading matrix, $\mathbf{S}_\mathrm{O}$ the corresponding **Y**-orthogonal score matrix, $\mathbf{L}_\mathrm{O}'$ the loading matrix of the **Y**-orthogonal components, and **E** is the residual matrix of **X**. Thus, OPLS-DA discriminates between the inter-class variation (class differences) shown by the predictive scores ($\boldsymbol{s}_{\mathrm{p},1}$) and intra-class variation (differences within a class) shown by the orthogonal scores ($\boldsymbol{s}_{\mathrm{o},1}$). The predictive loadings ($\boldsymbol{l}_{\mathrm{P},1}$), or in this case the weights ($\boldsymbol{w}^*$), can be analyzed to identify the *Q*-values (*i.e.*, the movement length scale) that gives rise to the classes. The absolute value of individual weight indicates the importance of that specific variable to the predefined class assignment, the larger value, the higher importance. The sign of the weight shows if the related variable is positively or negatively correlated to that particular class.

The quality and statistical significance of the multivariate models was determined from the proportion of the variation in the original data that was explained by the model, *i.e.*, the cumulative sum of squares of the entries ($R^2\mathbf{X}$(cum)) and eigenvalues (PCA), and the cumulative sum-of-squares of all the **y**-variables explained by the extracted OPLS-DA components ($R^2\mathbf{Y}$(cum)) (OPLS-DA). The models were further tested using leave-*n*-out cross-validation[54] (*n* equal to 1/7 of the data set) giving rise to a q^2(cum)-value comprising all model PCs or OPLS-DA components. Multivariate modeling was performed in the SIMCA software[55] and a more detailed description of PCA, OPLS-DA, and cross validation is presented in the ESI.†

3. Results and discussion

The dynamics of free (samples CT1 and CT2) and inhibited α-chymotrypsin (samples CT/CS and CT/TP) was measured on hydrated powders on the three instruments: IN6, IN13, and IN16B. The resulting EINS data were analyzed by extracting the intensities summed over all scattering angles (the entire *Q*-range) and MSD based on a shorter *Q*-range. Inspection of the three figures showing the summed intensities (Fig. 2) revealed the following: (i) negligible difference in the dynamics of the two samples of the free enzyme (CT1 and CT2) was visible on IN6. (ii) Covalent inhibitors affected the enzyme so that it became more flexible and this was true for both inhibited enzyme samples (CT/TP and CT/CS) and on all instruments. (iii) The differences in dynamics between the inhibited samples were within the error bars, but the tendency was that the effect of TPCK was larger than that of chymostatin. (iv) The results from IN6 and IN13 showed an effect of hydration; the CT2 sample had an increase in intensity on these instruments between 220 K and 270 K, which can be associated with water Bragg peaks. The hydration effect is seen in *Q*-ranges not included in the MSD calculations.

PCA was used to visualize general trends and groupings in the EINS data using the whole *Q*-range to visualize differences in the dynamics of the α-chymotrypsin samples. The PCA models were generally of high statistical quality where ∼99% of the original variation in the data was explained with a cross-validation q^2 of 0.98–0.99. Most of the variation was described by PC1 (97–99%) and a few percent by PC2 and PC3. Model details including statistics and number of samples and measurements are presented in the ESI,† Tables S1–S3. The main difference in enzyme dynamics between the different measurements, shown in PC1 (*e.g.*, for IN6 in Fig. 3a), was related to the different temperatures at which a sample was measured, but also the differences between free and inhibited α-chymotrypsin. This was true for all samples measured on all three instruments (Fig. 3 and 4).

The enzyme dynamics increased with the temperature; the intensities were negatively correlated with the temperature, as shown previously for proteins.[56] This relation was manifested in the PCA by the fact that the high *Q*:*s* were generally found at the right most in the loading plot, connected to the low temperatures to the right in the score plot. The different PCAs based on data of the three instruments, also showed groups consisting of the free (CT2) and inhibited α-chymotrypsin (CT/CS and CT/TP) and these groups were prominent mainly at higher temperatures (Fig. 3 and 4).

3.1 Measurement reproducibility

We analyzed the results of the two free α-chymotrypsin samples CT1 and CT2 on IN6 to determine the variation in measurement results on two similar samples prepared in parallel. Except for the Bragg-peak effects (see separate analysis below) the two samples appear equal within error bars from the summed intensities (Fig. 2a). The MSD and PCA showed that CT2 was slightly more dynamic in general (Fig. 2b and 3a) and more similar to the inhibited α-chymotrypsin samples in the MSD- and score plots, mainly at low temperatures. The difference between CT1 and CT2 may stem from differences in sample preparation, weights, and hydration. Nevertheless, we can conclude that the difference is

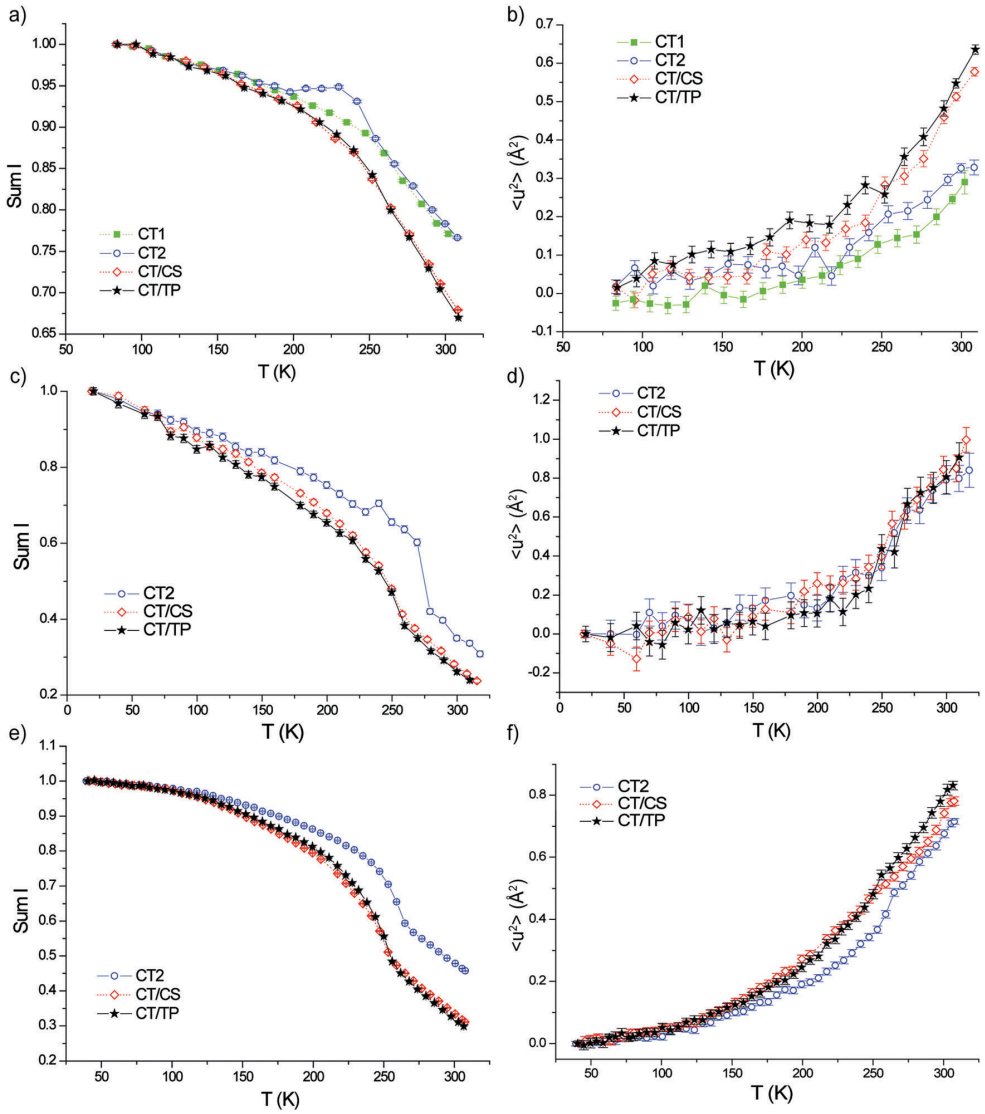


Fig. 2 Neutron intensities in arbitrary units summed over the available scattering angles (left column) and the extracted MSD (right column), both as function of temperature. Data are from IN6 (a and b), IN13 (c and d), and IN16B (e and f). The error bars were based on Poisson statistics applied to the raw data. The error bars of the MSD, extracted according to eqn (2), were obtained by linear regression when fitting the data.

small, and most importantly, considerably smaller than the difference between free and inhibited enzyme. Notably, the difference between CT1 and CT2 diminishes with increasing temperature, which is seen in both univariate (Fig. 2) and the multivariate analysis (Fig. 3a).

3.2 Effects of water phase transitions

On IN13 and IN16B, particularly for the CT2 sample, there was a sharp decrease in the summed intensities around 250–275 K (Fig. 2c and e) due to the melting of water. An effect of a water phase transition that is different from melting was seen for the CT2 sample measured on IN6 around 240 K (Fig. 2a and 3b). In fact, water may undergo several phase transitions but not all phases are thermodynamically stable. At a very low temperature, ice has an amorphous structure and at 150 K, it transforms into a quasi-cubic structure. Beyond 200 K the quasi-cubic structure passes into the hexagonal phase, which is very fast above 240 K, also reported by Koza *et al.*[57] Here, only the cubic to hexagonal phase transition makes fluctuations that contribute in a significant way to the dynamics probed by the neutrons. This effect was also weakly present in the other samples, which is only apparent by the multivariate analysis in Fig. 3b, where all samples show higher

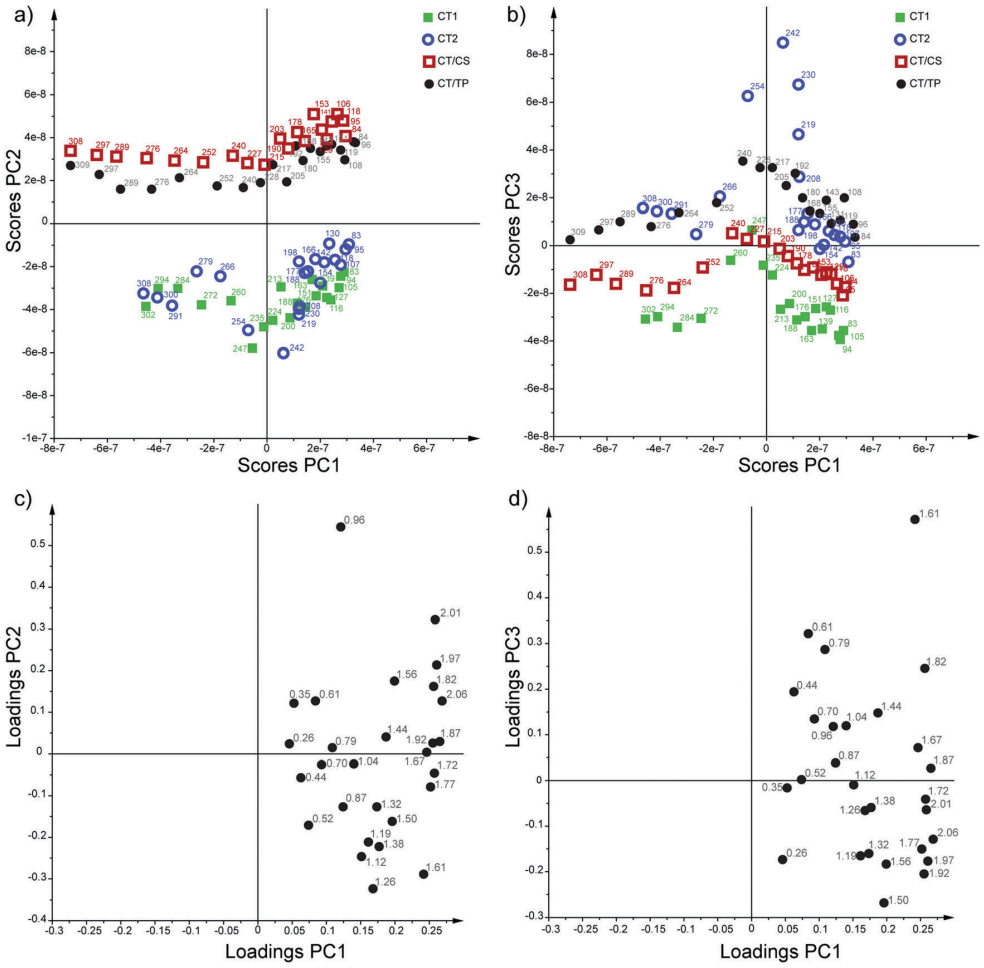


Fig. 3 PCA analysis of the measurements from IN6 showing in (a) and (b) score plots of the samples displaying the main differences between the samples (indicated with sample symbol and measurement temperature) based on neutron intensities, and in (c) and (d) loading plots with loading values for each variable Q indicated with the Q value in Å^{-1}. PC; principal component.

score values in principal component three (PC3) in that temperature range. The sharp increase in the intensities were mainly manifested in Q equal to 1.61 Å^{-1}, which matches with values for ice Bragg peaks that have been reported.[57] Such an effect is, however, negligible for the MSD extracted from the restricted Q-range, since the ice Bragg peaks appear in a Q-range not included in the MSD analysis. We concluded that the difference in dynamics between the two inhibitor-free samples CT1 and CT2 mainly occurred at mid-range temperatures due to Bragg peaks, and does not stem from a difference between the samples with any biological mechanistic relevance, but rather from slight sample dissimilarities.

3.3 Differences in dynamics between free and inhibited α-chymotrypsin

The summed intensities over the whole Q-range and the PCA of the neutron scattering data showed that there was a clear difference in the dynamics between the inhibited and free form of α-chymotrypsin; both covalent inhibitors affected the enzyme so that it became more flexible. The difference in enzyme motion is pronounced in the summed intensities measured on all instruments at mid- to high temperatures, but not at the lowest temperatures (Fig. 2a, c and e).

The MSD analysis of the IN6 data also showed clear differences between the inhibited and the free samples (Fig. 2b), where the MSD increases by ∼0.3 Å^2 at the highest temperature. In the IN13 data, the MSD had larger standard deviations, which did not allow a significant separation between inhibited and free α-chymotrypsin (Fig. 2d). On IN16B, the difference in the total elastic intensity was the largest between 200 K and 260 K. A significant difference in the MSD was also apparent at the highest temperatures (Fig. 2f), corresponding to a difference in MSD of ∼0.065 Å^2 between the CT/CS and CT2 samples and of ∼0.05 Å^2 between CT/TP and CT2 samples. In the PCA, notably

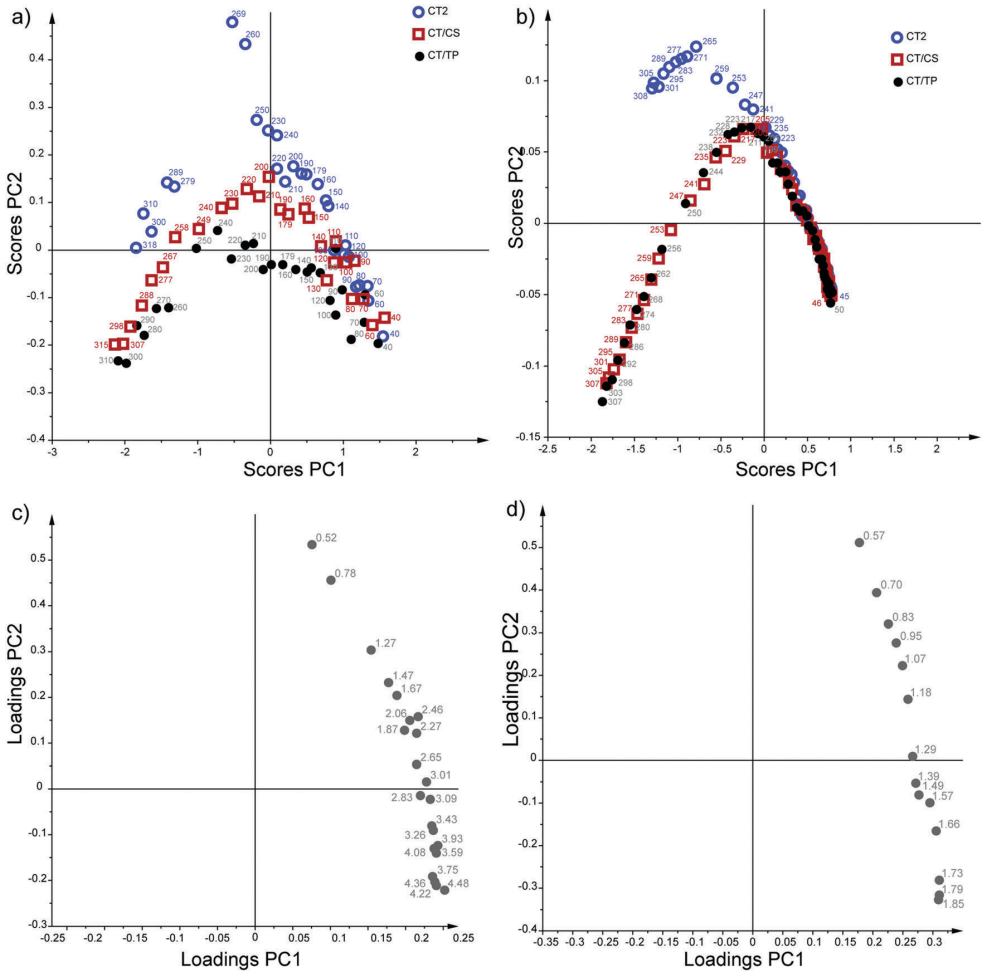


Fig. 4 PCA analysis of the measurements with score plots of the samples displaying the main differences between the samples (indicated with sample symbol and measurement temperature) based on neutron intensities from (a) IN13 and (b) IN16B, and loading plots with loading values for each variable Q indicated with the Q value in Å^{-1} from (c) IN13 and (d) IN16b. PC; principal component.

including intensities of the whole Q-range, the difference between the free and inhibited enzyme was pronounced across all temperatures in the IN6 measurements (Fig. 3a). From the IN13 and IN16B measurements, mainly at mid- to high temperatures (240–310 K), there are two groups of measurements including free (CT2) and inhibited (CT/CS and CT/TP) α-chymotrypsin samples (Fig. 4a and b). On IN13, the difference is less pronounced but clearly there, especially at higher temperatures (~280–315 K). To explore the cause of the higher mobility of the inhibited α-chymotrypsin further, and to more clearly pinpoint the kind of atomic motions that were involved, the samples were classified into two classes corresponding to the free and inhibited α-chymotrypsin sample(s). The analysis was focused on the data at the biologically relevant temperature interval 270–310 K.

The variation in Q intensities that could explain class separations (the $R^2\mathbf{X}$ of the OPLS-DA predictive component) amounted to 60%, 47%, and 76% of the total variation collected at IN6, IN13, and IN16B, respectively, showing that a substantial amount of the information in the EINS data was related to the class differences. Model statistics are presented in the ESI,† Tables S1–S3. All of the classification models showed, as expected, a clear separation between inhibited and free α-chymotrypsin, with the CT2 (or CT1 and CT2) samples to the left and CT/CS and CT/TP samples grouped to the right in the score plot (Fig. 5a, c and e). All the weights in right panels (Fig. 5b, d and f) had negative values and this relates to the position of the inhibitor-free samples to the left and negative side in the score plot. This meant that the free enzyme samples (at 270–310 K) generally had higher intensities in the whole Q-range and were thus less mobile compared to the inhibited enzymes. Note that the same spatial scale or Q-value seen on different instruments corresponds to different time scales

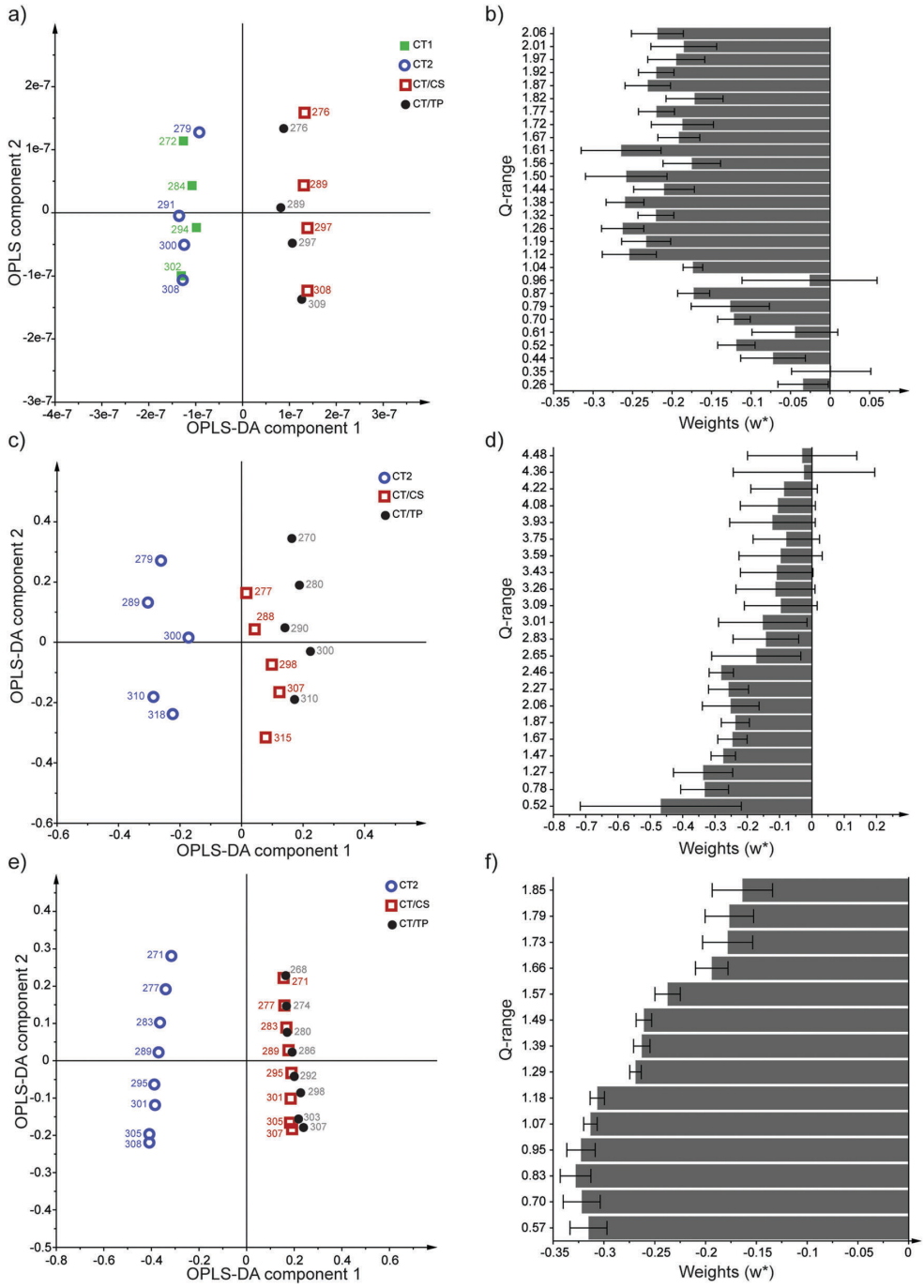


Fig. 5 Classification of inhibitor-free *versus* inhibited α-chymotrypsin in the temperature interval 270–310 K with scores, showing class separations, and weight for OPLS-DA component 1, showing scattering angle (*Q*-range) contributions to class separation, from measurements on IN6 (a and b, respectively), IN13 (c and d, respectively) and IN16B (e and f, respectively). Error bars are standard errors with a confidence level of 95% calculated by Jack-knifing.

according to the instrumental resolution. Furthermore, from the OPLS-DA weight plots (Fig. 5), it was clear that the high-range *Q*:s are most important for the motions seen on IN6. On IN13, the *Q*:s contributed similarly within the *Q*-ranges accessible on IN6 or

IN16B but higher-range *Q*:*s* decreased in weight, whereas a decrease with *Q* was clear on IN16B. In the 7.5 ps time window (IN6) the inhibited enzymes were more dynamic than the free form due to more small-amplitude high velocity motions (higher intensities at high scattering angles; Fig. 5a and b), such as the rotations of hydrogens bound to, for example, carbon in methyl groups. Water translational motions is monitored at a *Q*-range below ~1.5 Å^{-1} but the short time window on IN6 cannot capture this motion in total, as shown by small and non-significant weights in the OPLS-DA weights plot (Fig. 5b). The difference between the inhibited and free samples seen in the PCA analysis for the whole temperature range (Fig. 3a), could be corroborated by an OPLS-DA model including the whole temperature range that was statistically significant and showed a separation between the free and inhibited α-chymotrypsin (details are presented in the ESI,† Fig. S4).

