Study Conformational Dynamics of Intrinsically Disordered Proteins by Single-Molecule Spectroscopy

Dissertation

For the award of the degree

"Doctor rerum naturalium"

of the Georg-August-Universität Göttingen

within the doctoral program PBCS

of the Georg-August University School of Science (GAUSS)

submitted by

Man Zhou

From China

Göttingen 2016

Members of the Thesis Committee:

Prof. Dr. Jörg Enderlein (Reviewer)

3rd Institute of Physics - Biophysics, Faculty of Physics,

Georg-August-Universität Göttingen

Prof. Dr. Holger Stark (Reviewer)

Department of Structural Dynamics

Max Planck Institute for Biophysical Chemistry

Prof. Dr. Heinz Neumann

Department of Applied Synthetic Biology

Georg-August-Universität Göttingen

Date of oral examination: 01.07.2016

Affidavit

Hereby, I declare that the presented thesis has been written independently and with no other sources and aids than quoted.

Göttingen, 11.05.2016

Contents

Abstract

Intrinsically disordered proteins (IDPs) are proteins which lack a well-defined three-dimensional structure. The abundance and functional significance of IDPs has been recognized only recently. Due to their properties, IDPs play an important role in cellular functions. They serve as flexible inter-protein linkers, and participate in molecular recognition, molecular assembly, cellular signaling and regulation, or protein modification. Thus, genetically encoded alterations of IDPs are involved in many diseases, such as cancer, cardiovascular disease, amyloidosis, or neurodegeneration. Therefore the study and characterization of the conformational dynamics of IDPs are important to better understand the underlying mechanisms which lead to various pathologies.

FG repeats, rich in phenylalanine (F) and glycine (G), are one particular type of IDPs. FG repeats are located in the central channel of the nuclear pore complex (NPC), and they control the molecular transport between the nucleus and the cytoplasm. The way how FG repeats form or/and function as highly selective barriers in NPCs is not clear. In this thesis, the conformational dynamics of one FG repeats, Nsp1, is investigated by photo-induced electron-transfer fluorescence correlation spectroscopy (PET-FCS) and molecular dynamics simulation (MD simulation). Combination of PET-FCS and MD simulation offers a more comprehensive understanding of the relationship between functional mechanism and conformational dynamics of IDPs.

The results from PET-FCS measurements indicate that the N-terminus of Nsp1 tends to be more flexible than the C-terminus. Furthermore, short Nsp1 fragments (up to 50 amino acids) at low concentration (100 µM) do not tend to aggregate under physiological condition. These data indicate that the interaction between short FG repeats is not strong enough to solely generate the barrier. Interaction between long FG repeats or between many different kinds of FG repeats may also make a distinct contribution to the highly selective barrier of the NPC. To support MD simulations, hydrodynamic radii of various IDPs, FG repeats, GS repeats, Prothymosin alpha, and Integrase were measured by dual-focus fluorescence correlation spectroscopy (2fFCS). The

data showed that the conformations obtained by the force field CHARMM 22* and a charm-modified TIP3P water model agrees best with the experimental data. These results are important for further force field developments of MD simulation for IDPs in the future.

1. Introduction

One of the central dogmas in molecular biology is that the function of a protein depends critically on its three-dimensional, genetically determined structure. With other words, it is assumed that a well-defined three-dimensional structure is crucial for a proper function of proteins, which requires that a newly synthesized protein needs to fold into a unique and proper three-dimensional shape [1]. For example, the famous "lock-and-key model," proposed by chemist Emil Fischer in 1894, postulates that both the enzyme and substrate require specific complementary geometric shapes that fit exactly into each other [2]. However, recently exceptions to the rule have been discovered in the form of so called intrinsically disordered proteins (IDPs) that lack fixed three-dimensional structures, but have nonetheless fundamental and essential cellular functions.