Even smaller amplitude motions measured with the broader *Q*-range monitor (*Q* of 2.27–4.48 Å^{-1}) at the IN13 instrument (100 ps time window) were not affected upon covalent inhibition, as these OPLS-DA weights were non-significant in the classification model (Fig. 5b and c) and are thus not important in the differentiation between free and inhibited α-chymotrypsin.

Neutron scattering intensities measured during the longest time window, 1 ns, at IN16B, showed that the increased dynamics of α-chymotrypsin upon inhibition was not only due to small amplitude motions but due to larger, slower motions increased as well. This can be concluded because all OPLS-DA weights in this *Q*-range (0.57–1.85 Å^{-1}) were significant (Fig. 5f).

The classification method OPLS-DA[39,40] was particularly suitable in our case, as this method filtered the data from scattering intensity differences between measurements that were not related to class separation (*e.g.*, the temperature dependences). The fact that there was a marked difference between the free and the inhibited enzyme with respect to the high *Q*-values on IN6 (Fig. 4b and Fig. S4, ESI†) might be interpreted as follows: Compared to the free enzyme, the inhibited underwent the dynamical transition at lower temperature (*i.e.*, around 200 K)[58] and in a more cooperative manner, meaning more particles undergoing the transition simultaneously. We therefore speculate that the bound inhibitors induced small collective motions in the enzyme, possibly also disturbed the water network, so that less energy was required to undergo the dynamical transition. The motions captured by IN16B concerned scales above 3.3 Å monitored during 1 ns and corresponding to amino acid side chains- and water molecule movements, and the inhibited enzyme was more dynamic when considering these kinds of movements. Importantly, such distinctions were not possible to identify from the MSD solely, which is based on the slope of the intensities with respect to Q^2.

3.4 Effects of different inhibitors

There was a weak tendency that α-chymotrypsin became more flexible when inhibited by TPCK compared to chymostatin, as evident by the difference between the two samples in the summed intensities at the highest temperatures from the IN6 and IN16B measurements (Fig. 2b and f). A statistically significant difference in MSD could not be seen between CT/TP and CT/CS at IN13 and at the highest temperature. These results are summarized in Fig. 6, showing the MSD of the samples at 300 K measured on the three instruments. No significant overall increase in $\langle u^2 \rangle$ is noted in between the IN13 and IN16B results, and only judging from the MSD results, this would indicate that no additional motions were recorded between 100 ps and 1 ns.

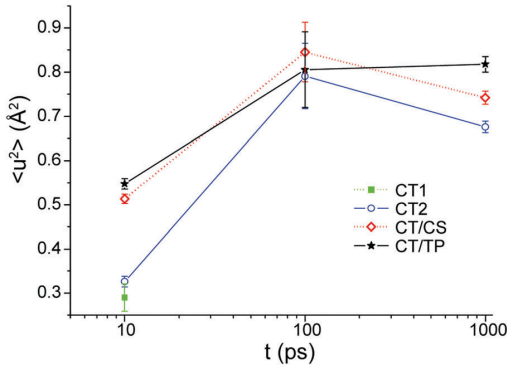


Fig. 6 MSD at 300 K for the different samples as function of the instrumental time window where 10, 100, and 1000 ps corresponds to the time window on IN6, IN13 and IN16B, respectively.

Nevertheless, the OPLS-DA (Fig. 5) showed that the *Q*-ranges contributing the most are not the same on the three instruments and thus only on average does the flexibility appear to be about the same on IN13 and IN16B. The more detailed multivariate analysis permits to shed light on this point. These findings might explain why different diffusion coefficients are often extracted from quasi-elastic neutron scattering data taken with different instrumental resolutions although the sample is identical.[59] The CT/TP sample showed a higher MSD compared to the CT/CS sample (on IN6 and IN16B) indicating that the former induced more motion in the α-chymotrypsin than the latter. Here, we note that TPCK is markedly smaller than chymostatin (Fig. 1), but still induce more motion. Nevertheless, the lack of structural data regarding the inhibitor-chymotrypsin complex and our data treatment does not allow for more precise conclusions, but we refer to the discussion in Peters *et al.*[15] where arguments are forwarded why inhibitors may act differently on the water network and how this could influence the global dynamics.

4. Conclusions

The dynamics of the serine protease α-chymotrypsin in its free form and when inhibited by covalently bound inhibitors TPCK and chymostatin was measured with EINS on the spectrometers IN6,[44] IN13,[45] and IN16B.[46] As expected, the dynamics for all samples increased with temperature, but the inhibited enzyme was shown to be more dynamical compared to the free, as was evident from both the summed intensities and the multivariate analysis including PCA and OPLS-DA. Other factors such as

sample preparation, hydration level, and sample weights could be ruled out as the cause for this difference. The increase in dynamics of the inhibited α-chymotrypsin was seen on all measured time-scales capturing movements such as methyl group rotations, water movements, and amino acid side chain movements. Contrary to the univariate methods, which average over effects and specific space dimensions the multivariate analysis could reveal subtle effects, such as the presence of Bragg peaks in all measurements and which *Q*-ranges and thus which type of motion constituted the difference in the enzyme dynamics. The OPLS-DA models pinpointed scattering angles that contained information on the difference in dynamics between inhibited and free enzyme, and gave statistical significance to this finding. At high temperatures on a short time-scale (IN6), mainly small amplitude motions, such as hydrogens rotating about carbon in methyl groups, contributed to the higher mobility of the inhibited enzyme. On IN16B, covering the longest time-scales, more motions that correspond to the movements of amino acid side chains could be seen in the inhibited enzyme. Thus, we conclude that the inhibited enzymes studied here underwent the dynamical transition at lower temperatures and in a more cooperative way leading to bigger amplitudes of motions. Here, we propose two simultaneous scenarios regarding the changes of the potential energy landscape of the enzymes. The inhibitors' interactions with the enzyme permitted more collective motions and more particles to overcome potential energy barriers in the conformational landscape simultaneously during the transition. Furthermore, the inhibitors influenced the water network around the enzymes in a way that permitted more degrees of motional freedom leading to a lowering of the potential energy barrier heights. Molecular dynamics simulations could eventually help to disentangle the different scenarios.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This work was funded by the Swedish Research Council (Dnr: 2014-4675). We are gratefully acknowledging the ILL for allocation of beam time.

References

1 V. Daggett, S. Schröder and P. Kollman, *J. Am. Chem. Soc.*, 1991, **113**, 8926–8935.
2 M. Kokkinidis, N. M. Glykos and V. E. Fadouloglou, *Adv. Protein Chem.*, 2012, **87**, 181–218.
3 A. Kohen, *Acc. Chem. Res.*, 2015, **48**, 466–473.
4 R. A. Copeland, M. R. Harpel and P. J. Tummino, *Expert Opin. Ther. Targets*, 2007, **11**, 967–978.
5 M. McAuley and D. J. Timson, *Appl. Biochem. Biotechnol.*, 2017, **181**, 83–90.
6 D. D. Boehr, H. J. Dyson and P. E. Wright, *Chem. Rev.*, 2006, **106**, 3055–3079.
7 R. M. Daniel, R. V. Dunn, J. L. Finney and J. C. Smith, *Annu. Rev. Biophys. Biomol. Struct.*, 2003, **32**, 69–92.
8 L. P. Yu, C. X. Zhu, Y. C. TseDinh and S. W. Fesik, *Biochemistry*, 1996, **35**, 9661–9666.
9 S. G. Yun, D. S. Jang, D. H. Kim, K. Y. Choi and H. C. Lee, *Biochemistry*, 2001, **40**, 3967–3973.
10 E. Balog, T. Becker, M. Oettl, R. Lechner, R. Daniel and J. Finney, *et al.*, *Phys. Rev. Lett.*, 2004, **93**, 028103.
11 A. Eletsky, A. Kienhofer, D. Hilvert and K. Pervushin, *Biochemistry*, 2005, **44**, 6788–6799.
12 Y. L. Miao, Z. Yi, C. Cantrell, D. C. Glass, J. Baudry and N. Jain, *et al.*, *Biophys. J.*, 2012, **103**, 2167–2176.
13 J. H. Davis and D. A. Agard, *Biochemistry*, 1998, **37**, 7696–7707.
14 M. Trapp, M. Trovaslet, F. Nachon, M. M. Koza, L. van Eijck and F. Hill, *et al.*, *J. Phys. Chem. B*, 2012, **116**, 14744–14753.
15 J. Peters, N. Martinez, M. Trovaslet, K. Scannapieco, M. M. Koza and P. Masson, *et al.*, *Phys. Chem. Chem. Phys.*, 2016, **18**, 12992–13001.
16 F. Gabel, M. Weik, P. Masson, F. Renault, D. Fournier and L. Brochier, *et al.*, *Biophys. J.*, 2005, **89**, 3303–3311.
17 L. Hedstrom, *Chem. Rev.*, 2002, **102**, 4501–4523.
18 N. D. Rawlings, F. R. Morton, C. Y. Kok, J. Kong and A. J. Barrett, *Nucleic Acids Res.*, 2007, **36**, D320–325.
19 B. Turk, *Nat. Rev. Drug Discovery*, 2006, **5**, 785–799.
20 G. B. Jones, M. O. Mitchell, J. S. Weinberg, A. V. D'Amico and G. J. Bubley, *Bioorg. Med. Chem. Lett.*, 2000, **10**, 1987–1989.
21 Y. Y. Ge, X. H. Wu, D. Z. Zhang and L. Q. Hu, *Bioorg. Med. Chem. Lett.*, 2009, **19**, 941–944.
22 J. Zhang, Y. F. Liu, L. Bo and C. A. Qiao, *J. Appl. Polym. Sci.*, 2011, **121**, 1992–1998.
23 C. S. Cummings, H. Murata, K. Matyjaszewski and A. J. Russell, *ACS Macro Lett.*, 2016, **5**, 493–497.
24 M. A. M. Latif, B. A. Tejo, R. Abedikargiban, M. B. A. Rahman and N. M. Micaelo, *J. Biomol. Struct. Dyn.*, 2014, **32**, 1263–1273.
25 R. Wang, Y. F. Zhang, D. N. Lu, J. Ge, Z. Liu and R. N. Zare, *Wiley Interdiscip. Rev.: Nanomed. Nanobiotechnol.*, 2013, **5**, 320–328.
26 B. W. Matthews, P. B. Sigler, R. Henderson and D. M. Blow, *Nature*, 1967, **214**, 652–656.
27 J. J. Birktoft and D. M. Blow, *J. Mol. Biol.*, 1972, **68**, 187–240.
28 G. Schoellmann and E. Shaw, *Biochemistry*, 1963, **2**, 252–255.
29 N. P. Tomkinson, I. J. Galpin and R. J. Beynon, *Biochem. J.*, 1992, **286**, 475–480.
30 J. J. Birktoft, D. M. Blow, R. Henderson and T. A. Steitz, *Philos. Trans. R. Soc., B*, 1970, **257**, 67–76.
31 A. Tulinsky and R. A. Blevins, *J. Biol. Chem.*, 1987, **262**, 7737–7743.
32 K. C. H. Chua, M. Pietsch, X. Z. Zhang, S. Hautmann, H. Y. Chan and J. B. Bruning, *et al.*, *Angew. Chem., Int. Ed.*, 2014, **53**, 7828–7831.
33 A. Razeto, B. Galunsky, V. Kasche, K. S. Wilson and V. S. Lamzin, *High resolution structure of bovine alpha-chymotrypsin*, 2006, DOI: 10.2210/pdb1yph/pdb.

34 V. F. Sears, *Neutron News*, 1992, **3**, 26–37.
35 J. Peters, M. Trovaslet, M. Trapp, F. Nachon, F. Hill and E. Royer, *et al.*, *Phys. Chem. Chem. Phys.*, 2012, **14**, 6764–6770.
36 S. W. Lovesey, *Theory of neutron scattering from condensed matter*, Oxford Science Publications, Oxford, UK, 1984, vol. I.
37 S. Wold, K. Esbensen and P. Geladi, *Chemom. Intell. Lab. Syst.*, 1987, **2**, 37–52.
38 J. E. Jackson, *A user's guide to principal components*, John Wiley & sons, Inc., New York, USA, 1991.
39 J. Trygg and S. Wold, *J. Chemom.*, 2002, **16**, 119–128.
40 M. Bylesjö, M. Rantalainen, O. Cloarec, J. K. Nicholson, E. Holmes and J. Trygg, *J. Chemom.*, 2006, **20**, 341–351.
41 S. Magazù, G. Maisano, F. Migliardo, G. Galli, A. Benedetto and D. Morineau, *et al.*, *J. Chem. Phys.*, 2008, **129**, 155103.
42 L. Hong, N. Smolin and J. C. Smith, *Phys. Rev. Lett.*, 2014, **112**, 158102.
43 A. Kumar and P. Venkatesu, *Chem. Rev.*, 2012, **112**, 4283–4307.
44 http://www.ill.eu/in6/home/, U.
45 F. Natali, J. Peters, D. Russo and F. Sonvico, *Neutron News*, 2008, **19**, 14–18.
46 B. Frick, E. Mamontov, L. van Eijck and T. Seydel, *Z. Phys. Chem.*, 2010, **224**, 33–60.
47 S. Magazù, F. Migliardo and A. Benedetto, *Rev. Sci. Instrum.*, 2011, **82**, 105115.
48 G. L. Squires, *Introduction to the theory of thermal neutron scattering*, Dover publications, Mineola, New York, USA, 1996.
49 J. A. McCammon and S. C. Harvey, *Dynamics of proteins and nuclear acids*, Cambridge University Press, Cambridge, UK, 1987.
50 A. Rahman, K. S. Singwi and A. Sjölander, *Phys. Rev.*, 1962, **126**, 986–996.
51 G. Zaccai, *Science*, 2000, **288**, 1604–1607.
52 H. H. Paalman and C. J. Pings, *J. Appl. Phys.*, 1962, **33**, 2635–2639.
53 LAMP, the Large Array Manipulation Program. http://www.ill.eu/data_treat/lamp/the-lamp-book/.
54 H. T. Eastment and W. J. Krzanowski, *Technometrics*, 1982, **24**, 73–77.
55 SIMCA 14, Umetrics AB. Box 7960, Umeå, Sweden.
56 F. Gabel, D. Bicout, U. Lehnert, M. Tehei, M. Weik and G. Zaccai, *Q. Rev. Biophys.*, 2002, **35**, 327–367.
57 M. M. Koza, B. Geil, H. Schober and F. Natali, *Phys. Chem. Chem. Phys.*, 2005, **7**, 1423–1431.
58 D. J. Bicout and G. Zaccai, *Biophys. J.*, 2001, **80**, 1115–1123.
59 M. Trapp, M. Tehei, M. Trovaslet, F. Nachon, N. Martinez and M. M. Koza, *et al.*, *J. R. Soc., Interface*, 2014, **11**, 20140372.

B.3. Influence of Enantiomeric Inhibitors on the Dynamics of Acetylcholinesterase Measured by Elastic Incoherent Neutron Scattering

To the following publication of Andersson et al. [99] my contribution was

- partly the neutron data reduction and corrections,
- the evaluation of the neutron data to obtain the MSD,MSPF and the summed intensities in collaboration with J. Peters (basis for Figures 2 and 4 in the publication).

THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

Cite This: J. Phys. Chem. B 2018, 122, 8516−8525

pubs.acs.org/JPCB

Influence of Enantiomeric Inhibitors on the Dynamics of Acetylcholinesterase Measured by Elastic Incoherent Neutron Scattering

C. David Andersson,[†] Nicolas Martinez,[‡,§] Dominik Zeller,[‡,§] Anders Allgardsson,[∥] Michael M. Koza,[‡] Bernhard Frick,[‡] Fredrik Ekström,[∥] Judith Peters,*[,‡,§] and Anna Linusson*[,†]

[†]Department of Chemistry, Umeå University, SE-90187 Umeå, Sweden

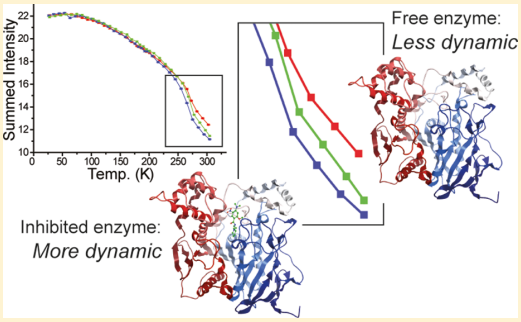
[‡]Institut Laue Langevin, F-38042 Grenoble Cedex 9, France

[§]Université Grenoble Alpes, IBS and LiPhy, F-38000 Grenoble, France

[∥]CBRN Defence and Security, Swedish Defence Research Agency, SE-90621 Umeå, Sweden

Supporting Information

ABSTRACT: The enzyme acetylcholinesterase (AChE) is essential in humans and animals because it catalyzes the breakdown of the nerve-signaling substance acetylcholine. Small molecules that inhibit the function of AChE are important for their use as drugs in the, for example, symptomatic treatment of Alzheimer's disease. New and improved inhibitors are warranted, mainly because of severe side effects of current drugs. In the present study, we have investigated if and how two enantiomeric inhibitors of AChE influence the overall dynamics of noncovalent complexes, using elastic incoherent neutron scattering. A fruitful combination of univariate models, including a newly developed non-Gaussian model for atomic fluctuations, and multivariate methods (principal component analysis and discriminant analysis) was crucial to analyze the fine details of the data. The study revealed a small but clear increase in the dynamics of the inhibited enzyme compared to that of the noninhibited enzyme and contributed to the fundamental knowledge of the mechanisms of AChE−inhibitor binding valuable for the future development of inhibitors.



INTRODUCTION

Acetylcholinesterase (AChE) is an essential enzyme that catalyzes the hydrolysis of the neurotransmitter acetylcholine (ACh) in the nervous system of higher organisms. AChE is an important target enzyme for the treatment of symptoms in Alzheimer's disease,[1] where the use of a reversible inhibitor of AChE leads to increased levels of ACh, slowing the decline in patient cognition and function.[2] Very few drug alternatives are available that target AChE, and donepezil is one of these drugs, whose adverse effects are mild to moderate but common, as recently reviewed.[3] No new drugs have been approved since the beginning of the 2000s, and roughly a dozen have failed clinical trials,[4] warranting new candidate drugs. In this quest, increased fundamental knowledge of the mechanisms of AChE−inhibitor binding and energetics of interactions are critical steps to efficient drug design.

Proteins, such as AChE, undergo transitions between conformational substates involving thermal motions of atoms. The overall energetics of these motions are an intrinsic factor in the binding free energy of a drug to a protein target. Compounds C5685-R and C5685-S (Figure 1, here abbreviated as 568R and 568S) inhibit the enzymatic activity of *Mus musculus* AChE (*m*AChE) by binding in the catalytic site of the enzyme with similar binding poses only differing by the position of the pyrrolidine ring carrying the chiral center, as shown in Figure 1. The inhibitors show the same potency, and their half-maximal inhibitory concentration (IC_{50}) was 0.7 μM. Investigations of the thermodynamics of binding using isothermal titration calorimetry (ITC) showed that they have a similar binding affinity: the Gibbs free energies of binding (ΔG) of 568R and 568S were −8.1 and −7.9 kcal mol^{-1}, respectively, but their enthalpy and entropy components differed, where the entropic contribution to ΔG was greater for 568R.[5] In the present study, we investigated if these noncovalent inhibitors with, from a drug perspective, a relatively low binding affinity for *m*AChE influence the overall dynamics of the complexes. If the inhibitors do influence the dynamics, do the differences in the enthalpy and entropy of binding translate into dynamical differences in AChE−inhibitor complexes?

Received: June 8, 2018
Revised: August 15, 2018
Published: August 15, 2018

DOI: 10.1021/acs.jpcb.8b05485
J. Phys. Chem. B 2018, 122, 8516−8525

The Journal of Physical Chemistry B Article

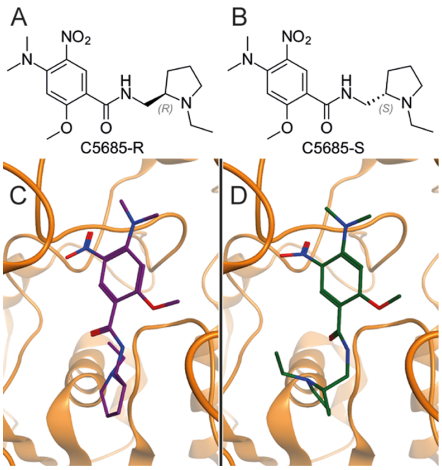


Figure 1. Compounds C5685-R (A) and C5685-S (B) and the enzyme *m*AChE showing the binding conformation of the compounds C5685-R, PDB code: 4ARA (C) and C5685-S, PDB code: 4ARB (D).

The influence of inhibitors on the dynamics of AChE has been evaluated by several different methodologies. Studies of the dynamics of AChE using molecular dynamics (MD) simulations showed that the local dynamics of the inhibitor-binding cavity can change upon binding to different inhibitors; some amino acid residues and substructures of inhibitors became more rigid,[6,7] whereas other molecular fragments, such as Phe330 (*Torpedo californica* AChE; *tc*AChE), displayed increased dynamics.[8] A recent study of the drug donepezil binding to *tc*AChE using microsecond long MD simulations showed that donepezil modulated the movements of AChE and affected the global dynamics.[9] A study using fluorescence spectroscopy to investigate huperzine A binding to *tc*AChE also indicated local changes in the dynamics of amino acid residues close to the inhibitor, but no changes in the global dynamics could be detected.[10] Using methods such as neutron scattering that specifically addresses the averaged atomic dynamics has shown that the changes in the AChE dynamics may be inhibitor-dependent. Huperzine A has been shown to not affect[11] or possibly increase[12,13] the dynamics of human AChE (*h*AChE), whereas the covalent inhibitor soman decreased its dynamics.[13] The same ligand-dependent trend has been observed for other protein−inhibitor complexes using nuclear magnetic resonance (NMR) or neutron scattering, where complex formation led to increased[14−16] or decreased dynamics.[17,18]

In this study, we have investigated the dynamics of atomic level motions in *m*AChE in complex with 568R and 568S, respectively, using elastic incoherent neutron scattering (EINS)[19] that measured the motions of hydrogens in a hydrated protein (or protein−inhibitor complex) powder sample. The fact that the neutron scattering cross section of hydrogen atoms is much larger than that of other atoms present in biological systems[20] leads to measurements that show the motions of hydrogens and indirectly the atoms to which hydrogens are bound in a sample. Enzymatic activities occur on a micro- to millisecond time scale, and there is support for a correlation between the catalytic activity and MD at a nanosecond time scale within the family of human cholinesterases.[21] To evaluate the EINS measurements of the AChE dynamics on the pico- to nanosecond time scale, the classical univariate analysis methods such as the sum of the intensities and the mean-square displacement (MSD)[22] and a newly developed model to derive the mean-square position fluctuations (MSPFs)[23] were used. In addition to these methods, we have recently presented and used multivariate methods including principal component analysis (PCA)[24,25] and orthogonal partial least-squares discriminant analysis (OPLS-DA)[26,27] to analyze the EINS data.[16] Here, the univariate and multivariate methods complemented each other and allowed for a detailed analysis of all available scattering angles (or momentum transfers), leading to statistically sound conclusions.

EXPERIMENTAL SECTION

Sample Preparation. *m*AChE was expressed as previously reported; the protein was harvested from the supernatant and purified using affinity chromatography.[28] Multiple batches of the enzyme were pooled, concentrated, and dialyzed in 20 mM ammonium acetate. *m*AChE (7.4 μmol) in an ammonium acetate/deuterated water (D_2O) buffer was split in three. Inhibitors C5685-S and C5685-R (Figure 1) were dissolved in D_2O, and 2.2 (mole) equivalents of each inhibitor were individually mixed with one of the AChE splits giving samples "*m*AChE-568S" and "*m*AChE-568R", respectively. The third AChE solution was mixed with only D_2O, yielding the inhibitor-free sample "*m*AChE". The three samples were each transferred to a flat neutron scattering sample holder (3 × 4 cm^2) of 1 mm thickness and was dried under vacuum twice until completely dry, monitoring the change in the sample weight. The three samples were rehydrated with D_2O to a total proportion of 0.4 g g^{-1}. The level of hydration allows for the monitoring of motion of the hydrogens in the sample, and it prevented whole enzyme rotation and translation in the samples, focusing measurements on internal protein dynamics.

Elastic Incoherent Neutron Scattering. The enzyme, in the presence or not of an inhibitor, was probed elastically by incoherent neutron scattering. We used two different spectrometers at the Institut Laue Langevin (ILL), France: the cold neutron time-of-flight spectrometer IN6[29] and the high-resolution backscattering IN16.[30] The experiments were done consecutively using the same samples. The highest temperature reached was around 310 K, corresponding to the mammalian body temperature; therefore, no denaturation of *m*AChE was expected. Each spectrometer has its specific energy resolution that corresponds to a characteristic time window, which can be evaluated through Heisenberg's uncertainty principle. IN16 permits to probe movements up to about 1 ns and thus to follow longer diffusional motions, and these motions are associated with torsional vibrations of the buried molecular subgroups corresponding to motional length scales between 3 and 30 Å. IN6 monitors short times up to 20 ps including fast local motions of small molecules or movements of functional groups such as the rotation of hydrogen atoms around carbon in a methyl group. IN6 permits to measure intensities at *Q*-values corresponding to length scales between 3 and 16 Å. The recorded data are corrected for the background through subtraction of the spectrum of an empty sample holder and normalized for a vanadium spectrum, which accounts for the detector efficiencies. Sample transmission was measured on the backscattering spectrometer IN13 at ILL,[31] which is equipped with two monitors to

DOI: 10.1021/acs.jpcb.8b05485
J. Phys. Chem. B 2018, 122, 8516−8525

The Journal of Physical Chemistry B Article

perform such measurements. As it was above 90% for all the samples, we neglected multiple scattering effects.

Univariate Data Analysis. Elastic scattering designates neutron intensities which are collected as a function of the scattering angle for a given energy resolution, ΔE, of the instrument without energy exchange between the neutron and the target. According to Bragg's equation, it is possible to convert the scattering angle into a momentum transfer Q, corresponding to the difference between the incoming and the outgoing wave vector of the neutron in the units of $\hbar$, as the used neutron wavelength is known, and only the direction of the neutron changes during the scattering process. It can be shown that the elastic neutron intensity S_{el} can be described within the Gaussian approximation,[32] where it is assumed that the atomic nuclei undergo harmonic motions around their equilibrium positions, through the equation

$$S_{el}(Q, \omega = 0 \pm \Delta E) \approx S_0 \exp\left(-\frac{1}{3}u^2Q^2\right) \quad (1)$$

where ω is the energy transfer in the units of $\hbar$ and $\langle u^2 \rangle$ are the MSDs. They describe the averaged amplitudes of motion and thus the flexibility of the sample, for example, at a given temperature. The MSD can be obtained quantitatively for each temperature by the slope of the semilogarithmic plot of the incoherent scattering function through the equation

$$u^2 \approx -3\frac{\mathrm{d}\ln S_{el}(Q, \omega = 0 \pm \Delta E)}{\mathrm{d}Q^2} \quad (2)$$

The Gaussian approximation is strictly valid for $Q \to 0$, and it holds up to $\langle u^2 \rangle\ Q_{max}{}^2 \approx 1$, restricting the Q-range that can be used for this type of analysis considerably. We determined the Gaussian approximation and thus MSD to be valid for the Q-ranges of 0.5−1.2 Å^{-1} on IN6 and 0.4−1.2 Å^{-1} on IN16, respectively.

As the scattering function S_{el} is not necessarily Gaussian, other univariate approaches going beyond this approximation were developed in the past. The first one was proposed by Doster et al.[33] describing the motional anharmonicity in the sample by a double-well potential. The Gaussian approximation uses only the first term of a cumulant expansion; thus, Yi et al.[34] included the second cumulant term to describe the standard deviation of the MSD. Another possibility consists of modeling the heterogeneity by different distributions: Nakagawa et al.[35] compared a bimodal, exponential, and Gaussian distribution. Meinhold et al.[36] used a Weibull distribution and Peters and Kneller[23] a Gamma distribution.[37] Most recently, a bimodal Gaussian distribution was introduced by Vural et al.[38] All of these methods represent interesting aspects and permit a better although more complex description of the data. In a representative manner, we applied the model recently developed by Peters and Kneller[23] that, unlike the MSD, permits the inclusion of data over the entire Q-range. Although within the Gaussian approximation only one atomic motion is representative for all hydrogens, this approach takes into account the motional heterogeneity of the amino acid side chains and their environment. The motional heterogeneity of the hydrogen atoms is described by a Gamma distribution, and we designated them as the neutron-weighted average of the atomic "MSPFs". The corresponding elastic intensity can be calculated analytically according to

$$S_{el}^{MSPF}(Q, \beta) \approx \frac{1}{\left(1 + \frac{Q^2\mathrm{MSPF}^2}{\beta}\right)^{\beta}} \quad (3)$$

where β accounts for the homogeneity in the atomic motions, for example, for $\beta \to \infty$, the Gaussian form is retrieved. As the two approaches rely on different approximations, the absolute values of the MSPF are different from the values of MSD. Finally, it is also possible to sum up the neutron intensities over all accessible scattering angles. This summed intensity has smaller error bars as it refers to a higher statistics and is inversely proportional to the square root of the MSD;[13] it is not limited to a certain Q-range. The summed intensities and the MSPF were calculated for the Q-range 0.3−2 Å^{-1} on IN6 and 0.4−1.7 Å^{-1} on IN16, respectively.

Multivariate Data Analysis. The procedure we used to analyze the EINS data with PCA[24,25] and OPLS-DA[26,27] followed a previously described protocol.[16] Briefly, the measured neutron intensities were collected as data points (values of the intensities) in a data matrix where matrix row vectors were the measured intensities at incrementally increasing temperatures and column vectors show at which Q-values at the detector (Q:s) the intensities were collected. This yielded the data matrix $\mathbf{X}(Q_i, T_j)$, where $i = 1, ..., M$ is the number of Q:s and $j = 1, ..., N$ is the number of different measurement temperatures. The intensities were normalized to the intensity measured at the lowest temperature (40 K) and were mean-centered. PCA[24,25] is a projection method that extracts the main variation in the data by reducing a multidimensional set of correlated variables (here Q:s) to a smaller set of variables, the so-called principal components (PC:s). The decomposition matrix containing principal components (PC:s) was calculated according to

$$\mathbf{X} = 1 \times \bar{\mathbf{X}} + \mathbf{S}\mathbf{L}' + \mathbf{E} \quad (4)$$

where $\bar{\mathbf{X}}$ is the $\mathbf{X}$ matrix average, $\mathbf{S}$ is the score matrix, $\mathbf{L}$ is the loading matrix, and $\mathbf{E}$ is the residual matrix. Loading values (l) and score values (s) are analyzed in two separate plots where the loading plot shows the Q:s and their weights on the PCs and the score plot shows how the EINS measurements differ in their intensities in the Q-range.

OPLS-DA[26,27] uses a user-defined matrix ($\mathbf{Y}$) containing class assignments for each measurement based on the sample identity. This information is used to decompose $\mathbf{X}$ containing the measured Q intensities in such a way that it describes the largest variation in $\mathbf{X}$ that correlates with $\mathbf{Y}$ according to

$$\mathbf{X} = \mathbf{S}_p\mathbf{L}_p' + \mathbf{S}_o\mathbf{L}_o' + \mathbf{E} \quad (5)$$

where $\mathbf{S}_p$ is the predictive score matrix, $\mathbf{L}_p'$ is the predictive loading matrix, $\mathbf{S}_o$ is the corresponding $\mathbf{Y}$-orthogonal score matrix, $\mathbf{L}_o'$ is the loading matrix of the $\mathbf{Y}$-orthogonal components, and $\mathbf{E}$ is the residual matrix of $\mathbf{X}$. OPLS-DA aims to discriminate between class differences, shown by the predictive scores ($s_{p,1}$), and differences within a class, determined by the orthogonal scores ($s_{o,1}$). The predictive loadings ($l_{p,1}$), or the weights (w^*) that describe the correlation between the $\mathbf{X}$ variables and the $\mathbf{Y}$ scores that we present here, can be analyzed to identify at which Q:s, that is, the inverse of the movement length scale, the intensities separate the classes.

The number of significant PC:s or OPLS-DA components and the internal prediction capacity of the models were determined by a leave-n-out cross-validation[39] (n equals 1/7 of

DOI: 10.1021/acs.jpcb.8b05485
J. Phys. Chem. B 2018, 122, 8516−8525

The Journal of Physical Chemistry B Article

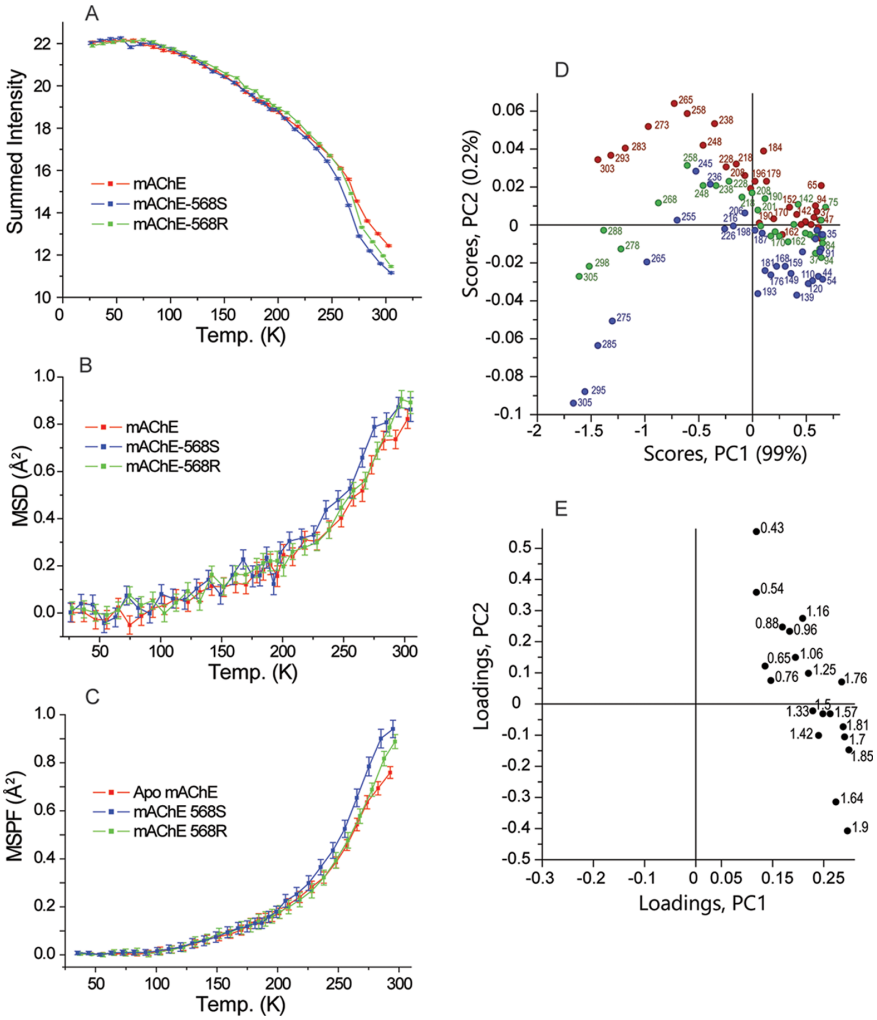


Figure 2. Neutron intensities from the IN16 measurements as arbitrary units summed over the available scattering angles (A), as the extracted MSD (B), and as the extracted MSPF (C). The error bars were based on Poisson statistics applied to the raw data and the summed intensities and linear regression to the fit when extracting the MSD from eq 2 and the MSPF from eq 3. PCA of the measurements from IN16 showing the score plot (*D*) of the samples displaying the main variation in sample neutron intensities (percent described variation within parenthesis) and in loading plots (*E*) with loading values for each variable *Q* indicated with the *Q* value in $Å^{-1}$. Samples are color-coded: *m*AChE (red), *m*AChE-568R (green), and *m*AChE-568S (blue). PC: principal component.

the data set) that gave a cumulative q^2 comprising all model components. q^2 is based on the prediction error sum of squares, which is the squared differences between the actual values and the estimated values for the data kept out of the model fitting. A significant PC in PCA had a q^2 higher than 0.05−0.06 (the limit increases with the number of PC:*s*) and a significant OPLS-DA component had a $q^2 > 0.01$. The validity of the OPLS-DA models was assessed with a permutation test to rule out that the models were a result of chance correlations. In short, the class assignments were randomized 50 times, and new OPLS-DA models were calculated for each new **Y**-matrix. These new models were validated based on $R^2\mathbf{Y}$ and q^2 and compared to the original model. The validity was further assessed by the size and statistical significance of the OPLS-DA coefficients according to Jack-knifing, where a high proportion of significant coefficients was indicative of a significant model. Multivariate modeling was performed using the SIMCA software.[40]

RESULTS

The dynamics of the enzyme *m*AChE in its free form and when in complex with inhibitor 568R or 568S (Figure 1) was measured on the instruments IN6[29] and IN16[30] at temperatures from 20 K to around 310 K. The resulting EINS data were analyzed by extracting the intensities from the instrument-specific scattering angle range, that is, the whole *Q*-range. The intensities exponentially decrease with the dynamics; thus, a low intensity is indicative of more motions in the enzyme.

Dynamics at Large Time Scales. IN16 monitors small-amplitude motions up to large motions such as torsional motions of amino acid side chains[41] within a 1 ns time window. The results from this instrument are shown in Figure

DOI: 10.1021/acs.jpcb.8b05485
J. Phys. Chem. B 2018, 122, 8516−8525

The Journal of Physical Chemistry B Article

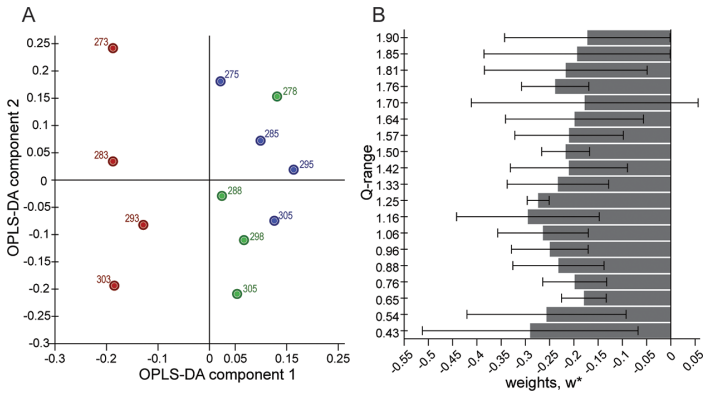


Figure 3. Classification of inhibitor-free vs inhibited *m*AChE in the temperature interval 270−305 K with OPLS-DA score components (A), showing class separations of samples *m*AChE (red), *m*AChE-568R (green), and *m*AChE-568S (blue), and weights (B) for OPLS-DA component 1, showing scattering angle (*Q*-range) contributions to class separation, from the measurements on IN16. Error bars are standard error with a confidence level of 95% calculated by Jack-knifing.

2, and, as expected, the motions in the *m*AChE samples increased with temperature. This can be concluded from the summed intensities in Figure 2A, where the intensities diminish with the increasing temperature. The measurements also revealed that the free form of *m*AChE was less dynamic than the inhibited form at higher temperatures, whereas no difference could be seen at low and mid-temperatures. Above ∼250 K, the samples with *m*AChE−inhibitor had lower summed intensities, and above the melting point (273 K), both showed more dynamics according to the MSD (Figure 2B) and MSPF (Figure 2C) analyses compared to the *m*AChE sample.

The same data were analyzed with PCA giving a two PC model that captured >99% of the original variation in the data (PC1: 99%, PC2: 0.2%); complete model statistics are presented in the Supporting Information (Table S1). That the inhibited *m*AChE was more dynamic was corroborated by the PCA. In the score plot (Figure 2D), the *m*AChE−inhibitor samples are positioned further to the left along the PC1 axis for each comparable measurement (temperature) compared to the free *m*AChE samples. The interpretation of this, based on the loading plot (Figure 2E), is that the inhibited samples had lower intensities (more dynamics) in the low *Q*-range (corresponding to larger amplitudes as, e.g., side-chain movements) and especially at temperatures above the melting point. The differences between the free and inhibited form of *m*AChE are also apparent along the PC2 axis where the noninhibited *m*AChE generally has higher score values (Figure 2D). At around 250 K, the inhibited samples start to deviate from the noninhibited ones in both PC1 and PC2 (Figure 2D). The small differences seen in PC2 based on 0.2% of the measured data are consistent with the differences seen in the summed intensities above 250 K. The effect of melting and thus increased dynamics can be seen in the interval 265−285 K as increasing distances between the measurements of each sample in the score plot in Figure 2D.

The measurement of the three samples also revealed that there was a difference between the two inhibitors in terms of dynamics; compound 568S seemed to induce more motion compared to 568R within the time scale up to 1 ns when in complex with *m*AChE. The measured data analyzed as summed intensities, MSD, and MSPF all showed that the dynamics in the *m*AChE-568S started to exceed that of the two other samples above 200 K, whereas the dynamics of the *m*AChE-568R sample was higher than that of the noninhibited *m*AChE sample from 280 K and upward. This indicates that *m*AChE-568S undergoes a dynamical transition at lower temperatures, meaning that the complex had a lower energy barrier at lower temperature. The barrier determines the onset of anharmonic motions where atoms no longer are vibrating around single equilibrium positions but can "jump" between states. We will come back to this point in the Discussion section. The summed intensities, the MSPF and the PCA showed that the samples were different at high temperatures; MSD did not show a difference at the highest temperature. An inspection of the raw data revealed that the differences between the two inhibited complexes were present even at the highest temperature, but it was not noticeable in the MSD analysis because of the included reduced *Q*-ranges.

The univariate analysis and the PCA showed that the differences between the inhibited and free *m*AChE arose at temperatures close to the melting point. To investigate this further, we classified the measurements above the melting point where the inhibited samples were assumed to belong to one class and the free *m*AChE to the other. Analysis of the classification using OPLS-DA showed that the variation in the *Q* intensities that related to class separation (the $R^2\mathbf{X}$ of the OPLS-DA component 1) was 41% of the total variation of the measured data. Hence, a substantial part of the EINS data was related to the differences between inhibited samples and the free *m*AChE sample (multivariate model statistics and permutation tests are shown in the Supporting Information in Table S1 and Figure S1, respectively). The remaining variation of the data was related to intraclass differences. This is shown in Figure 3A where the score plot shows the separation between the two classes on the OPLS-DA component 1 axis and the differences within the class on the OPLS-DA component 2 axis. All the weights displayed in Figure 3B were positively correlated with the *m*AChE sample on the OPLS-DA component 1 axis, demonstrating that the free *m*AChE had higher *Q*-intensities over the whole *Q*-range and was thus less dynamic compared to the inhibited enzyme in the investigated temperature range.