1.1 Properties of intrinsically disordered proteins (IDPs)

The structure of IDPs is very dynamic and fluctuates rapidly through a broad range of conformations, which covers a continuum in conformational space ranging from extended random coils to collapsed globules [3]. Over the past decades, numerous proteins have been recognized and identified as unfolded or intrinsically disordered based on experimental studies [4, 5]. Even a new database, called DisProt, was specifically established to collect information about disordered proteins, and it now contains more than 1300 identified IDPs [6]. Based on the prediction of some bioinformatics programs, 40% of all the human proteins are estimated to contain IDP fragments that are disordered under physiological conditions and fold into ordered structures only upon binding to their cellular targets [5], [7]-[9].

Compared to proteins with well-defined three-dimensional structure such as globular proteins, IDPs show the following properties: Firstly, IDPs have a biased amino acid composition and low sequence complexity, as well as low proportions of bulky hydrophobic amino acids, and high proportions of charged and hydrophilic amino acids [3]. Secondly, IDPs contain multiple

interaction motifs that mediate binding to diverse targets, thus they commonly function as central hubs in signaling networks. In this way, the use of the same IDP in multiple pathways can save resources for the cell.

IDPs can be generally classified into two groups: largely disordered and partially disordered proteins. The latter have extensive disordered regions (30-40 residues) embedded in the folded structure [10]. Two examples of largely and partially disordered IDPs are shown in figure 1.1.1. Thylakoid soluble phosphoprotein (TSP9), which is a plant-specific protein in the photosynthetic thylakoid membrane, contains a largely soluble flexible protein chain [11, 12]. Small ubiquitin-related modifier 1 (SUMO-1), which is a human-specific protein involved in the post-translational modification system, contains two disordered regions at two termini [13]-[15]. In general, partially disordered regions are more common and have been observed in numerous X-ray and NMR studies.

(A)   **(B)**
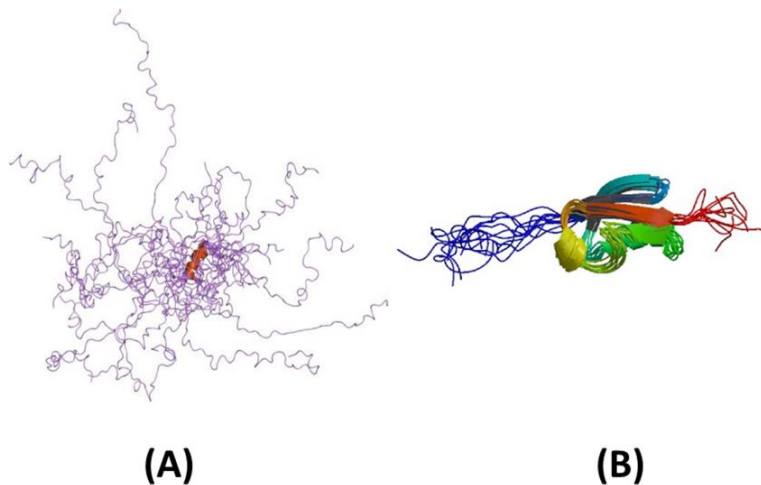
Figure 1.1.1 Examples of Structures for largely and practically disordered IDPs. (A) largely disordered structure of Thylakoid soluble phosphoprotein TSP9 (taken from ref. [12]), which contains a fully flexible, disordered protein chain (pink) with a small α-helix core (red). 20 structural conformations of the flexible parts are shown. (B) Partially disordered structure of the small ubiquitin-related modifier SUMO-1 (taken from ref. [16]), which contains an α-helix and a β-sheet in the core, and a small disordered part at the N- and C-terminus (blue and red). 10 structural conformations are shown in the structure.

1.2   Protein dynamics

By now, it is well accepted that many functional proteins do not need to adopt a fixed three-dimensional structure. Instead, they fluctuate between many conformational sub-states, and these fluctuations are key to their cellular functions. For example, protein dynamics has been found to play a major part in most tightly regulated biological processes such as signaling, transcription, and chromatin remodeling. The study of protein dynamics is concerned with the transitions between these conformational sub-states. To describe these transitions, the concept of an "energy landscape" was introduced to describe the relative probabilities of the conformational sub-states and the energy barriers between them [17]. This concept was proposed more than 30 year ago by Frauenfelder and co-workers [18]. In their study, the binding kinetics of carbon monoxide and oxygen to myoglobin was investigated by flash photolysis. Based on the observation of multiple energy barriers and non-exponential kinetics below the temperature of 230 K, the energy landscape concept was developed: it contains multiple conformational sub-states and barriers between energy wells of each sub-state. A sub-state is defined as a minimum in the energy surface, whereas a transition state is the maximum between the wells. Although these sub-states coexist in equilibrium with different populations, conformational transitions that cross the energy barriers between different sub-states can occur. A schematic diagram of a generic energy landscape is shown in figure 1.2.1. It shows multiple timescales and motional amplitudes of protein dynamics. Two large conformational sub-states (A&B) are present, and each sub-state is again subdivided into a large number of smaller sub-states. Transitions between tier-0 sub-states (A and B) are rare because of the high free-energy barriers. Transitions between tier-1 and tier-2 are more frequent because of the lower free-energy barriers.
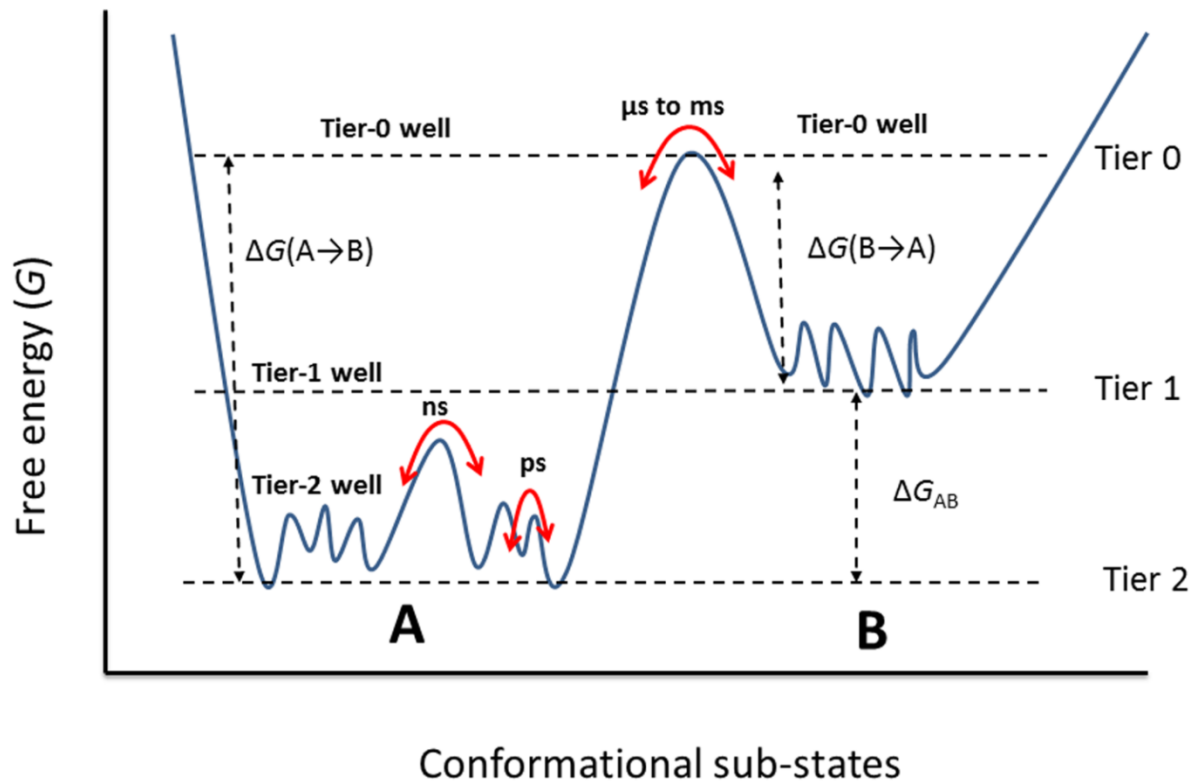
Figure 1.2.1 The energy landscape of protein dynamics. A one-dimensional cross section through the energy landscape is shown. The energy landscape is represented as a hierarchical organization of free energy wells (*i.e.*, the smallest tier-2 wells are within tier-1 wells and the tier-1 wells are within the largest tier-0 wells). The tier-2, tier-1 and tier-0 sub-states (A and B) are located within respective free energy wells. The populations of the tier-0 sub-states A and B are defined based on their difference in free energy (ΔG_{AB}). The barrier between these sub-states (ΔG) determines the transition rate. Lower tiers describe faster dynamics and higher tiers describe slower dynamics. Tier-2 dynamics on the ps timescale and tier-1 dynamics on the ns timescale are regarded as fast dynamics. Tier-0 dynamics on the µs timescale are regarded as slow dynamics. A change in the system, *e.g.* ligand binding, protein mutation and changes in external conditions, will shift the equilibrium between sub-states.