Dynamics at Short Time Scales Describing Small Atomic Motions. IN6 allows for the detection of small-

DOI: 10.1021/acs.jpcb.8b05485
J. Phys. Chem. B 2018, 122, 8516−8525

The Journal of Physical Chemistry B Article

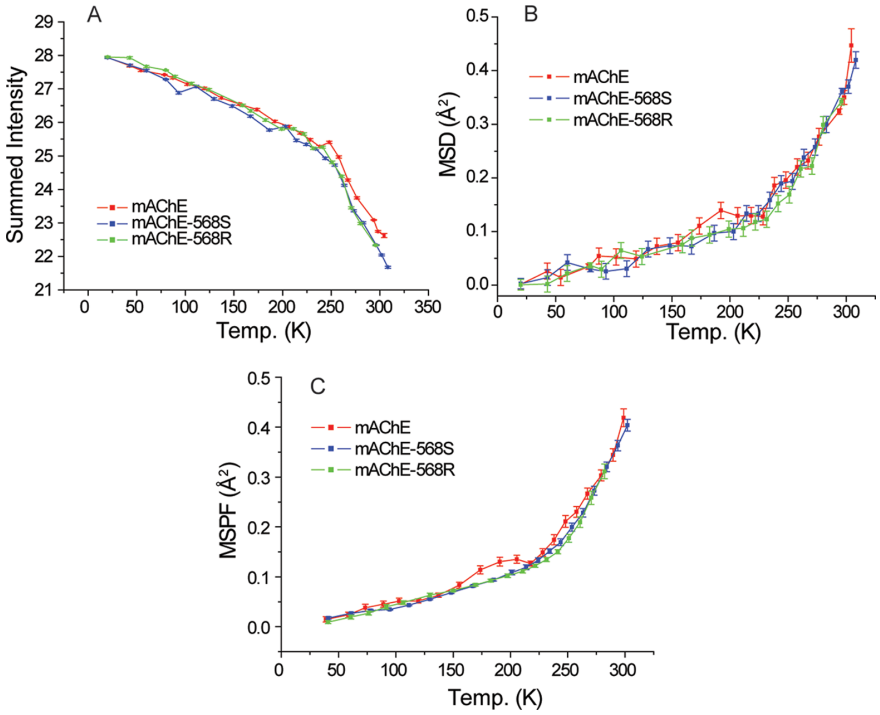


Figure 4. Neutron intensities from the IN6 measurements as arbitrary units summed over the available scattering angles (A), as the extracted MSD (B), and as the extracted MSPF (C). The error bars were based on Poisson statistics applied to the raw data, the summed intensities, and linear regression to the fit when extracting the MSD from eq 2 and the MSPF from eq 3.

amplitude motions such as rotations of hydrogen atoms around the carbon atoms in the methyl groups. Similar to the IN16 results, the summed intensities of the IN6 measurements (Figure 4A) showed that the small-amplitude motions increased with temperature and that the inhibited *m*AChE was slightly more dynamic compared to the free *m*AChE at temperatures above 250 K. Contrary to the analysis of the IN16 data, no differences could be observed between the inhibited and free *m*AChE or between the two inhibited *m*AChE samples based on the MSD or MSPF analyses (Figure 4B,C). The summed intensities have stronger statistics, and subtle differences between measurements might be hidden within error bars in the MSD and MSPF models. Because of beam time restrictions, the *m*AChE-568R sample was not measured at the two highest temperatures (310 and 315 K) but still seems to follow the trend of *m*AChE-568S. Notably, in the free *m*AChE sample, we detected weak traces of bulk water. In confined conditions, water is frozen into a solid amorphous structure when sufficiently quickly quenched. Upon heating, the amorphous structure evolves through a highly defective cubic-like crystal to the stable crystalline form of hexagonal symmetry ice (I_h; see the Supporting Information for details). This structural evolution of water has left fingerprints in the MSD and MSPF analyses (Figure 4B,C) as an increase of the signal at above 150 K, marking the amorphous to cubic ice transformation, and a decline above 200 K, at which the stable structure ice I_h is progressively formed. The signal of the amorphous structure is not noticeable in the applied approach, as it has been canceled out by the normalization of the data to the base temperature.

The IN6 data were analyzed with PCA, and the results are shown in Figure 5 (complete statistics are shown in the Supporting Information, Table S2). The PCA model, with three significant PC:*s* describing 99% of the total variation in the data (PC1: 98%, PC2: 0.7%, and PC3: 0.3%), showed that the dynamics increased at increasing temperatures (PC1), that there was a water effect in the measurements (PC2), and that there was a difference in dynamics between the free and inhibited forms of *m*AChE (PC3). The three samples were similar up to a temperature of around 240 K, illustrated by the similar positions of these measurements in the score plot (Figure 5A,B). At the highest temperatures, the inhibited samples were more dynamic shown by their lower score values in PC1 at each measurement in this temperature interval (i.e., they had low intensities in the *Q*:*s*) compared to the free *m*AChE. PC3 (Figure 5C) shows that the free *m*AChE is in fact different from the inhibited *m*AChE across all temperatures, which is subtle information contained in a few *Q*:*s* at the extremes of PC3 in the loading plot (Figure 5D). Between 260 and 270 K, there is a jump from high to low scores, which is assumingly related to the melting of water. This shift in the scores is visible in all PC:*s* but mostly in that PC2 is dominated by the cubic water phase transition at 240 K, primarily apparent in the free *m*AChE and *m*AChE-568R samples. This event is clearly associated with the *Q*:*s* around 1.59−1.81 $Å^{-1}$ (demonstrated by their high loadings in Figure 5B) and matches with the values for ice Bragg peaks (see the Supporting Information, Table S3) and have been reported before.[16,42] To summarize the PCA for the IN6 measurements, the largest variation in the data was the temperature trend, the second largest represents the ice Bragg reflections around 1.7

DOI: 10.1021/acs.jpcb.8b05485
J. Phys. Chem. B 2018, 122, 8516−8525

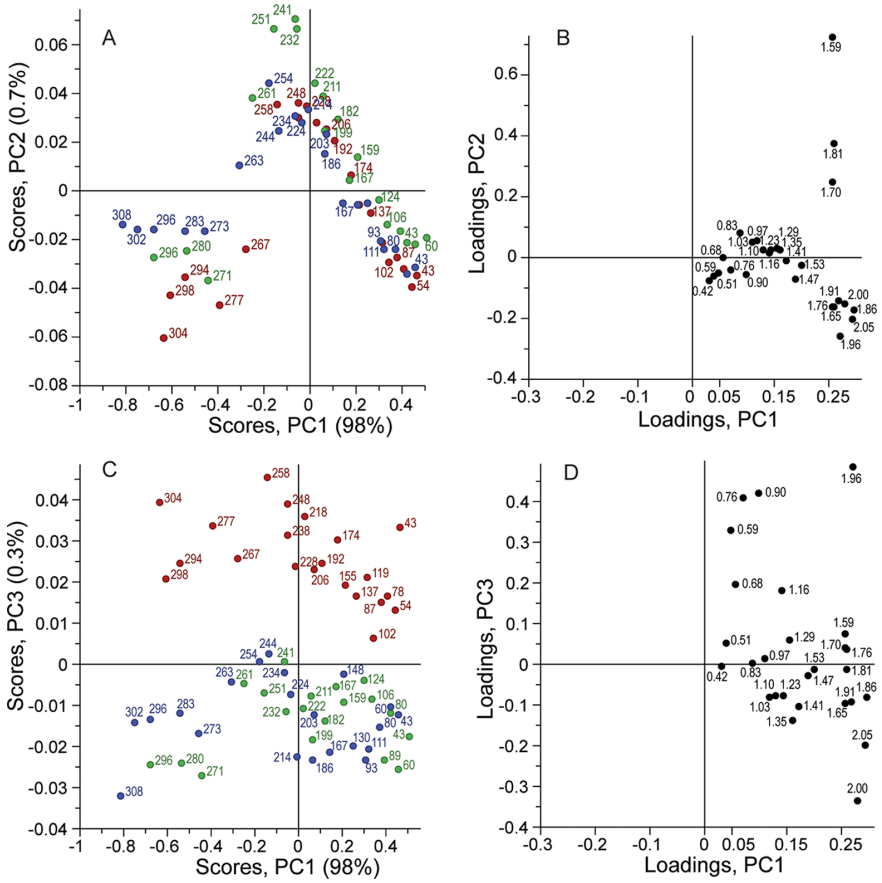


Figure 5. PCA of the measurements from IN6 showing the score plots (A,C) of the samples displaying the main differences in neutron intensities (percent described variation within parenthesis) and in loading plots (B,D) with loading values for each variable Q indicated with the Q value in $Å^{-1}$. Samples are color-coded: *m*AChE (red), *m*AChE-S68R (green), and *m*AChE-S68S (blue). PC: principal component.

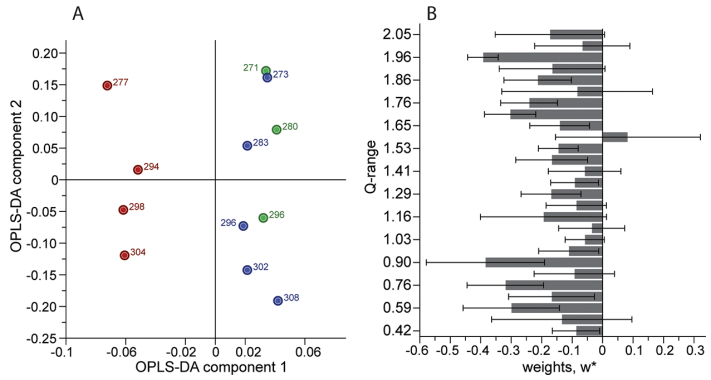


Figure 6. Classification of free vs inhibited *m*AChE based on the IN6 measurements in the temperature interval 270−305 K. (A) OPLS-DA score components, showing separations of the *m*AChE (red) class, the *m*AChE-S68R (green), and *m*AChE-S68S (blue) class. (B) OPLS-DA weights for score component 1, showing scattering angle (Q-range) contributions to class separation. Error bars are standard error with a confidence level of 95% calculated by Jack-knifing.

$Å^{-1}$ and the melting of water, and the third largest showed additional subtle difference between free and inhibited *m*AChE.

The OPLS-DA classification model for the IN6 measurements was deemed statistically significant and confirmed the small differences in the dynamics between free and inhibited *m*AChE (model statistics and permutation tests are shown in

DOI: 10.1021/acs.jpcb.8b05485
J. Phys. Chem. B 2018, 122, 8516−8525

The Journal of Physical Chemistry B Article

the Supporting Information, Table S2 and Figure S2, respectively). The class separation is shown in the OPLS-DA score plot (Figure 6A), and the negative sign of the weights showed that the inhibited *m*AChE was more dynamic compared to the free *m*AChE above the melting point (Figure 6B). Thus, the multivariate methods confirmed that there were differences in the dynamics between the inhibited and free *m*AChE samples at high temperatures, although the difference is very small. On this short time scale, up to 20 ps, the multivariate method could differentiate between the samples and show subtle differences.

DISCUSSION

The objective of this study was to compare the effects of binding by two enantiomeric, noncovalent inhibitors on the MD of the enzyme *m*AChE. We investigated it by EINS on two different spectrometers of the ILL, giving access to various time windows and sizes of atomic motions. We anticipated that the effects on the enzyme dynamics of these small, noncovalent, relatively weak inhibitors might be minor; nevertheless, we could clearly identify the differences between the inhibited and free enzyme. A fruitful combination of univariate (Gaussian approximation and a non-Gaussian model permitting to exploit the whole *Q*-range) and multivariate methods (PCA and OPLS-DA) was crucial to analyze the fine details of the data.

AChE Becomes More Dynamic When Inhibitors Are Bound. The collective conclusion from both instruments is that the inhibited *m*AChE became more dynamic compared to the noninhibited *m*AChE, especially at higher temperatures. This was deduced from the summed intensities and the multivariate modeling of the data resulting from the measurements on IN16 and IN6 and the MSD and MSPF based on the data from IN16. Previous EINS measurements have shown that *m*AChE is less dynamic overall compared to the human form,[19] though they are structurally similar (~89% sequence identity). In our study, *m*AChE showed less overall dynamics compared to a previous study,[19] but remarkably, the changes in dynamics upon inhibitor binding is evident in the experimental data, mainly from the IN16 measurements. Thus, even though 568R and 568S are noncovalent inhibitors with a modest potency, they do affect the dynamics of the enzyme. Increased dynamics was generally apparent at temperatures above water melting, which might indicate that liquid water may help mediate the motions induced by the inhibitors. Huperzine A, which is a strong inhibitor of human AChE (the inhibitory constant, K_i, of 40 nM),[43] has been evaluated for its effect on AChE dynamics using EINS.[13] Its inhibition led to a small but significant increase in dynamics shown by the MPSF model[13] of measurements on the instrument IN13.[31] Similar to our study, the effect on dynamics was mainly observed above the melting temperature of water and especially at temperatures around 300 K. The extremely potent noncovalent inhibitor syn-TZ2PA6 with a dissociation constant (K_d) of 0.41 pM[44] in complex with *m*AChE has been investigated using MD simulation.[6] This study showed that the *m*AChE−inhibitor complex displayed smaller structural fluctuations and deviations, especially in the active site, compared to the noninhibited enzyme. Notably, 27 residues on the surface of the enzyme unexpectedly showed increased dynamics,[6] but the limited simulation time and lack of reproduced data make it unclear for us as to whether these effects can be translated into overall dynamical changes.

Small motions (e.g., methyl hydrogen rotations) at short time scales (up to 20 ps) specifically monitored on IN6 up to larger motions (side chains or domains) up to about 1 ns monitored by IN16 were all affected by inhibition. Thus, the combination of several spectrometers gives the most complete picture of MD. Concerning the small motions, the main difference between inhibited and noninhibited dynamics was seen in the lowest and highest *Q*:*s* in the *Q*-range on IN6, seen by the loadings in Figure 5D and weights in Figure 6B. This indicates that both the largest and the smallest amplitude motions detected within the IN6 time window increased in the inhibited *m*AChE. Translation of water and D_2O molecules would be monitored at *Q*:*s* below 1.5 $Å^{-1}$, but the short 20 ps time window of IN6 would not fully capture this, so the motions are of enzyme origin (although they can be affected by the motions of water). On IN16, differences emerged at temperatures around 240 K where the inhibited enzyme, especially by 568S, displayed higher intensities across the whole *Q*-range, that is, an increased dynamics in all amplitudes monitored at the 1 ns time window. These results show that the motions of amino acid side chains and even loop movements also increased. AChE has been suggested to exhibit a "breathing" motion including an opening and a contraction of the middle region of the active site gorge.[9,45] Inhibitors have been shown by MD simulations to reduce this motion and to stabilize the gorge in an open position,[9] as well as the motions of catalytic site residues.[6] Our data and the experimental results of others suggest an inhibitor-induced increase in the overall motion of AChE, and future MD simulations of our inhibitors and *m*AChE may shed light on the structural origin of these motions.

Magnitude of Increased Enzyme Dynamics Is Inhibitor-Dependent. The inhibitor 568S induced more dynamics at the 1 ns time scale and underwent dynamical transition at about 10 K earlier compared to the inhibitor 568R, shown by the fact that *m*AChE-568S changes slope direction (summed intensities on IN16) at an earlier temperature than *m*AChE-568R and noninhibited *m*AChE. The summed intensities, PCA, MSD, and MSPF, at the 1 ns time scale all show that the *m*AChE-568S complex was the most dynamic at temperatures from around 250 K and upward.

Inhibition Alters the Water Phase Transition Temperatures. When analyzing the IN6 data by PCA, we could identify a water phase transition at around 240 K (Figure 5A). This corresponded to ice Bragg reflections, especially apparent for 568R, but too subtle to show up in the univariate analysis. The effect was associated with *Q*:*s* around 1.7 $Å^{-1}$. Phase transitions, including water melting, may be spotted in the summed intensities as a change in the direction of the slope. In biological systems, water can exist in different solid structures when cooled strongly below the freezing point. Above 200 K, the water may adopt the ordinary water ice of hexagonal symmetry up until the melting point at around 270 K (see the Supporting Information for a detailed description). In a recent study by us, the enzyme chymotrypsin in its uninhibited form showed a clear increase in intensity in between 220 and 270 K in measurements on IN6, which was associated with water Bragg peaks.[16] Thus, this phenomenon seems to be present in all measurements but to a varying degree and not always detectable by the univariate analysis.

Differences in Dynamics in the Complexes Are Not Related to the Entropy of Binding. The previous ITC experiments to derive the ΔG of binding of the two inhibitors

DOI: 10.1021/acs.jpcb.8b05485
J. Phys. Chem. B 2018, 122, 8516−8525

The Journal of Physical Chemistry B Article

to *m*AChE showed that the binding was both enthalpically and entropically favorable. 568R had a larger entropic contribution ($-T\Delta S$) to the ΔG of −2.5 kcal mol^{-1} compared to 568S of −0.4 kcal mol^{-1}, whereas 568S displayed a larger enthalpy of binding −7.5 kcal mol^{-1} compared to −5.6 kcal mol^{-1} for 568R.[5] A larger $-T\Delta S$ can be interpreted as an increase in entropy in the system upon complex formation, possibly including increased protein motions. Furthermore, crystallographic data of the complexes and quantum mechanical calculations showed that 568R, when binding to *m*AChE, had more conformations compared to 568S and that these additional binding conformations showed fewer noncovalent interactions, possibly reflected in the lower enthalpy contribution to ΔG.[5] Thus, the ITC experiments and the crystallographic data indicated that *m*AChE-568R might be more dynamic than *m*AChE-568S. However, the EINS data presented here investigating the total averaged dynamics of the two complexes showed that the *m*AChE-568S complex was more dynamic than *m*AChE-568R. It appears as the inhibition of *m*AChE by 568R and especially 568S led to an increase in dynamics in the enzyme, whereas the inhibitor 568R in itself is more flexible than 568S in complex with *m*AChE. One hypothesis is that an inhibitor with strong enthalpy contributions such as 568S (i.e., involved in hydrogen bonds, salt bridges, etc., to the protein) will disturb/perturb the motions of the enzyme and thus result in increased total average motions and increased dynamics.

CONCLUSIONS

The effects on the enzyme dynamics of two enantiomeric noncovalent inhibitors of *m*AChE were measured using EINS. Both univariate and multivariate data analyses of the measurements showed that the motions of the enzymes increased when bound to an inhibitor; this was true for all kinds of motions measured at different time windows. The effects of increased dynamics were generally seen at temperatures above the melting point of water, indicating that liquid water may help mediate the motions. Whereas the univariate data analysis approaches could show these general tendencies, the multivariate methods permitted a more detailed description highlighting specific information such as the appearance of Bragg peaks or a significant contribution from low or large Q values. This approach is strengthened by the fact that conclusions drawn regarding the observed physical phenomenon do not rely on any single model or assumption about the data and the additional statistical verification of differences between samples. Thus, we have shown that even small and noncovalent inhibitors with a moderate potency from a drug perspective can have measurable effects on the dynamics of AChE. We could also see a difference in the influence on the dynamics between the two inhibitors even though they are enantiomers (mirror images of each other's structures) and have the same binding affinities. An interesting future work would include computer simulations to pinpoint the areas and substructures in the enzyme responsible for the increase in overall dynamics of the enzymes upon inhibitor complex formation.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpcb.8b05485.

Multivariate model statistics and solid water structures at ambient pressure (PDF)

AUTHOR INFORMATION

Corresponding Authors

*E-mail: jpeters@ill.fr (J.P.).
*E-mail: anna.linusson@umu.se (A.L.).

ORCID

C. David Andersson: 0000-0001-8198-1688
Anna Linusson: 0000-0003-0063-8912

Author Contributions

A.L., F.E., A.A., J.P., N.M., and M.M.K. performed the experiments. C.D.A., J.P., D.Z., N.M., and A.L. performed the data analysis. C.D.A., J.P., and A.L. did the major writing, and all authors contributed to the comments on the manuscript. All authors have given approval to the final version of the manuscript.

Funding

This work was funded by the Swedish Research Council (Dnr: 2014-4675).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the ILL for the allocation of beam time. Anna Linusson would like to thank the Swedish Research Council for the financial support.

REFERENCES

(1) Talesa, V. N. Acetylcholinesterase in Alzheimer's disease. *Mech. Ageing Dev.* **2001**, *122*, 1961−1969.
(2) Tan, C.-C.; Yu, J.-T.; Wang, H.-F.; Tan, M.-S.; Meng, X.-F.; Wang, C.; Jiang, T.; Zhu, X.-C.; Tan, L. Efficacy and safety of donepezil, galantamine, rivastigmine, and memantine for the treatment of Alzheimer's disease: A systematic review and meta-analysis. *J. Alzheim. Dis.* **2014**, *41*, 615−631.
(3) Mohammad, D.; Chan, P.; Bradley, J.; Lanctôt, K.; Herrmann, N. Acetylcholinesterase inhibitors for treating dementia symptoms - a safety evaluation. *Expet Opin. Drug Saf.* **2017**, *16*, 1009−1019.
(4) Galimberti, D.; Scarpini, E. Old and new acetylcholinesterase inhibitors for Alzheimer's disease. *Expet Opin. Invest. Drugs* **2016**, *25*, 1181−1187.
(5) Berg, L.; Niemiec, M. S.; Qian, W.; Andersson, C. D.; Wittung-Stafshede, P.; Ekström, F.; Linusson, A. Similar but different: Thermodynamic and structural characterization of a pair of enantiomers binding to acetylcholinesterase. *Angew. Chem., Int. Ed.* **2012**, *51*, 12716−12720.
(6) Senapati, S.; Bui, J. M.; McCammon, J. A. Induced fit in mouse acetylcholinesterase upon binding a femtomolar inhibitor: A molecular dynamics study. *J. Med. Chem.* **2005**, *48*, 8155−8162.
(7) Bennion, B. J.; Essiz, S. G.; Lau, E. Y.; Fattebert, J.-L.; Emigh, A.; Lightstone, F. C. A wrench in the works of human acetylcholinesterase: Soman induced conformational changes revealed by molecular dynamics simulations. *PLoS One* **2015**, *10*, No. e0121092.
(8) Xu, Y.; Colletier, J.-P.; Weik, M.; Jiang, H.; Moult, J.; Silman, I.; Sussman, J. L. Flexibility of aromatic residues in the active-site gorge of acetylcholinesterase: X-ray versus molecular dynamics. *Biophys. J.* **2008**, *95*, 2500−2511.
(9) Cheng, S.; Song, W.; Yuan, X.; Xu, Y. Gorge motions of acetylcholinesterase revealed by microsecond molecular dynamics simulations. *Sci. Rep.* **2017**, *7*, 1−12.
(10) Boyd, A. E.; Dunlop, C. S.; Wong, L.; Radić, Z.; Taylor, P.; Johnson, D. A. Nanosecond dynamics of acetylcholinesterase near the active center gorge. *J. Biol. Chem.* **2004**, *279*, 26612−26618.

DOI: 10.1021/acs.jpcb.8b05485
J. Phys. Chem. B 2018, 122, 8516−8525

(11) Trapp, M.; Trovaslet, M.; Nachon, F.; Koza, M. M.; van Eijck, L.; Hill, F.; Weik, M.; Masson, P.; Tehei, M.; Peters, J. Energy landscapes of human acetylcholinesterase and its huperzine a-inhibited counterpart. *J. Phys. Chem. B* **2012**, *116*, 14744−14753.
(12) Trapp, M.; Tehei, M.; Trovaslet, M.; Nachon, F.; Martinez, N.; Koza, M. M.; Weik, M.; Masson, P.; Peters, J. Correlation of the dynamics of native human acetylcholinesterase and its inhibited huperzine A counterpart from sub-picoseconds to nanoseconds. *J. R. Soc., Interface* **2014**, *11*, 20140372.
(13) Peters, J.; Martinez, N.; Trovaslet, M.; Scannapieco, K.; Koza, M. M.; Masson, P.; Nachon, F. Dynamics of human acetylcholinesterase bound to non-covalent and covalent inhibitors shedding light on changes to the water network structure. *Phys. Chem. Chem. Phys.* **2016**, *18*, 12992−13001.
(14) Yun, S.; Jang, D. S.; Kim, D.-H.; Choi, K. Y.; Lee, H. C. 15N NMR Relaxation Studies of Backbone Dynamics in Free and Steroid-Bound 5-3-Ketosteroid Isomerase from Pseudomonas testosteroni. *Biochemistry* **2001**, *40*, 3967−3973.
(15) Balog, E.; Becker, T.; Oettl, M.; Lechner, R.; Daniel, R.; Finney, J.; Smith, J. C. Direct determination of vibrational density of states change on ligand binding to a protein. *Phys. Rev. Lett.* **2004**, *93*, 028103.
(16) Andersson, C. D.; Martinez, N.; Zeller, D.; Rondahl, S. H.; Koza, M. M.; Frick, B.; Ekström, F.; Peters, J.; Linusson, A. Changes in dynamics of α-chymotrypsin due to covalent inhibitors investigated by elastic incoherent neutron scattering. *Phys. Chem. Chem. Phys.* **2017**, *19*, 25369−25379.
(17) Eletsky, A.; Kienhöfer, A.; Hilvert, D.; Pervushin, K. Investigation of Ligand Binding and Protein Dynamics in Bacillus subtilis Chorismate Mutase by Transverse Relaxation Optimized Spectroscopy−Nuclear Magnetic Resonance. *Biochemistry* **2005**, *44*, 6788−6799.
(18) Miao, Y.; Yi, Z.; Cantrell, C.; Glass, D. C.; Baudry, J.; Jain, N.; Smith, J. C. Coupled flexibility change in cytochrome P450cam substrate binding determined by neutron scattering, NMR, and molecular dynamics simulation. *Biophys. J.* **2012**, *103*, 2167−2176.
(19) Trovaslet, M.; Trapp, M.; Weik, M.; Nachon, F.; Masson, P.; Tehei, M.; Peters, J. Relation between dynamics, activity and thermal stability within the cholinesterase family. *Chem.-Biol. Interact.* **2013**, *203*, 14−18.
(20) Sears, V. F. Neutron scattering lengths and cross sections. *Neutron News* **1992**, *3*, 26−37.
(21) Peters, J.; Trovaslet, M.; Trapp, M.; Nachon, F.; Hill, F.; Royer, E.; Gabel, F.; van Eijck, L.; Masson, P.; Tehei, M. Activity and molecular dynamics relationship within the family of human cholinesterases. *Phys. Chem. Chem. Phys.* **2012**, *14*, 6764−6770.
(22) Lovesey, S. W. *Theory of Neutron Scattering from Condensed Matter*; Oxford Science Publications: Oxford, U.K., 1984.
(23) Peters, J.; Kneller, G. R. Motional heterogeneity in human acetylcholinesterase revealed by a non-Gaussian model for elastic incoherent neutron scattering. *J. Chem. Phys.* **2013**, *139*, 165102.
(24) Wold, S.; Esbensen, K.; Geladi, P. Principal component analysis. *Chemom. Intell. Lab. Syst.* **1987**, *2*, 37−52.
(25) Jackson, J. E. *A User's Guide to Principal Components*; John Wiley & sons, Inc.: New York, USA, 1991.
(26) Trygg, J.; Wold, S. Orthogonal projections to latent structures (O-PLS). *J. Chemom.* **2002**, *16*, 119−128.
(27) Bylesjö, M.; Rantalainen, M.; Cloarec, O.; Nicholson, J. K.; Holmes, E.; Trygg, J. OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. *J. Chemom.* **2006**, *20*, 341−351.
(28) Ekström, F.; Akfur, C.; Tunemalm, A.-K.; Lundberg, S. Structural Changes of Phenylalanine 338 and Histidine 447 Revealed by the Crystal Structures of Tabun-Inhibited Murine Acetylcholinesterase. *Biochemistry* **2006**, *45*, 74−81.
(29) *Cold Neutron Time-Focussing Time-of-Flight Spectrometer IN6-Sharp*; Institut Laue-Langevin: Grenoble, France. https://www.ill.eu/instruments-support/instruments-groups/instruments/in6-sharp/description/instrument-layout/, (accessed Aug 14, 2018).
(30) Frick, B.; Gonzalez, M. Five years operation of the second generation backscattering spectrometer IN16-a retrospective, recent developments and plans. *Phys. B* **2001**, *301*, 8−19.
(31) Francesca, N.; Peters, J.; Russo, D.; Sonvico, F. IN13 backscattering spectrometer at ILL: Looking for motions in biological macromolecules and organisms. *Neutron News* **2008**, *19*, 14−18.
(32) Rahman, A.; Singwi, K. S.; Sjölander, A. Theory of Slow Neutron Scattering by Liquids. I. *Phys. Rev.* **1962**, *126*, 986−996.
(33) Doster, W.; Cusack, S.; Petry, W. Dynamical transition of myoglobin revealed by inelastic neutron scattering. *Nature* **1989**, *337*, 754−756.
(34) Yi, Z.; Miao, Y.; Baudry, J.; Jain, N.; Smith, J. C. Derivation of mean-square displacements for protein dynamics from elastic incoherent neutron scattering. *J. Phys. Chem. B* **2012**, *116*, 5028−5036.
(35) Nakagawa, H.; Kamikubo, H.; Tsukushi, I.; Kanaya, T.; Kataoka, M. Protein dynamical heterogeneity derived from neutron incoherent elastic scattering. *J. Phys. Soc. Jpn.* **2004**, *73*, 491−495.
(36) Meinhold, L.; Clement, D.; Tehei, M.; Daniel, R.; Finney, J. L.; Smith, J. C. Protein dynamics and stability: The distribution of atomic fluctuations in thermophilic and mesophilic dihydrofolate reductase derived using elastic incoherent neutron scattering. *Biophys. J.* **2008**, *94*, 4812−4818.
(37) Kneller, G. R.; Hinsen, K. Quantitative model for the heterogeneity of atomic position fluctuations in proteins: A simulation study. *J. Chem. Phys.* **2009**, *131*, 045104.
(38) Vural, D.; Smith, J. C.; Glyde, H. R. Determination of dynamical heterogeneity from dynamic neutron scattering of proteins. *Biophys. J.* **2018**, *114*, 2397−2407.
(39) Eastment, H. T.; Krzanowski, W. J. Cross-validatory choice of the number of components from a principal component analysis. *Technometrics* **1982**, *24*, 73−77.
(40) *SIMCA*, version 14.1; Sartorius Stedim Biotech (SSB): Otto-Brenner-Straße 20, 37079 Göttingen, Germany, 2015.
(41) McCammon, J. A.; Harvey, S. C. *Dynamics of Proteins and Nuclear Acids*; Cambridge University Press: Cambridge, U.K., 1987.
(42) Koza, M. M.; Geil, B.; Schober, H.; Natali, F. Absence of molecular mobility on nano-second time scales in amorphous ice phases. *Phys. Chem. Chem. Phys.* **2005**, *7*, 1423−1431.
(43) Ashani, Y.; Peggins, J. O.; Doctor, B. P. Mechanism of inhibition of cholinesterases by huperzine A. *Biochem. Biophys. Res. Commun.* **1992**, *184*, 719−726.
(44) Lewis, W. G.; Green, L. G.; Grynszpan, F.; Radić, Z.; Carlier, P. R.; Taylor, P.; Finn, M. G.; Sharpless, K. B. Click chemistry in situ: Acetylcholinesterase as a reaction vessel for the selective assembly of a femtomolar inhibitor from an array of building blocks. *Angew. Chem., Int. Ed.* **2002**, *41*, 1053−1057.
(45) Shen, T.; Tai, K.; Henchman, R. H.; McCammon, J. A. Molecular dynamics of acetylcholinesterase. *Acc. Chem. Res.* **2002**, *35*, 332−340.

DOI: 10.1021/acs.jpcb.8b05485
J. Phys. Chem. B 2018, 122, 8516−8525

B.4. Structural stability of human butyrylcholinesterase under high hydrostatic pressure

To the following publication of Kangur et al. [142] my contribution was

- some help with experiments,
- data treatment to calculate the pressure inside the sample cell.

BBA - Proteins and Proteomics 1867 (2019) 107–113

Contents lists available at ScienceDirect

BBA - Proteins and Proteomics

journal homepage: www.elsevier.com/locate/bbapap

Structural stability of human butyrylcholinesterase under high hydrostatic pressure

Liina Kangur[a], Kõu Timpmann[a], Dominik Zeller[b,c], Patrick Masson[d], Judith Peters[b,c], Arvi Freiberg[a,e,*]

[a] Institute of Physics, University of Tartu, W. Ostwald Str. 1, Tartu 51011, Estonia
[b] Université Grenoble Alpes, CNRS, LiPhy, Grenoble 38000, France
[c] Institut Laue Langevin, 71 avenue des Martyrs, Grenoble Cedex 38042, France
[d] Neuropharmacology Laboratory, Kazan Federal University, Kazan, Russia
[e] Institute of Molecular and Cell Biology, University of Tartu, Riia 23, Tartu 51014, Estonia

ABSTRACT

Human butyrylcholinesterase is a nonspecific enzyme of clinical, pharmacological and toxicological significance. Although the enzyme is relatively stable, its activity is affected by numerous factors, including pressure. In this work, hydrostatic pressure dependence of the intrinsic tryptophan fluorescence in native and salted human butyrylcholinesterase was studied up to the maximum pressure at ambient temperature of about 1200 MPa. A correlated large shift toward long wavelengths and broadening observed at pressures between 200 and 700 MPa was interpreted as due to high pressure-induced denaturation of the protein, leading to an enhanced exposure of tryptophan residues into polar solvent environment. This transient process in native butyrylcholinesterase presumably involves conformational changes of the enzyme at both tertiary and secondary structure levels. Pressure-induced mixing of emitting local indole electronic transitions with quenching charge transfer states likely describes the accompanying fluorescence quenching that reveals different course from spectral changes. All the pressure-induced changes turned irreversible after passing a mid-point pressure of about 400 ± 50 MPa. Addition of either 0.1 M ammonium sulphate (a kosmotropic salt) or 0.1 M lithium thiocyanate (a chaotropic salt) to native enzyme similarly destabilized its structure.

1. Introduction

Butyrylcholinesterase (BChE; EC. 3.1.1.8; P06276) is a nonspecific cholinesterase that hydrolyses a variety of esters. Alternative names for this globular protein are acylcholine acylhydrolase, choline esterase II and pseudocholinesterase. The major form of human BChE is a tetramer comprised of four identical 85 kDa subunits, each containing 574 amino acids and 9 N-linked glycans [1]. The 40 amino acids in the C-terminal side constitute a tetramerization domain where 4 alpha helices are joined by noncovalent interaction with a short polyproline-rich peptide [2–5]. Human plasma BChE is synthesized in the liver. It is present in high amounts in serum, intestine, liver and lung [6]. BChE in human plasma has a concentration of 5 mg/l and a half-life of 12 days [7]. A function of BChE is to detoxify ingested or inhaled ester-containing poisons, such as cocaine [6]. BChE is being developed as a bioscavenger of organophosphates for treatment of acute poisoning [8]. Its known physiological functions are hydrolysis of the neurotransmitter acetylcholine either as surrogate or backup of acetylcholinesterase (AChE) in the central nervous system and at neuromuscular junctions [9–12] as well as hydrolysis of the hunger hormone, octanoyl-ghrelin [13,14]. Furthermore, BChE deficiency and genetic variations of the BChE gene result in delayed metabolism of the muscle relaxants succinylcholine and mivacurium [15,16].

A model of the 3D structure [15] shows four C-terminal alpha helices, one from each BChE monomer, covalently bound through one Cys571-Cys571 bridge per dimer. The four alpha-helix bundle tightly interacts with a short polyproline peptide. The tetrameric BChE structure contains a large number of aromatic amino acids: 4 × 18 Trp residues, 4 × 39 Phe, and 4 × 20 Tyr. In the following, we will pay special attention to Trp residues, as chromophores, which dominate optical absorption/emission spectra of the enzyme around 300 nm. From 18 Trp residues per monomer, 4 belong to the tetramerization domain. In the globular monomer domain there are 14 Trp residues: Trp82, Trp112, Trp231 and Trp430, lining the active site gorge are not exposed to the bulk solvent; Trp96, Trp177, Trp471 and Trp522 are deeply buried into the hydrophobic protein core and the other 6 Trp residues (Trp52; Trp56; Trp376, Trp412; Trp433; Trp490) protrude the protein surface. Being partially solvent-exposed, it is conceivable that specifically these 4 × 6 Trp chromophores (per tetramer), considered as optical probes, must be most sensitive to changes in their local environment.

The denaturing effect of high hydrostatic pressure on the spatial

* Corresponding author at: Institute of Physics, University of Tartu, W. Ostwald Str. 1, Tartu 51011, Estonia.
E-mail address: arvi.freiberg@ut.ee (A. Freiberg).

https://doi.org/10.1016/j.bbapap.2018.11.001
Received 24 July 2018; Received in revised form 2 November 2018; Accepted 5 November 2018
Available online 07 November 2018

structure of human BChE is well documented [17–22]. Electrophoresis and 8-anilino-1-naphthalenesulphonate binding under pressure provided evidence for a transition toward a molten-globule state around 150 MPa [19]. In [20], the protein molten globule was detected by transient increase in protein size between 125 and 150 MPa. Beyond 160 MPa, BChE loses capability to bind reversible inhibitors, and its enzymatic activity is irreversibly lost [23,24]. These former data did not reveal tetramer dissociation under hydrostatic pressure up to 4 kbar. In fact, recent data of molecular modelling, SAXS and cryo-EM [25] show that the tetramer is a dimer of dimers, whereby each dimer is formed of disulfide-bonded monomer. The two dimers interact through glycan chains. The glycan interactions involve numerous H-bonds that are rather insensitive to pressure, unlike hydrophobic interactions that are easily broken by pressure. This probably explains why the protein does not easily dissociate under pressure.

A high-pressure study of the amide I band at 1650 cm^{-1} by Fourier-transform IR (FT-IR) spectroscopy in soman-aged BChE [18] showed that there is no significant change in the protein secondary structure below 3 kbar. However, at higher pressures an irreversible unfolding was observed with the midpoint pressure of denaturation $P_{1/2} = 550 \pm 60$ MPa. The experiment was performed at 25 °C in D_2O of pD = 7 using relatively high protein concentration. Pressure dependence of the Trp emission spectrum in the UV range was studied at the same time, but unfortunately, under quite different conditions in terms of enzymatic state (soman-aged enzyme and native enzyme), solvent (10 mM Tris/HCl buffer pH = 8 in water), temperature (20 °C), enzyme concentration (10 μg/ml for fluorescence measurements versus 10–30 mg/ml for FT-IR), and pressure range (up to 3 kbar for fluorescence and up to 11 kbar for FT-IR). These measurements unexpectedly showed no change in emission spectrum shape upon increasing pressure. At the same time, intensity of the emission decreased rapidly, so that already at 300 MPa no emission could be detected. The fluorescence spectrum excited at 295 nm had a maximum at 338 nm at ambient pressure. Compared with the maximum position of about 325 nm measured in the current work (see below), this suggests a rather disturbed initial protein structure, widely open to polar environment. After pressure release, the emission signal failed to recover to its initial value. The authors suspected pressure-induced aggregation of the proteins as the main source of irreversibility.

Former studies also showed that salts may induce opposite conformational/hydrational shifts of human forms of BChE at the origin of hysteretic catalytic behaviour with certain substrates [26]. Salts that stabilize protein structures (weakly hydrated cations and strongly hydrated anions) are generally called kosmotropes; salts that destabilize/denature proteins (strongly hydrated cations and weakly hydrated anions) are called chaotropes.

In the present work, intrinsic fluorescence of native (plain/unsalted) and salted human BChE selectively excited into the Trp absorption band at 295 ± 1 nm was studied between 310 and 500 nm under hydrostatic pressures varying from atmospheric pressure of 0.1 MPa to 1200 MPa. The maximum applied pressure was determined by pressure-induced gelification of the sample and accompanying loss of hydrostaticity. Significant differences between the data reported in ref. [18] and those of the present study were observed starting already from the fluorescence spectrum measured at ambient pressure.

2. Materials and methods

2.1. Sample preparation

The tetrameric enzyme in phosphate buffered saline was highly purified from Cohn fraction IV-4, an enriched source of pooled human plasma BChE [27,28]. Pure and 100% active human BChE displays a specific activity of 700 units/mg with 1 mM butyrylthiocholine as the substrate at pH 7.0 and 25 °C. Units of activity, defined as micromoles butyrylthiocholine hydrolyzed per min, were measured in 0.1 M potassium phosphate pH 7.0 at 25 °C with 1 mM butyrylthiocholine and 0.5 mM dithiobisnitrobenzoic acid. The human BChE used in the present work was 99% pure and had a specific activity of 500 units/mg. This means that the enzyme preparation was 71% active. As appropriate for pressure measurements [29], the phosphate buffer was exchanged by dialysis to Tris (50 mM, pH 7.5) buffer. The BChE concentration was adjusted to ~5 mg/ml; at 280 nm, 1 mg/ml BChE corresponds to an absorbance of 1.8. To modulate the conformational stability of BChE, ammonium sulphate ($(NH_4)_2SO_4$, a stabilizing kosmotrope) or lithium thiocyanate (LiSCN, a destabilizing chaotrope) was added by dilution of the BChE solution to ~1.7 mg/ml. Optical density of the samples in the high-pressure cell was adjusted to < 0.1 at the excitation wavelength to avoid fluorescence reabsorption effects.

2.2. Spectroscopy

Absorption and CD spectra were measured with a resolution of 1 nm, using a Chirascan Plus spectrophotometer (Applied Photophysics, UK). Fluorescence spectra were recorded with a resolution of 2 nm, using a spectrograph (DTMc300, Bentham Instruments Ltd.) coupled with a CCD camera (iDus 416, Andor Technology). Fluorescence was excited at 295 nm by the third harmonic of a tunable femtosecond pulsed Ti: Sapphire laser with a pulse repetition rate of 76 MHz (Coherent Mira Optima 900-F). The average intensity of the excitation light was kept sufficiently low (~25 mW/cm^2) to avoid non-linear excitation effects. The spectral width of the excitation pulse determined as full width at half maximum (fwhm) was 2 nm. A long pass filter (AHF Analysentechnik LP300) was used to block excitation laser scattering.

2.3. High-pressure measurements

Samples of 0.2–0.3 μl were placed in a thermostatted diamond anvil cell D-02 (Diacell Products Ltd) as described in [30–32]. All measurements were performed at 25 ± 0.1 °C. A stainless steel gasket approximately 0.35 mm thick was used to contain the sample in the cell. Pressure was determined in real time using a ruby-microbead pressure sensor (RSA Le Rubis SA) directly mounted into the sample chamber and excited at 532 nm by a Nd:YAG laser. A thermoelectrically cooled DV420-OE1 CCD camera (Andor Technology) attached to a 1.5 m Jobin-Yvon TH150 spectrograph with spectral resolution of 0.07 nm measured the shift of rubies' emission spectral line with a 100 MPa maximum at 694.2 nm. Measurement uncertainty for pressure was ± (10 − 20) MPa, depending on the measurement. Pressure was generally increased step by step with an average rate of 6–20 MPa per minute. In special relaxation dynamics measurements described in Figs. 7 and 8, different waiting times were applied, as necessary for equilibration of the signal at each pressure point. Several measurements were carried out for each sample to ensure data reproducibility.

2.4. Data analysis

Fluorescence spectra were corrected for the background signals and spectral sensitivity of the measuring set-up. Emission spectrum plotted in energy scale at each pressure was characterized by its maximum position, width (defined by its fwhm), and integral intensity. The band position and width were determined using Origin curve-fitting algorithms (Microcal Software, Inc.). The apparent protein unfolding midpressure was determined by fitting unfolding profiles by sigmoidal curves as described, for example, in [33].

3. Results and discussion

3.1. Reference spectra recorded at ambient pressure

Overlaid in Fig. 1 are UV absorbtion (blue) and circular dichroism (CD) (black) spectra of native BChE at ambient pressure. All three

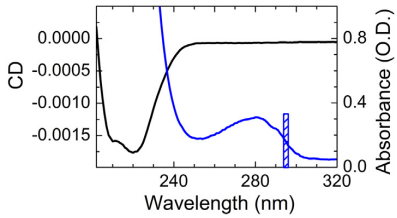


Fig. 1. Absorption (blue, right scale) and CD (black, left scale) spectra of native BChE at ambient pressure. Vertical column indicates the position and spectral width of the excitation laser pulse.

aromatic amino acid chromophores (Trp, Phe, and Tyr) are contributing to the weak absorption band around 280 nm. Trp residues are known to control the red side slope of this band. The CD spectrum corresponding to absorbance of aromatic amino acids is rather weak, and thus, dominated by a single structured band around 210–220 nm. This band related to peptide bond chromophores is characteristic of α-helical fold. Similar measurements on salted enzymes are, unfortunately, impractical because of the strong absorbance of salts in the absorption range of peptide chromophores.

Fig. 2 shows emission spectra of human BChE selectively excited at 295 nm and measured at different pressures. They have typical to fluorescence spectra of Trp residues uniform shapes.

A brief comparison of the main spectral characteristics for all samples measured at ambient pressure (Table 1) reveals that the spectrum of native BChE with no added salt has the shortest wavelength spectrum. The position and width of the Trp emission band is frequently regarded as a measure of structure and polarity of the microenvironment that surrounds the chromophore probes in proteins [34–36]. It can thus be concluded that the structure of native BChE is compact, with Trp residues least exposed to the polar water solvent. This conclusion is in agreement with the crystal structure of human BChE (PDB entry 1P0I). The prototypical high-order structure of the native BChE is further confirmed by the structured CD spectrum (Fig. 1).

Salting either by stabilizing or destabilizing salts leads to a small but regular red shift of the spectrum. The spectral red shift (i.e., toward longer wavelengths) is consistently accompanied by a broadening of the spectrum, which may indicate a salting-induced destabilization of the protein. The chaotropic salt has a stronger effect than the kosmotropic one, already observed at low concentration. A 1 M concentration of LiSCN led to complete denaturation of the enzyme and could thus not be measured. We also note that all the samples studied in this work show at ambient pressure the emission peak at considerably shorter wavelengths than that reported in ref. [18] (338 nm).

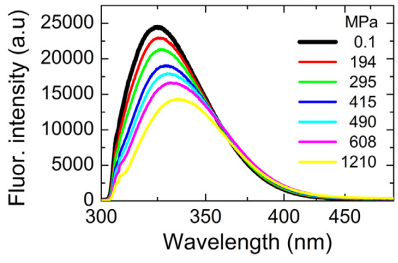


Fig. 2. Pressure-dependent auto-fluorescence spectra of native human BChE excited at 295 nm in 50 mM Tris-HCl pH 7.5. The ambient-pressure spectrum is highlighted by a black bold line. Note the reciprocal (linear in energy) wavelength scale. The spectra are deformed in the short-wavelength range below 305 nm due to the excitation cut-off filter used.

3.2. Pressure dependent emission spectra of BChE in the presence and absence of lyotropic salts

Upon sample compression, the Trp fluorescence bands shift to longer wavelengths and broaden (see Figs. 3-5). This is in striking contrast to the data of ref. [18], where no shift or shape change of the emission band was observed. The greatest shift detected at maximum experimental pressure of 1213 MPa in native BChE was 10.3 nm or 945 cm^{-1} (Fig. 3A). The peak wavelength achieved at this pressure is thus 335.3 nm. In the apparently most disordered sample (with 0.1 M LiSCN), the peak at 336.6 nm was recorded at 1189 MPa. These wavelengths, clearly corresponding to samples denatured by high pressure (see below), are within the experimental uncertainty rather close to that reported in ref. [18].