Based on the timescale, the protein dynamics can be classified into following classes:

(1) Fast dynamics (ps to ns).

Fast dynamics can be described as small-scale conformational fluctuations between different sub-states separated by a very low-energy barrier. Previous studies demonstrated that fast dynamics depends strongly on hydration level and temperature [19]-[21]. Local atomic fluctuation and rotations of side chain on the picosecond timescale, and loop motions on the nanosecond timescale can be regarded as fast dynamics (tier-2 and tier-1 dynamics shown in figure 1.2.1). Some protein dynamics processes take also place on ps to ns timescales. For example, the relaxation time of hydration water dynamics in hydrated proteins at room temperature is about 10-50 ps [22]. MD simulation studies suggest that jump-like motions in proteins are on the ns timescale. These motions represent large atomic displacements about 8 Å [23]. Recent studies suggest that protein dynamics can be as slow as ~100 ns; these dynamics are slaved to the fluctuations in the bulk solvent and the hydration shell, which consists of about 2 layers of water that surround proteins [24].

(2) Slow dynamics (μs)

The dynamics on a μs timescale includes the main backbone motions and transitions between well-defined basins in the potential energy landscape of the protein structure, shown as tier-0 dynamics in figure 1.2.1 [25]. Many biological processes in the cell, such as chromatin remodeling, transcription and translation, happen on this time scale. Therefore, the protein dynamics on the μs timescale is directly connected to protein function [22].

(3) Dynamics at longer times

Many single molecular studies suggest that some protein dynamics occur on much longer timescales than μs [26]-[28]. These dynamics are controlled by the fast dynamics described above. Although the atomistic picture of many macromolecules is well-understood, the relationship between dynamics at atomistic level and dynamics on these longer timescales remains speculative.

1.3   Important functions of IDPs

IDPs play an important role in cellular functions, broadly including the flexible linking different proteins or protein parts, molecular recognition, molecular assembly, cellular signaling and regulation, or protein modification [29]. These functions can be structurally grouped into six different functional classes: entropic chains, display sites, chaperones, effectors, assemblers, and scavengers [30]. A single protein may consist of several disordered regions that belong to different functional classes [31]. The following text will give examples of these important functions.

(1) Entropic chains benefit directly from their conformational disorder, *i.e.* they function without fixed structures. Most flexible linkers between protein domains are partially disordered. These linkers are often composed of flexible residues like glycine and serine so that the adjacent protein domains are free to move and do not interfere with each other. One example is the family of flexible linkers in the 70 kDa subunit of replication protein A (RPA70) [32].