Significant broadening of the fluorescence band in native BChE is observed starting from about 160–170 MPa, and in salted samples earlier still. The pressure dependencies of fluorescence peak positions (Fig. 3A) and widths (Fig. 3B) are well correlated with each other. Both dependencies reveal three phases with relatively low sensitivity at low and high pressures and high sensitivity at intermediate pressures between 200 and 700 MPa (see below). In the latter pressure range, the rapid downshift of the fluorescence peak frequency coincides with the corresponding band broadening. These remarkable tendencies in different samples can be better compared in relative fluorescence band shift and broadening scales, as shown in Fig. 4.

Assuming the red shift of the Trp fluorescence band is related to enhanced exposure of the Trp residues to polar medium, the band shift and broadening data together could be interpreted as due to pressure-induced conformational changes of the protein structure. By previous FT-IR measurements [18] of vibrational amide I band, it was suggested that conformational changes that take place in the soman-aged-BChE samples involve secondary structure elements. The corresponding transition was shifted by ~100 MPa toward higher pressures compared with our experiments, which show a transition mid-pressure around 400 MPa (dashed line in Fig. 4). This difference is, perhaps, natural because of sample variation (soman-aged enzyme vs. native enzyme), and especially, because the present UV fluorescence and former IR vibrational absorption techniques probe variant substructures of the enzyme (tertiary and secondary, correspondingly). Furthermore, the protein concentration used in vibrational spectroscopy experiments [18] was an order of magnitude greater compared with current measurements, which might cause aggregation of the protein. In our experience, aggregation of proteins normally pushes the denaturing pressure toward higher values [37].

The fluorescence band broadening generally provides evidence for an increased flexibility of the protein structure in the vicinity of probe chromophores. In this context, from all samples studied, the native BChE appears most ordered (Figs. 3B and 4B). It shows by far the narrowest band width at normal pressure, only reaching the initial broadening level of the salt-treated protein spectra upon compression, see Fig. 3B. The denaturing broadening is largest for the sample with added “stabilizing” salt (0.1 M $(NH_4)_2SO_4$), see Fig. 4B. It is worth noting an excellent agreement with ref. [20], where the protein molten globule state appeared between 125 and 150 MPa, while in this work the fluorescence band broadening in native BChE started at 160–170 MPa. Even more significantly, the latter numbers almost coincide with the data of refs. [23, 24], which confirmed an irreversible loss of enzymatic activity beginning from 160 MPa.

Fig. 5 shows the pressure dependence of integrated fluorescence intensity, measured as area under the properly corrected emission spectrum. For technical reasons, the signal was not corrected for possible modification of the absorption spectrum (Fig. 1) with pressure. Yet pressure-induced modification of Trp absorption band position and width are expected to be smaller than those in fluorescence spectra, because of smaller dipole moment in the ground electronic state. Moreover, these spectral adjustments hardly influence comparative

Table 1
Fluorescence band positions and widths as well as relative band shifts (with respect to the native BChE with no salt) measured at ambient pressure.

Sample	Peak position (nm) (cm^{-1})		Relative band shift (cm^{-1})	Band width ± 10 (cm^{-1})
Native BChE	325.0	30,767 ± 8		4495
+ 0.1 M $(NH_4)_2$ SO_4	325.6	30,709 ± 11	−58	4645
+ 1 M $(NH_4)_2$ SO_4	326.1	30,665 ± 7	−102	4685
+ 0.1 M LiSCN	326.8	30,603 ± 14	−164	4689

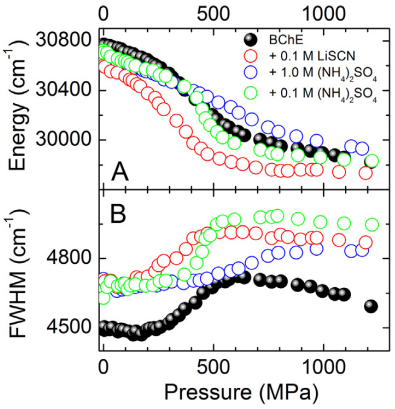


Fig. 3. Pressure dependence of the transition energy (A) and width (B) for BChE in the presence and absence of lyotropic salts.

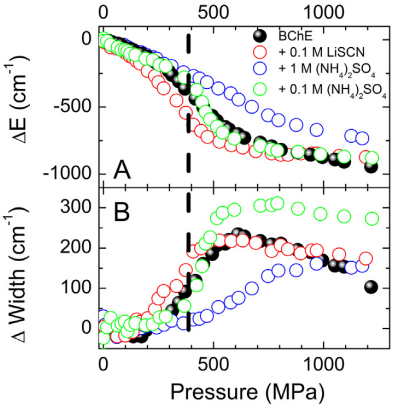


Fig. 4. Pressure dependence of the fluorescence band shift (relative transition energy, ΔE = E(P)-E(0)) (A) and broadening (relative fwhm, ΔWidth = Width (P)-Width(0)) (B) for BChE in the presence and absence of lyotropic salts. Vertical dashed line indicates the apparent transition mid-pressure for native BChE.

measurements, as long as they have been performed under similar conditions.

The fluorescence in all samples rather monotonously drops with increasing pressure (Fig. 5), though the quenching in native BChE seems to be delayed compared with salted samples. The overall intensity decrease is the smallest in the case of native enzyme (36%) and the greatest (48%) in BChE that contains 1 M $(NH_4)_2SO_4$ stabilizing salt. Although some of these effects may be due to modifications of the absorption spectrum, which requires additional study, general tendency remains – the salt-treated samples appear less ordered than the native one.

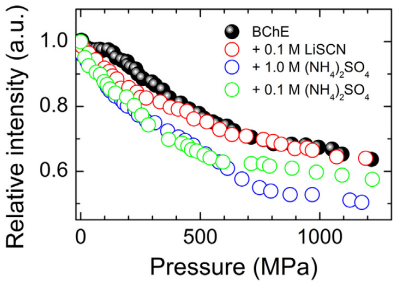


Fig. 5. Pressure dependence of integrated fluorescence intensity normalized with respect to the signal at ambient-pressure.

Due to qualitatively different pressure-dependent courses of fluorescence band position and width from one side (Figs. 3 and 4), and of fluorescence intensity from another side (Fig. 5), it is conceivable that the mechanisms of these dependencies are basically different (though some inter-dependency cannot be excluded). The quenching of Trp fluorescence has commonly been related with photoinduced electron transfer from indole chromophores to backbone amides [38,39]. High sensitivity of short-range charge transfer states to protein structural changes occurring under hydrostatic compression have also been demonstrated [40]. Pressure-induced modulation of the energies of quenching charge transfer states with respect to energies of local indole electronic transitions (^1L_a and ^1L_b) may thus explain the observed fluorescence quenching under external hydrostatic pressure.

3.3. Recovery of emission spectra upon decompression

A recovery of the Trp fluorescence spectra for native BChE is shown in Fig. 6. It is evident that the spectra do not mend completely as well as not promptly upon decompression. The decompressed spectra remain up to 4–5 nm (in different native samples) or 2–5 nm (in salted samples) more red shifted compared with respective initial spectra, being also about 100–250 cm^{-1} broader. However, while the salted samples

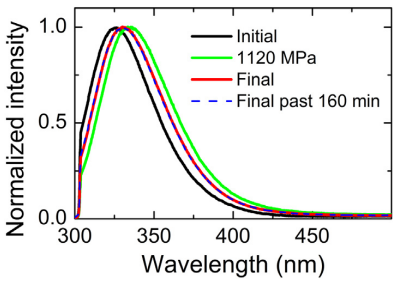


Fig. 6. Normalized emission spectra for native BChE, measured before the pressure cycle (black), at 1120 MPa (green), immediately after pressure release from 1120 MPa (red), and after additional waiting time of 160 min (dashed blue line overlapping with the solid red line). The spectra are deformed in the short-wavelength range below about 305 nm due to the excitation cut-off filter used.

continue slow relaxation toward fully recovered spectrum over a considerable time (~2 h, data not shown), spectral characteristics of the native sample measured at 1 bar stay steady both prior and after compression. This is despite the fact that the decompressed spectrum of the native enzyme retains significant deviations from the initial spectrum, see Fig. 6.

Slow but significant (several fold) increase in fluorescence intensity of the salted sample was also observed prior to as well as after compression. No such abnormality was detected in salt-free BChE. This slow kinetic change can be caused by a combination of effects such as (i) association/dissociation of salt ions with/from the enzyme structure; (ii) equilibrium change between different enzyme conformational states [41]; (iii) equilibrium changes resulting from alteration of the protein hydration due to the presence of lyotropic salts. Further studies are required to disentangle these complex effects.

3.4. Pressure dependence of the relaxation dynamics

Having demonstrated great temporal stability of the fluorescence from native enzymes both prior to compression and after the pressure release Fig. 6, we next studied pressure dependences of the Trp fluorescence peak position and width, as in Fig. 3, but waiting at each pressure point to see whether steady-state condition (dynamical equilibrium) for the parameter values had been achieved. Results of this study are shown in Figs. 7 and 8.

Data of Fig. 7 give evidence for strong dependence of structural flexibility of the enzyme on applied hydrostatic pressure. At low (below ~200 MPa) and high (above ~700 MPa) pressures the repeating measurements performed at constant pressure always yielded within the experimental uncertainty one and the same result for the fluorescence band position and width. Between 200 and 700 MPa, however, the protein is malleable, showing maximum flexibility at pressures around 400 ± 100 MPa. As noted above, this narrow pressure range corresponds to the transition range from native to unfolded enzyme as detected by Trp fluorescence. Similar data have been obtained for salted samples (data not shown).

In Fig. 8, the data of Fig. 7 (shown by blue balls) are plotted in parallel to the earlier data (black balls), measured according to our regular protocol, by increasing pressure step by step with an average rate of 6–20 MPa per min (see Materials and Methods). Note also that the two data sets were measured half a year apart using different samples. This explains some variation of the initial (0.1 MPa) data (which, for example, in case of width still remains within ~2% of its original value).

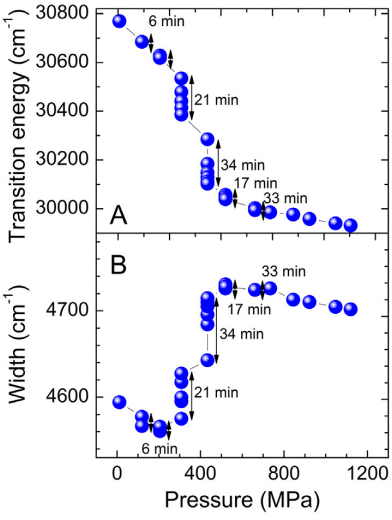


Fig. 7. Pressure dependence of the Trp fluorescence band maximum (A) and width (B) in native BChE. Multiple symbols at constant pressure correspond to consecutive measurements and manifest the presence of slow enzyme equilibration dynamics at these particular pressures. Numbers at selected pressures indicate corresponding waiting times. Solid lines connecting experimental points are for guidance of the eyes.

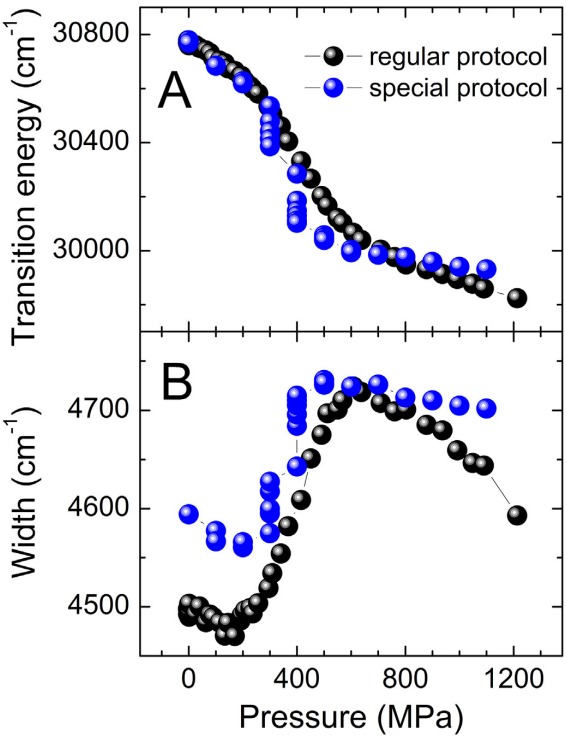


Fig. 8. Pressure dependence of the Trp fluorescence maximum position (A) and width (B) for native BChE. Shown with black symbols are data recorded according to our regular protocol (same as in Fig. 3) and to a special relaxation measurement protocol, by waiting at each pressure point as long as it takes to reach a steady-state condition (blue balls), see text for further explanations. Lines connecting experimental points are for guidance of the eyes.

Data of Fig. 8, obtained by applying two measuring protocols, follow qualitatively similar routes. In fact, at low pressures up until ~200 MPa, the two routs practically overlap. Yet the distribution of conformational states reached at high pressures using the different measurement protocols obviously deviate. Multiplicity of denatured states also supports the contention that "native" BChE may exist as multiple stable conformations. From the literature, it is known that some functional conformations may lead to bi- or multiple- stability behaviour.

4. Summary and conclusions

Pressure dependence of the Trp fluorescence spectrum in native and salted human BChE was studied up to about 1200 MPa.

A correlated large shift to longer wavelengths and broadening of the fluorescence band observed at pressure range of 200–700 MPa was interpreted as due to high-pressure induced denaturation of the proteins, leading to an enhanced exposure of a number of Trp residues into polar solvent environment. This transient process with transition mid-pressure of 400 ± 50 MPa for the native human BChE presumably involves

conformational changes of the enzyme at both tertiary and (taken into account results of ref. [18]) secondary structure levels. The enzyme structure, being highly responsive to pressure in the transition range, is rather inflexible outside this range. The pressure effects generally lacked reversibility when the transition pressure was exceeded. Almost complete recovery of the spectrum upon decompression was only observed in the sample with added stabilizing salt (0.1 M $(NH_4)_2SO_4$). This is the well-known salting-in effect that protects the protein against formation of a “scrambled” structure upon pressure release. The recovery was slow taking in average longer than 1 h at ambient temperature.

The above interpretations should be to a certain extent considered initial because decoding structural information from Trp spectroscopic data is generally rather challenging [34–36]. Specifically, there are known direct physical effects of pressure on fluorescence properties of Trp water solutions [42]. These effects, although considerably weaker than those observed in this study, should be taken into account in any quantitative evaluation of the protein stability. Furthermore, investigations of protein conformational stability under perturbing physical conditions more frequently than not give results that are strongly dependent on the history of protein preparations. Though active and inactive BChE forms resulting from purification processes do not show significant differences in electrophoretic mobility, isoelectric point, size and absorption spectra, they may display clear changes in conformational stability and fluorescence spectrum. The lyotropic salts by modifying the structure of water interacting with the solvent-exposed surface of the enzyme, may affect its sensitivity to compression. Parallel studies involving explicit molecular level quantum chemical modelling are, therefore, required for a detailed understanding of the enzyme properties, as described by pressure dependence of the Trp fluorescence.

Acknowledgements

This work was supported by the Estonian-France bilateral research cooperation program “G. F. PARROT” 2017-2018 and, partly, by the Estonian Research Council (grant IUT02-28). P. M. was supported by the Russian Science Foundation (grant 17-14-01097). The authors thank Oksana Lockridge from the University of Nebraska Medical Centre (Omaha, USA) for kindly providing the samples for these experiments and for useful comments.

References

[1] O. Lockridge, C.F. Bartels, T.A. Vaughan, C.K. Wong, S.E. Norton, L.L. Johnson, Complete amino acid sequence of human serum cholinesterase, J. Biol. Chem. 262 (1987) 549–557 http://www.ncbi.nlm.nih.gov/pubmed/3542989.
[2] C.V. Altamirano, O. Lockridge, Association of tetramers of human butyrylcholinesterase is mediated by conserved aromatic residues of the carboxy terminus, Chem. Biol. Interact. 119-120 (1999) 53–60 10.1016/S0009-2797(99)00013-7.
[3] C.V. Altamirano, O. Lockridge, Conserved aromatic residues of the C-terminus of human butyrylcholinesterase mediate the association of tetramers, Biochemistry 38 (1999) 13414–13422, https://doi.org/10.1021/bi991475.
[4] M.R. Blong, E. Bedows, O. Lockridge, Tetramerization domain of human butyrylcholinesterase is at the C-terminus, Biochem. J. 327 (1997) 747–757, https://doi.org/10.1042/bj3270747.
[5] H. Li, Lawrence M. Schopfer, P. Masson, O. Lockridge, Lamellipodin proline rich peptides associated with native plasma butyrylcholinesterase tetramers, Biochem. J. 411 (2008) 425–432, https://doi.org/10.1042/bj20071551.
[6] O. Jbilo, C.F. Bartels, A. Chatonnet, J.-P. Toutant, O. Lockridge, Tissue distribution of human acetylcholinesterase and butyrylcholinesterase messenger RNA, Toxicon 32 (1994) 1445–1457, https://doi.org/10.1016/0041-0101(94)90416-2.
[7] D. ØStergaard, J. Viby–Mogensen, H.K. Hanel, L.T. Skovgaard, Half–life of plasma cholinesterase, Acta Anaesthesiol. Scand. 32 (1988) 266–269, https://doi.org/10.1111/j.1399-6576.1988.tb02727.x.
[8] P. Masson, O. Lockridge, Butyrylcholinesterase for protection from organophosphorus poisons: Catalytic complexities and hysteretic behavior, Arch. Biochem. Biophys. 494 (2010) 107–120, https://doi.org/10.1016/j.abb.2009.12.005.
[9] E.G. Duysen, C.F. Bartels, O. Lockridge, Wild-type and A328W mutant human butyrylcholinesterase tetramers expressed in chinese hamster ovary cells have a 16-hour half-life in the circulation and protect mice from cocaine toxicity, J. Pharmacol. Exp. Ther. 302 (2002) 751–758, https://doi.org/10.1124/jpet.102.033746.
[10] B. Li, J.A. Stribley, A. Ticu, W. Xie, L.M. Schopfer, P. Hammond, S. Brimijoin, S.H. Hinrichs, O. Lockridge, Abundant tissue butyrylcholinesterase and its possible function in the acetylcholinesterase knockout mouse, J. Neurochem. 75 (2000) 1320–1331, https://doi.org/10.1046/j.1471-4159.2000.751320.x.
[11] M.M. Mesulam, A. Guillozet, P. Shaw, A. Levey, E.G. Duysen, O. Lockridge, Acetylcholinesterase knockouts establish central cholinergic pathways and can use butyrylcholinesterase to hydrolyze acetylcholine, Neuroscience 110 (2002) 627–639, https://doi.org/10.1016/S0306-4522(01)00613-3.
[12] W. Xie, J.A. Stribley, A. Chatonnet, P.J. Wilder, A. Rizzino, R.D. McComb, P. Taylor, S.H. Hinrichs, O. Lockridge, Postnatal developmental delay and supersensitivity to organophosphate in gene-targeted mice lacking acetylcholinesterase, J. Pharmacol. Exp. Ther. 293 (2000) 896–902 http://jpet.aspetjournals.org/content/293/3/896.
[13] S. Brimijoin, Y. Gao, L. Geng, V.P. Chen, Treating cocaine addiction, obesity, and emotional disorders by viral gene transfer of butyrylcholinesterase, Front. Pharmacol. 9 (2018), https://doi.org/10.3389/fphar.2018.00112.
[14] V.P. Chen, Y. Gao, L. Geng, R.J. Parks, Y.-P. Pang, S. Brimijoin, Plasma butyrylcholinesterase regulates ghrelin to control aggression, Proc. Natl. Acad. Sci. 112 (2015) 2251–2256, https://doi.org/10.1073/pnas.1421536112.
[15] O. Lockridge, Review of human butyrylcholinesterase structure, function, genetic variants, history of use in the clinic, and potential therapeutic uses, Pharmacol. Ther. 148 (2015) 34–46, https://doi.org/10.1016/j.pharmthera.2014.11.011.
[16] Y. Pan, J.L. Muzyka, C.-G. Zhan, Model of human butyrylcholinesterase tetramer by homology modeling and dynamics simulation, J. Phys. Chem. B 113 (2009) 6543–6552, https://doi.org/10.1021/jp8114995.
[17] C. Clery, N. Bec, C. Balny, V.V. Mozhaev, P. Masson, Kinetics of butyrylcholinesterase in reversed micelles under high pressure, Biochim. Biophys. Acta Protein Struct. Mol. Enzymol. 1253 (1995) 85–93, https://doi.org/10.1016/0167-4838(95)00137-J.
[18] C. Clery, K. Goossens, G. Hui Bon, K. Hoa, C. Heremans, P. Masson Balny, Pressure - induced structural modifications of butyrylcholinesterase, in: W.J.J. van den Tweel, A. Harder, R.M. Buitelaar (Eds.), Studies in Organic Chemistry, Elsevier, 1993, pp. 255–260 Place Published.
[19] C. Cléry, F. Renault, P. Masson, Pressure-induced molten globule state of cholinesterase, FEBS Lett. 370 (1995) 212–214, https://doi.org/10.1016/0014-5793(95)00787-A.
[20] P. Masson, P. Gouet, C. Clery, Pressure and propylene carbonate denaturation of native and "aged" phosphorylated cholinesterase, J. Mol. Biol. 238 (1994) 466–478, https://doi.org/10.1006/jmbi.1994.1305.
[21] A. Weingand-Ziade, F. Renault, P. Masson, Differential effect of pressure and temperature on the catalytic behaviour of wild-type human butyrylcholinesterase and its D70G mutant, Eur. J. Biochem. 264 (1999) 327–335, https://doi.org/10.1046/j.1432-1327.1999.00609.x.
[22] A. Weingand-Ziadé, F. Renault, P. Masson, Combined pressure/heat-induced inactivation of butyrylcholinesterase, Biochim. Biophys. Acta Protein Struct. Mol. Enzymol. 1340 (1997) 245–252, https://doi.org/10.1016/S0167-4838(97)00051-4.
[23] P. Masson, C. Balny, Conformational plasticity of butyrylcholinesterase as revealed by high pressure experiments, Biochim. Biophys. Acta Protein Struct. Mol. Enzymol. 1041 (1990) 223–231, https://doi.org/10.1016/0167-4838(90)90276-L.
[24] P. Masson, N. Bec, M.-T. Froment, F. Nachon, C. Balny, O. Lockridge, L.M. Schopfer, Rate-determining step of butyrylcholinesterase-catalyzed hydrolysis of benzoylcholine and benzoylthiocholine, Eur. J. Biochem. 271 (2004) 1980–1990, https://doi.org/10.1111/j.1432-1033.2004.04110.x.
[25] K.M. Boyko, T.N. Baymukhametov, Y.M. Chesnokov, M. Hons, S.V. Lushchekina, P.V. Konarev, A.V. Lipkin, A.L. Vasiliev, P. Masson, V.O. Popov, M.V. Kovalchuk, 3D structure of the natural tetrameric form of human butyrylcholinesterase as revealed by cryoEM, SAXS and MD, Biochimie 156 (2018) 196–205, https://doi.org/10.1016/j.biochi.2018.10.017.
[26] P. Masson, M.-T. Froment, S. Fort, F. Ribes, N. Bec, C. Balny, L.M. Schopfer, Butyrylcholinesterase-catalyzed hydrolysis of N-methylindoxyl acetate: analysis of volume changes upon reaction and hysteretic behavior, Biochim. Biophys. Acta Protein Struct. Mol. Enzymol. 1597 (2002) 229–243, https://doi.org/10.1016/S0167-4838(02)00265-0.
[27] S. Onder, E. David, O. Tacal, L.M. Schopfer, O. Lockridge, Hupresin retains binding capacity for butyrylcholinesterase and acetylcholinesterase after sanitation with sodium hydroxide, Front. Pharmacol. 8 (2017) 713, https://doi.org/10.1111/j.1399-6576.1988.tb02727.x.
[28] A. Saxena, P. Tipparaju, C. Luo, B.P. Doctor, Pilot-scale production of human serum butyrylcholinesterase suitable for use as a bioscavenger against nerve agent toxicity, Process Biochem. 45 (2010) 1313–1318, https://doi.org/10.1016/j.procbio.2010.04.021.
[29] R.C. Neuman, W. Kauzmann, A. Zipp, Pressure dependence of weak acid ionization in aqueous buffers, J. Phys. Chem. 77 (1973) 2687–2691, https://doi.org/10.1021/j100640a025.
[30] A. Freiberg, L. Kangur, D. John, C.N. Hunter Olsen, Structural implications of hydrogen-bond energetics in membrane proteins revealed by high-pressure spectroscopy, Biophys. J. 103 (2012) 2352–2360, https://doi.org/10.1016/j.bpj.2012.10.030.
[31] L. Kangur, M.R. Jones, A. Freiberg, Hydrogen bonds in the vicinity of the special pair of the bacterial reaction center probed by hydrostatic high-pressure absorption spectroscopy, Biophys. Chem. 231 (2017) 27–33, https://doi.org/10.1016/j.bpc.2017.04.003.
[32] L. Kangur, K. Leiger, A. Freiberg, Evidence for high-pressure-induced rupture of hydrogen bonds in LH2 photosynthetic antenna pigment-protein complexes, J.

Phys. 121 (2008) 112004 http://stacks.iop.org/1742-6596/121/i=11/a=112004.
[33] E. Monsellier, H. Bedouelle, Quantitative measurement of protein stability from unfolding equilibria monitored with the fluorescence maximum wavelength, Protein Eng. Des. Sel. 18 (2005) 445–456, https://doi.org/10.1093/protein/gzi046.
[34] J. Hixon, Y. Reshetnyak, Algorithm for the analysis of tryptophan fluorescence spectra and their correlation with protein structural parameters, Algorithms 2 (2009) 1155, https://doi.org/10.3390/a2031155.
[35] T.Q. Luong, S. Kapoor, R. Winter, Pressure—A gateway to fundamental insights into protein solvation, dynamics, and function, ChemPhysChem 16 (2015) 3555–3571, https://doi.org/10.1002/cphc.201500669.
[36] J.L. Silva, A.C. Oliveira, T.C.R.G. Vieira, G.A.P. de Oliveira, M.C. Suarez, D. Foguel, High-pressure chemical biology and biotechnology, Chem. Rev. 114 (2014) 7239–7267, https://doi.org/10.1021/cr400204z.
[37] L. Kangur, K. Timpmann, A. Freiberg, Stability of integral membrane proteins under high hydrostatic pressure: The LH2 and LH3 antenna pigment−protein complexes from photosynthetic bacteria, J. Phys. Chem. B 112 (2008) 7948–7955, https://doi.org/10.1021/jp801943w.
[38] P.R. Callis, [7] 1La and 1Lb transitions of tryptophan: Applications of theory and experimental observations to fluorescence of proteins, Methods in Enzymology, Academic Press, 1997, pp. 113–150 Place Published.
[39] P.R. Callis, Exploring the Electrostatic Landscape of Proteins with Tryptophan Fluorescence, Reviews in Fluorescence 2007, Springer New York, 2009, pp. 199–248 Place Published.
[40] K. Timpmann, A. Ellervee, T. Pullerits, R. Ruus, V. Sundström, A. Freiberg, Short-range exciton couplings in LH2 photosynthetic antenna proteins studied by high hydrostatic pressure absorption spectroscopy, J. Phys. Chem. B 105 (2001) 8436–8444, https://doi.org/10.1021/jp003496f.
[41] P. Masson, Time-dependent kinetic complexities in cholinesterase-catalyzed reactions, Biochem. Mosc. 77 (2012) 1147–1161, https://doi.org/10.1134/S0006297912100070.
[42] K. Ruan, S. Tian, R. Lange, C. Balny, Pressure Effects on Tryptophan and Its Derivatives, Biochem. Biophys. Res. Commun. 269 (2000) 681–686, https://doi.org/10.1006/bbrc.2000.2345.

List of Figures

List of Tables

Abbreviations

A-L	Alpha-Lactalbumin (bovine)
A-L_{ca}	Alpha-Lactalbumin with Ca^{2+} (bovine)
A-L_{dep}	Ca^{2+} depleted Alpha-Lactalbumin (bovine)
BS	backscattering
BSS	backscattering spectrometer
Do model	EINS model used in Doster et al. [2], p. 19
DSF	dynamic scattering function or structure factor
EI	elastic intensity (experimental), defined in Eq. (5.1), p. 46
EINS	elastic incoherent neutron scattering
EISF	elastic incoherent structure factor
ESS	European Spallation Source
FRM-II	FRM-II reactor source
FWHM	full width at half maximum
GA model	Gaussian approximation model for EINS, p. 16
ILL	Institute Laue Langevin (reactor source)
IN13	neutron instrument at ILL
INS	inelastic neutron scattering
IRIS	neutron instrument at ISIS
ISF	intermediate scattering function
ISIS	ISIS Pulsed Neutron and Muon Source (STFC)
MSD	mean square displacement
MSPF	mean square position fluctuation, equal to the time independent (or static) MSD
OSIRIS	neutron instrument at ISIS
PK model	EINS model used in Peters and Kneller [10], p. 18
QENS	quasi-elastic neutron scattering
SPHERES	neutron instrument at FRM-II
ToF	Time-of-Flight
ToF-BSS	Time-of-Flight backscattering spectrometer
Yi model	EINS model used in Yi et al. [9], p. 19

Bibliography

[1] Benno P. Schoenborn. "Neutron Diffraction Analysis of Myoglobin". *Nature* 224 (1969), pp. 143–146. DOI: 10.1038/224143a0.

[2] W. Doster, S. Cusack, and W. Petry. "Dynamical transition of myoglobin revealed by inelastic neutron scattering". *Nature* 337.6209 (1989), pp. 754–6. ISSN: 0028-0836 (Print) 0028-0836 (Linking). DOI: 10.1038/337754a0. URL: http://www.ncbi.nlm.nih.gov/pubmed/2918910.

[3] Wolfgang Doster. "The dynamical transition of proteins, concepts and misconceptions". *European Biophysics Journal* 37.5 (Feb. 2008), pp. 591–602. DOI: 10.1007/s00249-008-0274-3.

[4] Jörg Pieper et al. "Temperature-dependent vibrational and conformational dynamics of photosystem II membrane fragments from spinach investigated by elastic and inelastic neutron scattering". *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 1817.8 (Aug. 2012), pp. 1213–1219. DOI: 10.1016/j.bbabio.2012.03.020.

[5] Leonid Rusevich et al. "Perspectives in biological physics: The nDDB project for a neutron Dynamics Data Bank for biological macromolecules". *The European Physical Journal E* 36.7 (July 2013), p. 80. ISSN: 1292-895X. DOI: 10.1140/epje/i2013-13080-5. URL: https://doi.org/10.1140/epje/i2013-13080-5.

[6] M. Bée. *Quasielastic Neutron Scattering:Principles and Applications in Solid State Chemistry, Biology and Materials Science*. Adam Hilger, Philadelphia, 1988.

[7] A. Rahman, K. S. Singwi, and A. Sjolander. "Theory of Slow Neutron Scattering by Liquids .1." *Physical Review* 126.3 (1962), pp. 986–996. ISSN: 0031-899X. DOI: 10.1103/PhysRev.126.986. URL: %3CGo%20to%20ISI%3E://WOS:A19621470C00047.

[8] G. Squires. *Introduction to the Theory of Thermal Neutron Scatering*. Dover Publications, 1978.

[9] Z. Yi et al. "Derivation of mean-square displacements for protein dynamics from elastic incoherent neutron scattering". *J Phys Chem B* 116.16 (2012), pp. 5028–36. ISSN: 1520-5207 (Electronic) 1520-5207 (Linking). DOI: 10.1021/jp2102868. URL: http://www.ncbi.nlm.nih.gov/pubmed/22471396.

[10] J. Peters and G. R. Kneller. "Motional heterogeneity in human acetylcholinesterase revealed by a non-Gaussian model for elastic incoherent neutron scattering". *J Chem Phys* 139.16 (2013), p. 165102. ISSN: 1089-7690 (Electronic) 0021-9606 (Linking). DOI: 10.1063/1.4825199. URL: https://www.ncbi.nlm.nih.gov/pubmed/24182083.

[11] Lars Meinhold et al. "Protein Dynamics and Stability: The Distribution of Atomic Fluctuations in Thermophilic and Mesophilic Dihydrofolate Reductase Derived Using Elastic Incoherent Neutron Scattering". *Biophysical Journal* 94.12 (2008), pp. 4812–4818. ISSN: 0006-3495. DOI: https://doi.org/10.1529/biophysj.107.121418. URL: http://www.sciencedirect.com/science/article/pii/S0006349508703485.

[12] Atsushi Tokuhisa et al. "Non-Gaussian behavior of elastic incoherent neutron scattering profiles of proteins studied by molecular dynamics simulation". *Phys. Rev. E* 75 (4 Apr. 2007), p. 041912. DOI: 10.1103/PhysRevE.75.041912. URL: https://link.aps.org/doi/10.1103/PhysRevE.75.041912.

[13] Derya Vural et al. "Motional displacements in proteins: The origin of wave-vector-dependent values". *Phys. Rev. E* 91 (5 May 2015), p. 052705. DOI: 10.1103/PhysRevE.91.052705. URL: https://link.aps.org/doi/10.1103/PhysRevE.91.052705.

[14] Helmut Schober. "An introduction to the theory of nuclear neutron scattering in condensed matter." *Journal of Neutron Research* 17.3-4 (2014), pp. 109–357. DOI: 10.3233/JNR-140016.

[15] M. Tanabashi et al. "Review of Particle Physics". *Phys. Rev. D* 98 (3 Aug. 2018), p. 030001. DOI: 10.1103/PhysRevD.98.030001. URL: https://link.aps.org/doi/10.1103/PhysRevD.98.030001.

[16] V. F. Sears. "Neutron scattering lengths and cross sections". *Neutron News* 3 (1992), pp. 26–37.

[17] Ana M. Gaspar et al. "Using polarization analysis to separate the coherent and incoherent scattering from protein samples". *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* 1804.1 (2010). Includes Special Section: Protein-Water Interactions, pp. 76–82. ISSN: 1570-9639. DOI: https://doi.org/10.1016/j.bbapap.2009.06.024. URL: http://www.sciencedirect.com/science/article/pii/S1570963909001629.

[18] Mark T.F. Telling et al. "Mean squared displacement analysis of an-harmonic behaviour in lyophilised proteins". *Chemical Physics* 424 (2013). Neutron Scattering Highlights on Water and Biological Systems, pp. 32–36. ISSN: 0301-0104. DOI: https://doi.org/10.1016/j.chemphys.2013.05.008. URL: http://www.sciencedirect.com/science/article/pii/S0301010413002218.

[19] W. Doster et al. "Elastic resolution spectroscopy: a method to study molecular motions in small biological samples". *Physica B* 301.1-2 (2001), pp. 65–68. ISSN: 0921-4526. DOI: Doi10.1016/S0921-4526(01)00513-0. URL: %3CGo%20to%20ISI%3E://WOS:000169563000011.

[20] W. Doster, H. Nakagawa, and M.S. Appavou. "Scaling analysis of bio-molecular dynamics derived from elastic incoherent neutron scattering experiments". *J. Chem. Phys.* 139 (2013), p. 045105.

[21] Jean Pierre Boon and Sidney Yip. *Molecular Hydrodynamics*. New York : Dover, 1991. ISBN: 0486669491.

[22] V. Réat et al. "Functional Dynamics in Purple Membranes". In: *Biological Macromolecular Dynamcis, Proceedings of a Workshop on Inelastic and Quasielastic Neutron Scattering in Biology*. 1996, pp. 117–122.

[23] Gerald R. Kneller and Guillaume Chevrot. "Impact of anisotropic atomic motions in proteins on powder-averaged incoherent neutron scattering intensities". *The Journal of Chemical Physics* 137.22 (2012), p. 225101. DOI: 10.1063/1.4769782. eprint: https://doi.org/10.1063/1.4769782. URL: https://doi.org/10.1063/1.4769782.

[24] Zhuo Liu et al. "Dynamical Transition of Collective Motions in Dry Proteins". *Phys. Rev. Lett.* 119 (4 July 2017), p. 048101. DOI: 10.1103/PhysRevLett.119.048101. URL: https://link.aps.org/doi/10.1103/PhysRevLett.119.048101.

[25] W. Doster. "Are Proteins Dynamically Heterogeneous? Neutron Scattering Analysis of Hydrogen Displacement Distributions." *Int J Mol Theor Phy.* 2 (1 2018), pp. 1–14.

[26] Gerald R. Kneller and Konrad Hinsen. "Quantitative model for the heterogeneity of atomic position fluctuations in proteins: A simulation study". *The Journal of Chemical Physics* 131.4 (2009), p. 045104. DOI: 10.1063/1.3170941. eprint: https://doi.org/10.1063/1.3170941. URL: https://doi.org/10.1063/1.3170941.

[27] Hiroshi Nakagawa et al. "Protein Dynamical Heterogeneity Derived from Neutron Incoherent Elastic Scattering". *Journal of the Physical Society of Japan* 73.2 (2004), pp. 491–495. DOI: 10.1143/jpsj.73.491.

[28] Derya Vural, Jeremy C. Smith, and Henry R. Glyde. "Determination of Dynamical Heterogeneity from Dynamic Neutron-Scattering of Proteins". *Biophysical Journal* (2018). ISSN: 0006-3495. DOI: https://doi.org/10.1016/j.bpj.2018.02.024. URL: http://www.sciencedirect.com/science/article/pii/S000634951830256X.

[29] *NIST Digital Library of Mathematical Functions*. http://dlmf.nist.gov/, Release 1.0.18 of 2018-03-27. F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller and B. V. Saunders, eds. URL: http://dlmf.nist.gov/.

[30] Torsten Becker and Jeremy C. Smith. "Energy resolution and dynamical heterogeneity effects on elastic incoherent neutron scattering from molecular systems". *Phys. Rev. E* 67 (2 Feb. 2003), p. 021904. DOI: 10.1103/PhysRevE.67.021904. URL: https://link.aps.org/doi/10.1103/PhysRevE.67.021904.

[31] A. Rahman. "Correlations in the Motion of Atoms in Liquid Argon". *Physical Review* 136.2A (Oct. 1964), A405–A411. DOI: 10.1103/physrev.136.a405.

[32] K. Sköld et al. "Coherent- and Incoherent-Scattering Laws of Liquid Argon". *Physical Review A* 6.3 (Sept. 1972), pp. 1107–1131. DOI: 10.1103/physreva.6.1107.

[33] D. Zeller et al. "Analysis of elastic incoherent neutron scattering data beyond the Gaussian approximation". *The Journal of Chemical Physics* 149.23 (Dec. 2018), p. 234908. DOI: 10.1063/1.5049938.

[34] Helmut Schober. "Neutron Scattering Instrumentation". In: *Neutron Applications in Earth, Energy and Environmental Sciences. Neutron Scattering Applications and Techniques*. Ed. by L. Liang, R. Rinaldi, and H. Schober. 1st ed. Springer, 2009. Chap. 3, pp. 37–104. ISBN: 978-0-387-09416-8. DOI: https://doi.org/10.1007/978-0-387-09416-8_3.

[35] MLZ. *Neutron Source*. URL: https://www.mlz-garching.de/englisch/neutron-research/neutron-source.html (visited on 09/04/2018).

[36] Harald Conrad. "Spallation – Neutrons Beyond Nuclear Fission". In: *Handbook of Particle Detection and Imaging*. Ed. by Claus Grupen and Irène Buvat. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 719–757. ISBN: 978-3-642-13271-1. DOI: 10.1007/978-3-642-13271-1_30. URL: https://doi.org/10.1007/978-3-642-13271-1_30.

[37] PSI. *The Swiss Spallation Neutron Source (SINQ)*. URL: https://www.psi.ch/sinq/ (visited on 09/04/2018).

[38] Roland Garoby et al. "The European Spallation Source Design". *Physica Scripta* 93.1 (2018), p. 014001. DOI: https://doi.org/10.1088/1402-4896/aa9bff. URL: http://stacks.iop.org/1402-4896/93/i=1/a=014001.

[39] B. Frick. "Neutron Backscattering Spectroscopy". In: *Neutron and X-ray Spectroscopy*. Ed. by F. Hippert et al. Springer, 2006, pp. 483–527. DOI: 10.1007/1-4020-3337-0_15.

[40] F. Natali et al. "IN13 Backscattering Spectrometer at ILL: Looking for Motions in Biological Macromolecules and Organisms". *Neutron News* 19.4 (2008), pp. 14–18. ISSN: 1044-8632 1931-7352. DOI: 10.1080/10448630802474083. URL: http://dx.doi.org/10.1080/10448630802474083%20http://www.tandfonline.com/doi/abs/10.1080/10448630802474083.

[41] ILL. *IN13 Description*. URL: https://www.ill.eu/users/instruments/instruments-list/in13/description/instrument-layout/ (visited on 08/24/2018).

[42] J. Wuttke et al. "SPHERES, Julich's high-flux neutron backscattering spectrometer at FRM II". *Rev Sci Instrum* 83.7 (2012), p. 075109. ISSN: 1089-7623 (Electronic) 0034-6748 (Linking). DOI: 10.1063/1.4732806. URL: http://www.ncbi.nlm.nih.gov/pubmed/22852726.

[43] Heinz Maier-Leibnitz Zentrum. "SPHERES: Backscattering spectrometer". *Journal of large-scale research facilities* 1 (2015), A30. DOI: http://dx.doi.org/10.17815/jlsrf-1-38.

[44] Joachim Wuttke and Michaela Zamponi. "Simulation-guided optimization of small-angle analyzer geometry in the neutron backscattering spectrometer SPHERES". *Review of Scientific Instruments* 84.11 (2013), p. 115108. DOI: 10.1063/1.4831815. eprint: https://doi.org/10.1063/1.4831815. URL: https://doi.org/10.1063/1.4831815.

[45] B. Frick et al. "Recent Backscattering Instrument Developments at the ILL and SNS". *Zeitschrift Fur Physikalische Chemie-International Journal of Research in Physical Chemistry & Chemical Physics* 224.1-2 (2010), pp. 33–60. ISSN: 0942-9352. DOI: 10.1524/zpch.2010.6091. URL: %3CGo%20to%20ISI%3E://WOS:000275275000003.

[46] A. Meyer et al. "The high-flux backscattering spectrometer at the NIST Center for Neutron Research". *Review of Scientific Instruments* 74.5 (May 2003), pp. 2759–2777. DOI: 10.1063/1.1568557.

[47] C. J. Carlile and M. A. Adams. "The design of the IRIS inelastic neutron spectrometer and improvements to its analysers". *Physica B: Condensed Matter* 182.4 (Dec. 1992), pp. 431–440. DOI: 10.1016/0921-4526(92)90047-v.

[48] Mark. T. F. Telling and Ken. H. Andersen. "Spectroscopic characteristics of the OSIRIS near-backscattering crystal analyser spectrometer on the ISIS pulsed neutron source". *Phys. Chem. Chem. Phys.* 7 (6 2005), pp. 1255–1261. DOI: 10.1039/B413934H. URL: http://dx.doi.org/10.1039/B413934H.

[49] E. Mamontov and K. W. Herwig. "A time-of-flight backscattering spectrometer at the Spallation Neutron Source, BASIS". *Review of Scientific Instruments* 82.8 (2011), p. 085109. DOI: 10.1063/1.3626214. eprint: https://doi.org/10.1063/1.3626214. URL: https://doi.org/10.1063/1.3626214.

[50] STFC. *IRIS*. URL: https://www.isis.stfc.ac.uk/Pages/Iris.aspx (visited on 08/23/2018).

[51] Eugene A. Permyakov and Lawrence J. Berliner. "alpha-Lactalbumin: structure and function". *FEBS Letters* 473.3 (), pp. 269–274. DOI: 10.1016/S0014-5793(00)01546-5. eprint: https://febs.onlinelibrary.wiley.com/doi/pdf/10.1016/S0014-5793%2800%2901546-5. URL: https://febs.onlinelibrary.wiley.com/doi/abs/10.1016/S0014-5793%2800%2901546-5.

[52] Katsutoshi Nitta and Shintaro Sugai. "The evolution of lysozyme and alpha-lactalbumin". *European Journal of Biochemistry* 182.1 (), pp. 111–118. DOI: 10.1111/j.1432-1033.1989.tb14806.x. eprint: https://febs.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1432-1033.1989.tb14806.x. URL: https://febs.onlinelibrary.wiley.com/doi/abs/10.1111/j.1432-1033.1989.tb14806.x.

[53] R. Shinozaki and M. Iwaoka. "Effects of Metal Ions, Temperature, and a Denaturant on the Oxidative Folding Pathways of Bovine alpha-Lactalbumin". *International Journal of Molecular Sciences* 18.9 (2017), p. 1996. ISSN: 1422-0067. DOI: 10.3390/ijms18091996. URL: http://dx.doi.org/10.3390/ijms18091996.

[54] Eugene A. Permyakov et al. "Calcium binding to α-lactalbumin: Structural rearrangement and association constant evaluation by means of intrinsic protein fluorescence changes". *Biochemical and Biophysical Research Communications* 100.1 (May 1981), pp. 191–197. DOI: 10.1016/s0006-291x(81)80081-2.

[55] D.A. Dolgikh et al. "α-lactalbumin: compact state with fluctuating tertiary structure?" *FEBS Letters* 136.2 (Dec. 1981), pp. 311–315. DOI: 10.1016/0014-5793(81)80642-4.

[56] Y. Hiraoka and S. Sugai. "Equilibrium and kinetic study of sodium-and potassium-induced conformational changes of apo-α-lactalbumin". *International Journal of Peptide and Protein Research* 26.3 (Jan. 1985), pp. 252–261. DOI: 10.1111/j.1399-3011.1985.tb03203.x.