(2) Partially disordered IDPs which serve as display sites are heavily involved in cellular signaling and regulation. The multiplicity of protein-interaction motifs in IDPs and their capacity for regulation through post-translational modification bring important advantages to the signaling and regulation process. For example, the same amino acid sequence can be used in different contexts and in response to different signals to turn different signaling pathways on or off and thus induce different cellular responses. These attributes also contribute to pathway crosstalk and to the operation of positive- and negative-feedback circuits [3]. Well-characterized samples are histones, p53, and the cyclin-dependent kinase regulator p27 [33]-[35].

(3) Chaperones are proteins that assist RNA and proteins to reach their functionally folded states [36,37]. Disordered regions constitute over one-half of the sequences of RNA chaperones and over one-third of protein chaperones [38, 39].

(4) Effectors interact with other proteins and modify their activity. Upon binding to their interaction partners, IDPs often undergo a disorder-to-order transition [40, 41]. Typical

examples of effectors are p21 and p27, which regulate different cyclin-dependent kinases that are responsible for the control of cell-cycle progression in mammals [42].

(5) Disordered assemblers bring together multiple binding partners to promote the formation of high-order protein complexes [43, 44], such as the ribosome complex [45], activated T-cell receptor complex [30], the RIP1/RIP3 necrosome complex [46], or the transcription pre-initiation complex [47]. The open structure of disordered assemblers is largely preserved upon scaffolding their partner proteins, resulting in a large binding interface that enables multiple proteins to bind to a single IDP [48].

(6) IDPs play also a role in the scavenger system, which stores and neutralizes small ligands. For example, chromogranin A is a random coil in both the isolated form as well as the cellular environment. It functions as a scavenger by storing ATP and adrenaline in the medulla of the adrenal gland [49].

Due to the important functions of IDPs, mutations in IDPs or alterations in their abundance cause mis-identification, mis-signaling, or unnatural or nonnative folding in cellular processes [50]. These aberrations are commonly associated with disease, such as cancer, cardiovascular disease, amyloidosis, neurodegenerative diseases, and diabetes [51, 52]. For example, the tumor suppressor p53, a transcription factor, is involved in cancer development [53]. Breast cancer type 1 susceptibility protein (BRCA1) is involved in many signaling processes such as DNA damage response, transcription, and cell cycle checkpoint control. The mutations in BRCA1 predispose to breast, ovarian and other cancers [54]. The aberration of the blood protein thrombin, which is involved in coagulation and clotting, will cause cardiovascular disease [51]. α-synuclein and τ protein are involved in neurodegenerative diseases (*e.g.* Alzheimer) characterized by abnormal protein aggregation and deposition in the brain [55, 56]. Amylin is a polypeptide produced by pancreatic β-cells. It has several functions associated with the regulation of energy metabolism. Mutation of amylin is associated with the development of type II diabetes [57].

1.4   Methods to characterize IDPs

Nowadays, there is no doubt that IDPs plays an important role in cellular functions and that the aberration of IDPs is involved in many diseases. Therefore, it is very important to study the structural property of IDPS which could serve as potential pharmaceutical targets. However, due to the lack of well-defined structures, the characterization of the conformational structure, properties, and function of IDPs is very challenging.

Over the last couple of decades, several biophysical methods have been developed to characterize IDPs. These are NMR spectroscopy, X-ray crystallography, small-angle X-ray scattering (SAXS), circular dichroism (CD) spectroscopy, fluorescence spectroscopy, and protease sensitivity assays [1]. Each method has strong and weak points.