[57] Yuri V. Griko and David P. Remeta. "Energetics of solvent and ligand-induced conformational changes in α-lactalbumin". *Protein Science* 8.3 (Dec. 1999), pp. 554–561. DOI: 10.1110/ps.8.3.554.

[58] A. Hakansson et al. "Apoptosis induced by a human milk protein." *Proceedings of the National Academy of Sciences* 92.17 (Aug. 1995), pp. 8064–8068. DOI: 10.1073/pnas.92.17.8064.

[59] M. Svensson et al. "Conversion of alpha -lactalbumin to a protein inducing apoptosis". *Proceedings of the National Academy of Sciences* 97.8 (Apr. 2000), pp. 4221–4226. DOI: 10.1073/pnas.97.8.4221.

[60] Helen M. Berman et al. "The Protein Data Bank". *Nucleic Acids Research* 28.1 (Jan. 2000), pp. 235–242. DOI: 10.1093/nar/28.1.235.

[61] Z. Bu et al. "A view of dynamics changes in the molten globule-native folding step by quasielastic neutron scattering". *J Mol Biol* 301.2 (2000), pp. 525–36. ISSN: 0022-2836 (Print) 0022-2836 (Linking). DOI: 10.1006/jmbi.2000.3978. URL: http://www.ncbi.nlm.nih.gov/pubmed/10926525.

[62] Zimei Bu, Jeremy Cook, and David J.E Callaway. "Dynamic regimes and correlated structural dynamics in native and denatured alpha-lactalbumin". *Journal of Molecular Biology* 312.4 (Sept. 2001), pp. 865–873. DOI: 10.1006/jmbi.2001.5006.

[63] Mounir Tarek, Dan A. Neumann, and Douglas J. Tobias. "Characterization of sub-nanosecond dynamics of the molten globule state of α-lactalbumin using quasielastic neutron scattering and molecular dynamics simulations". *Chemical Physics* 292.2-3 (Aug. 2003), pp. 435–443. DOI: 10.1016/s0301-0104(03)00291-x.

[64] Samapan Sikdar, J. Chakrabarti, and Mahua Ghosh. "A microscopic insight from conformational thermodynamics to functional ligand binding in proteins". *Mol. BioSyst.* 10.12 (Sept. 2014), pp. 3280–3289. DOI: 10.1039/c4mb00434e.

[65] Z. F. Brotzakis et al. "Dynamics of Hydration Water around Native and Misfolded α-Lactalbumin". *The Journal of Physical Chemistry B* 120.21 (May 2016), pp. 4756–4766. DOI: 10.1021/acs.jpcb.6b02592.

[66] Dariya S. Glazer, Randall J. Radmer, and Russ B. Altman. "Improving Structure-Based Function Prediction Using Molecular Dynamics". *Structure* 17.7 (July 2009), pp. 919–929. DOI: 10.1016/j.str.2009.05.010.

[67] S. Perticaroli et al. "Description of Hydration Water in Protein (Green Fluorescent Protein) Solution". *J Am Chem Soc* 139.3 (2017), pp. 1098–1105. ISSN: 1520-5126 (Electronic) 0002-7863 (Linking). DOI: 10.1021/jacs.6b08845. URL: https://www.ncbi.nlm.nih.gov/pubmed/27783480.

[68] D. Richard, M. Ferrand, and G. J. Kearley. "Analysis and visualisation of neutron-scattering data". *Journal of Neutron Research* 4.1 (1996), pp. 33–39. ISSN: 1023-8166. DOI: 10.1080/10238169608200065. URL: http://dx.doi.org/10.1080/10238169608200065%20http://iospress.metapress.com/content/27p31m235g721313/fulltext.pdf.

[69] O. Arnold et al. "Mantid—Data analysis and visualization package for neutron scattering and μ SR experiments". *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 764 (2014), pp. 156–166. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2014.07.029. URL: http://www.sciencedirect.com/science/article/pii/S0168900214008729.

[70] Dominik Zeller and Judith Peters. "Internal time on IN13". *Institut Laue-Langevin (ILL)* (2016). DOI: DOI:10.5291/ill-data.inter-337.

[71] Dominik Zeller et al. "Testing the Validity of Current Models to Describe the Protein Dynamics from EFWS data". *Institut Laue-Langevin (ILL)* (2018). DOI: DOI:10.5291/ill-data.crg-2485.

[72] M. Settles and W. Doster. "Iterative Calculation of the Vibrational Density of States from Incoherent Neutron Scattering Data with the Account of Double Scattering". In: *Biological Macromolecular Dynamics*. Ed. by S. Cusack et al. Adenine Press, NY, USA, 1997.

[73] S. Busch and T. Unruh. "The slow short-time motions of phospholipid molecules with a focus on the influence of multiple scattering and fitting artefacts". *Journal of Physics-Condensed Matter* 23.25 (2011), pp. 254205–. ISSN: 0953-8984. DOI: 10.1088/0953-8984/23/25/254205. URL: %3CGo%20to%20ISI%3E://WOS:000291433000008.

[74] Jörg Pieper et al. "Light-induced Modulation of Protein Dynamics During the Photocycle of Bacteriorhodopsin". *Photochemistry and Photobiology* 85.2 (Mar. 2009), pp. 590–597. DOI: 10.1111/j.1751-1097.2008.00501.x.

[75] A. M. Stadler et al. "Cytoplasmic water and hydration layer dynamics in human red blood cells". *J Am Chem Soc* 130.50 (2008), pp. 16852–3. ISSN: 1520-5126 (Electronic) 0002-7863 (Linking). DOI: 10.1021/ja807691j. URL: https://www.ncbi.nlm.nih.gov/pubmed/19053467.

[76] M. Trapp et al. "Hydration dependent studies of highly aligned multilayer lipid membranes by neutron scattering". *J Chem Phys* 133.16 (2010), p. 164505. ISSN: 1089-7690 (Electronic) 0021-9606 (Linking). DOI: 10.1063/1.3495973. URL: http://www.ncbi.nlm.nih.gov/pubmed/21033803.

[77] Helen Thompson et al. "Proton dynamics in lithium-ammonia solutions and expanded metals". *The Journal of Chemical Physics* 124.2 (Jan. 2006), p. 024501. DOI: 10.1063/1.2145745.

[78] Joachim Wuttke. "Multiple-scattering effects on smooth neutron-scattering spectra". *Phys. Rev. E* 62 (5 Nov. 2000), pp. 6531–6539. DOI: 10.1103/PhysRevE.62.6531. URL: https://link.aps.org/doi/10.1103/PhysRevE.62.6531.

[79] Reiner Zorn. "On the evaluation of neutron scattering elastic scan data". *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 603.3 (May 2009), pp. 439–445. DOI: 10.1016/j.nima.2009.02.040.

[80] Matthew Newville et al. *LMFIT: Non-Linear Least-Square Minimization and Curve-Fitting for Python*. Sept. 2014. DOI: 10.5281/zenodo.11813. URL: https://doi.org/10.5281/zenodo.11813.

[81] F. Gabel et al. "Proteins dynamics studied by neutron scattering". *Quaterly reviews of biophysics* 35.4 (2002), pp. 327–367.

[82] M. Trapp et al. "High hydrostatic pressure effects investigated by neutron scattering on lipid multilamellar vesicles". *Physical Chemistry Chemical Physics* 15.48 (2013), pp. 20951–20956. ISSN: 1463-9076. DOI: 10.1039/c3cp52762j. URL: %3CGo%20to%20ISI%3E://WOS:000327467000020.

[83] J. Marion et al. "Pressure-induced molten globule state of human acetylcholinesterase: structural and dynamical changes monitored by neutron scattering". *Physical Chemistry Chemical Physics* 17.5 (2015), pp. 3157–3163. ISSN: 1463-9076. DOI: DOI10.1039/c4cp02992e. URL: %3CGo%20to%20ISI%3E://WOS:000348203200024.

[84] G. Zaccai. "How Soft Is a Protein? A Protein Dynamics Force Constant Measured by Neutron Scattering". *Science* 288.5471 (2000), pp. 1604–1607. ISSN: 00368075 10959203. DOI: 10.1126/science.288.5471.1604.

[85] Bjarne F. Rasmussen et al. "Crystalline ribonuclease A loses function below the dynamical transition at 220 K". *Nature* 357.6377 (June 1992), pp. 423–424. DOI: 10.1038/357423a0.

[86] M. Tarek and D. J. Tobias. "Role of protein-water hydrogen bond dynamics in the protein dynamical transition". *Phys Rev Lett* 88.13 (2002), p. 138101. ISSN: 0031-9007 (Print) 0031-9007 (Linking). URL: http://www.ncbi.nlm.nih.gov/pubmed/11955127.

[87] S.-H. Chen et al. "Observation of fragile-to-strong dynamic crossover in protein hydration water". *Proceedings of the National Academy of Sciences* 103.24 (June 2006), pp. 9012–9016. DOI: 10.1073/pnas.0602474103.

[88] W. Doster. "The protein-solvent glass transition". *Biochim Biophys Acta* 1804.1 (2010), pp. 3–14. ISSN: 0006-3002 (Print) 0006-3002 (Linking). DOI: 10.1016/j.bbapap.2009.06.019. URL: http://www.ncbi.nlm.nih.gov/pubmed/19577666.

[89] S. Khodadadi et al. "The origin of the dynamic transition in proteins". *The Journal of Chemical Physics* 128.19 (May 2008), p. 195106. DOI: 10.1063/1.2927871.

[90] S. Magazu, F. Migliardo, and A. Benedetto. "Puzzle of protein dynamical transition". *J Phys Chem B* 115.24 (2011), pp. 7736–43. ISSN: 1520-5207 (Electronic) 1520-5207 (Linking). DOI: 10.1021/jp111421m. URL: http://www.ncbi.nlm.nih.gov/pubmed/21612196.

[91] Andrew L. Lee and A. Joshua Wand. "Microscopic origins of entropy, heat capacity and the glass transition in proteins". *Nature* 411.6836 (May 2001), pp. 501–504. DOI: 10.1038/35078119.

[92] J.H. Roh et al. "Onsets of Anharmonicity in Protein Dynamics". *Phys. Rev. Lett.* 95 (2005), pp. 38101–38104.

[93] Liang Hong et al. "Elastic and Conformational Softness of a Globular Protein". *Physical Review Letters* 110.2 (Jan. 2013). DOI: 10.1103/physrevlett.110.028104.

[94] M. Tehei, R. Daniel, and G. Zaccai. "Fundamental and biotechnological applications of neutron scattering measurements for macromolecular dynamics". *Eur Biophys J* 35.7 (2006), pp. 551–8. ISSN: 0175-7571 (Print) 0175-7571 (Linking). DOI: 10.1007/s00249-006-0082-6. URL: http://www.ncbi.nlm.nih.gov/pubmed/16868745.

[95] U. Lehnert et al. "Thermal motions in bacteriorhodopsin at different hydration levels studied by neutron scattering: Correlation with kinetics and light-induced conformational changes". *Biophysical Journal* 75.4 (1998), pp. 1945–1952. ISSN: 0006-3495. URL: %3CGo%20to%20ISI%3E://WOS:000076144000033.

[96] J. M. Zanotti, M. C. Bellissent Funel, and S.H. Chen. "Experimental evidence of a liquid-liquid transition in interfacial water". *Europhys. Lett.* 71 (2005), pp. 91–97.

[97] S. Combet and J. M. Zanotti. "Further evidence that interfacial water is the main "driving force" of protein dynamics: a neutron scattering study on perdeuterated C-phycocyanin". *Phys Chem Chem Phys* 14.14 (2012), pp. 4927–34. ISSN: 1463-9084 (Electronic) 1463-9076 (Linking). DOI: 10.1039/c2cp23725c. URL: https://www.ncbi.nlm.nih.gov/pubmed/22388956.

[98] J. D. Nickels et al. "Dynamics of protein and its hydration water: neutron scattering studies on fully deuterated GFP". *Biophys J* 103.7 (2012), pp. 1566–75. ISSN: 1542-0086 (Electronic) 0006-3495 (Linking). DOI: 10.1016/j.bpj.2012.08.046. URL: http://www.ncbi.nlm.nih.gov/pubmed/23062349.

[99] C. David Andersson et al. "Influence of Enantiomeric Inhibitors on the Dynamics of Acetylcholinesterase Measured by Elastic Incoherent Neutron Scattering". *The Journal of Physical Chemistry B* 122.36 (Aug. 2018), pp. 8516–8525. DOI: 10.1021/acs.jpcb.8b05485.

[100] J. Peters et al. "Dynamics of human acetylcholinesterase bound to non-covalent and covalent inhibitors shedding light on changes to the water network structure". *Phys Chem Chem Phys* 18.18 (2016), pp. 12992–3001. ISSN: 1463-9084 (Electronic) 1463-9076 (Linking). DOI: 10.1039/c6cp00280c. URL: http://www.ncbi.nlm.nih.gov/pubmed/27109895.

[101] S. Sacquin-Mora et al. "Probing the flexibility of the bacterial reaction center: The wild-type protein is more rigid than two site-specific mutants". *Biochemistry* 46.51 (2007), pp. 14960–14968. ISSN: 0006-2960. DOI: Doi10.1021/Bi7004416. URL: %3CGo%20to%20ISI%3E://WOS:000251734500023.

[102] M. Erlkamp et al. "Influence of pressure and crowding on the sub-nanosecond dynamics of globular proteins". *J Phys Chem B* 119.14 (2015), pp. 4842–8. ISSN: 1520-5207 (Electronic) 1520-5207 (Linking). DOI: 10.1021/acs.jpcb.5b01017. URL: https://www.ncbi.nlm.nih.gov/pubmed/25781064.

[103] A. Paciaroni et al. "Conditioning action of the environment on the protein dynamics studied through elastic neutron scattering". *European Biophysics Journal with Biophysics Letters* 35.7 (2006), pp. 591–599. ISSN: 0175-7571. DOI: DOI10.1007/s00249-006-0073-7. URL: %3CGo%20to%20ISI%3E://WOS:000240362200007.

[104] Wolfgang Doster and Marcus Settles. "Protein–water displacement distributions". *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* 1749.2 (June 2005), pp. 173–186. DOI: 10.1016/j.bbapap.2005.03.010.

[105] M.A. González. "Force fields and molecular dynamics simulations". *École thématique de la Société Française de la Neutronique* 12 (2011), pp. 169–200. DOI: 10.1051/sfn/201112009.

[106] M. Karplus and G. A. Petsko. "Molecular dynamics simulations in biology". *Nature* 347 (1990), pp. 631–639.

[107] B. R. Brooks et al. "CHARMM: The biomolecular simulation program". *Journal of Computational Chemistry* 30.10 (July 2009), pp. 1545–1614. DOI: 10.1002/jcc.21287.

[108] Romelia Salomon-Ferrer, David A. Case, and Ross C. Walker. "An overview of the Amber biomolecular simulation package". *Wiley Interdisciplinary Reviews: Computational Molecular Science* 3.2 (Sept. 2012), pp. 198–210. DOI: 10.1002/wcms.1121.

[109] James C. Phillips et al. "Scalable molecular dynamics with NAMD". *Journal of Computational Chemistry* 26.16 (2005), pp. 1781–1802. DOI: 10.1002/jcc.20289.

[110] Chris Oostenbrink et al. "A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6". *Journal of Computational Chemistry* 25.13 (2004), pp. 1656–1676. DOI: 10.1002/jcc.20090.

[111] A. D. MacKerell et al. "All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins†". *The Journal of Physical Chemistry B* 102.18 (Apr. 1998), pp. 3586–3616. DOI: 10.1021/jp973084f.

[112] Jay W. Ponder and David A. Case. "Force Fields for Protein Simulations". In: *Protein Simulations*. Elsevier, 2003, pp. 27–85. DOI: 10.1016/s0065-3233(03)66002-x.

[113] Alexander D. Mackerell, Michael Feig, and Charles L. Brooks. "Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations". *Journal of Computational Chemistry* 25.11 (Aug. 2004), pp. 1400–1415. DOI: 10.1002/jcc.20065.

[114] X. Huang et al. "Anisotropic surface strain in single crystalline cobalt nanowires and its impact on the diameter-dependent Young's modulus". *Nanoscale* 5.23 (2013), pp. 11643–8. ISSN: 2040-3372 (Electronic) 2040-3364 (Linking). DOI: 10.1039/c3nr81284g. URL: https://www.ncbi.nlm.nih.gov/pubmed/24096984.

[115] Robert B. Best et al. "Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone phi, ψ and Side-Chain χ1 and χ2 Dihedral Angles". *Journal of Chemical Theory and Computation* 8.9 (Aug. 2012), pp. 3257–3273. DOI: 10.1021/ct300400x.

[116] George A. Kaminski et al. "Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides†". *The Journal of Physical Chemistry B* 105.28 (July 2001), pp. 6474–6487. DOI: 10.1021/jp003919d.

[117] Bertrand Guillot. "A reappraisal of what we have learnt during three decades of computer simulations on water". *Journal of Molecular Liquids* 101.1-3 (Nov. 2002), pp. 219–260. DOI: 10.1016/s0167-7322(02)00094-6.

[118] C. Vega et al. "What ice can teach us about water interactions: a critical comparison of the performance of different water models". *Faraday Discussions* 141 (2009), pp. 251–276. DOI: 10.1039/b805531a.

[119] H. J. C. Berendsen et al. "Molecular dynamics with coupling to an external bath". *The Journal of Chemical Physics* 81.8 (Oct. 1984), pp. 3684–3690. DOI: 10.1063/1.448118.

[120] Hans C. Andersen. "Molecular dynamics simulations at constant pressure and/or temperature". *The Journal of Chemical Physics* 72.4 (Feb. 1980), pp. 2384–2393. DOI: 10.1063/1.439486.

[121] Shuichi Nosé. "A unified formulation of the constant temperature molecular dynamics methods". *The Journal of Chemical Physics* 81.1 (July 1984), pp. 511–519. DOI: 10.1063/1.447334.

[122] S. Nosé. "A molecular dynamics method for simulations in the canonical ensemble". *Molecular Physics* 52.2 (June 1984), pp. 255–268. DOI: 10.1080/00268978400101201.

[123] William G. Hoover. "Canonical dynamics: Equilibrium phase-space distributions". *Physical Review A* 31.3 (Mar. 1985), pp. 1695–1697. DOI: 10.1103/physreva.31.1695.

[124] Glenn J. Martyna, Michael L. Klein, and Mark Tuckerman. "Nosé-Hoover chains: The canonical ensemble via continuous dynamics". *The Journal of Chemical Physics* 97.4 (Aug. 1992), pp. 2635–2643. DOI: 10.1063/1.463940.

[125] Giovanni Bussi, Davide Donadio, and Michele Parrinello. "Canonical sampling through velocity rescaling". *The Journal of Chemical Physics* 126.1 (Jan. 2007), p. 014101. DOI: 10.1063/1.2408420.

[126] Simone Melchionna, Giovanni Ciccotti, and Brad Lee Holian. "Hoover NPT dynamics for systems varying in shape and size". *Molecular Physics* 78.3 (Feb. 1993), pp. 533–544. DOI: 10.1080/00268979300100371.

[127] Evangelia D. Chrysina, Keith Brew, and K. Ravi Acharya. "Crystal Structures of Apo- and Holo-bovine α-Lactalbumin at 2.2-Å: Resolution Reveal an Effect of Calcium on Inter-lobe Interactions". 275.47 (July 2000), pp. 37021–37029. DOI: 10.1074/jbc.m004752200.

[128] Hans W. Horn et al. "Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew". *The Journal of Chemical Physics* 120.20 (May 2004), pp. 9665–9678. DOI: 10.1063/1.1683075.

[129] David Van Der Spoel et al. "GROMACS: Fast, flexible, and free". *Journal of Computational Chemistry* 26.16 (2005), pp. 1701–1718. DOI: 10.1002/jcc.20291.

[130] Mark James Abraham et al. "GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers". *SoftwareX* 1-2 (Sept. 2015), pp. 19–25. DOI: 10.1016/j.softx.2015.06.001.

[131] Ulrich Essmann et al. "A smooth particle mesh Ewald method". *The Journal of Chemical Physics* 103.19 (Nov. 1995), pp. 8577–8593. DOI: 10.1063/1.470117.

[132] Berk Hess. "P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation". *Journal of Chemical Theory and Computation* 4.1 (Jan. 2008), pp. 116–122. DOI: 10.1021/ct700200b.

[133] Pan Tan et al. "Gradual Crossover from Subdiffusion to Normal Diffusion: A Many-Body Effect in Protein Surface Water". *Physical Review Letters* 120.24 (June 2018). DOI: 10.1103/physrevlett.120.248101.

[134] E. Pellegrini et al. "MDANSE a versatile application for analysing molecular dynamics data". *manuscript in preparation* (). URL: www.mdanse.org (visited on 11/07/2018).

[135] Mounir Tarek and Douglas J. Tobias. "Environmental Dependence of the Dynamics of Protein Hydration Water". *Journal of the American Chemical Society* 121.41 (Oct. 1999), pp. 9740–9741. DOI: 10.1021/ja990643i.

[136] B. Aoun et al. "Direct comparison of elastic incoherent neutron scattering experiments with molecular dynamics simulations of DMPC phase transitions". *European Physical Journal E* 39.48 (2016), pp. 1–10. ISSN: 1292-8941. DOI: 10.1140/epje/i2016-16048-y. URL: %3CGo%20to%20ISI%3E://WOS:000374985200001.

[137] L. Van Hove. "Correlations in Space and Time and Born Approximation Scattering in Systems of Interacting Particles". *Phys. Rev.* 95.1 (1954), pp. 249–262.

[138] Hans Frauenfelder, Robert D. Young, and Paul W. Fenimore. "The role of momentum transfer during incoherent neutron scattering is explained by the energy landscape model". *Proceedings of the National Academy of Sciences* 114.20 (2017), pp. 5130–5135. ISSN: 0027-8424. DOI: 10.1073/pnas.1612267114. eprint: http://www.pnas.org/content/114/20/5130.full.pdf. URL: http://www.pnas.org/content/114/20/5130.

[139] Gerald R. Kneller. "Franck–Condon picture of incoherent neutron scattering". *Proceedings of the National Academy of Sciences* (2018). ISSN: 0027-8424. DOI: 10.1073/pnas.1718720115. eprint: http://www.pnas.org/content/early/2018/08/29/1718720115.full.pdf. URL: http://www.pnas.org/content/early/2018/08/29/1718720115.

[140] Genscript®. *Amino Acid Chart and Reference Table*. URL: https://www.genscript.com/amino_acid_structure.html (visited on 09/07/2018).

[141] C. D. Andersson et al. "Changes in dynamics of [small alpha]-chymotrypsin due to covalent inhibitors investigated by elastic incoherent neutron scattering". *Phys. Chem. Chem. Phys.* 19 (37 2017), pp. 25369–25379. DOI: 10.1039/C7CP04041E. URL: http://dx.doi.org/10.1039/C7CP04041E.

[142] Liina Kangur et al. "Structural stability of human butyrylcholinesterase under high hydrostatic pressure". *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* 1867.2 (Feb. 2019), pp. 107–113. DOI: 10.1016/j.bbapap.2018.11.001.