NMR spectroscopy is a powerful technique to determine the three-dimensional structure of molecules. Generally, it determines the magnetic properties of certain atomic nuclei that can provide detailed information about the chemical environment and the structure of molecules. Since it does not require crystallization, NMR spectroscopy is a suitable technique to study IDPs and in particular to develop an ensemble description of IDPs. Under favorable conditions, NMR can even provide motional information on a residue-by-residue basis by performing pulse sequence experiments on a variety of different isotopic labels [58]. For example, a particularly useful method is ^{15}N-^1H heteronuclear NOE measurement, which gives positive values for ordered residues and negative values for disordered residues [59]. Therefore, high structural resolution can be achieved down to the single residues level. However, relatively few IDP structures have been characterized by NMR compared with ordered proteins, which indicates the critical issue of NMR: it requires high sample concentrations (> 1mM) which often lead to aggregation of IDPs. The high cost and the time-consuming isotopic labeling, as well as the complex setup and analysis, do also limit the application of NMR to the study of IDPs.

X-ray crystallography is a useful tool to identify the three-dimensional molecular structure of a crystal. Generally, purified proteins at high concentration are crystallized, and the crystals are exposed to an X-ray beam. Then, the resulting diffraction patterns can be processed to obtain

the structural information. In X-ray crystallography, which uses the coherent scattering of X-rays, the absence of a scattering signal from a protein structure signals the presence of disordered regions [60]. However, this may also be caused by crystal defects or even by unintentional proteolytic removal during protein purification. Moreover, most of the time, obtaining a high-quality crystal suitable for X-ray scattering is very difficult and is thus the core bottleneck of X-ray crystallography. Even more, the successful formation of protein crystals can be hindered by the presence of highly flexible and disordered regions [60].

Small-angle X-ray scattering (SAXS) provides only low-resolution information on the shape and structural transitions of IDPs in solution. Generally, SAXS experiments are simple: the sample solution is placed into a quartz capillary illuminated by a collimated monochromatic X-ray beam, and subsequently, the intensity of scattered X-rays is recorded by an X-ray detector. The scattering pattern of pure solvent is collected as well and subtracted from the pattern of the sample solution. From the resulting difference scattering pattern, the overall shape and size of the protein of interest is reconstructed [61]. With high-intensity synchrotron sources, time-resolved SAXS studies can also yield unique information about kinetics and interactions [62]. However, SAXS has many restrictions [61]: (1) It requires high sample concentrations (typically > 1 mg/mL). (2) The sample must be highly pure. (3) Sample aggregation must be strictly avoided. (4) Accurate concentration determination is crucial for good data analysis. (5) X-ray-induced radiation damage arises from bond breakage and free radical formation in the samples and can be a source of radiation-induced aggregation. Therefore, due to these limitations, SAXS is often complemented with other techniques such as NMR, circular dichroism (CD) spectrometry, or fluorescence spectroscopy to study IDPs [63].

Circular dichroism (CD) spectroscopy can provide information about the secondary structure of proteins and thus distinguish between ordered and disordered proteins [1]. However, like all spectroscopic techniques, the CD signal reflects an average of the entire molecular population; thus, it cannot determine which specific residues are involved in the secondary structure. Therefore, this method cannot provide quantitative structural information and needs to be combined with other methods.

Last but not least, protease sensitivity assays are used to study protein structure and their flexibility by determining possible digestion sites [64]. However, the requirement for protease-sensitive residues limits the demarcation of order/disorder boundaries which can be discerned by this method [1].

1.4.1 Molecular dynamics (MD) simulation to study IDPs

Molecular dynamics (MD) simulation is a computational technique to investigate the relationship between the structure and function of biological macromolecules at the atomic level. MD simulation was first introduced by Alder and Wainwright in the late 1950s and by Rahman in the early 1960s to study the atomic movement of liquids [65, 66]. Due to the advancements in computer technology and algorithmic improvements, MD has become a powerful tool in many areas of physics and chemistry. Since the 1970s, MD has been used widely to study the structure and dynamics of macromolecules, such as proteins and nucleic acids.

In molecular dynamics simulation, a "force field" is required to describe the relationship between structure and energy of the system of interest. Simply, the force field describes variations in energy due to bond and angle stretching, torsional rotations, dispersion, exchange and long-range electrostatics. Therefore, both the protein and the solvent have to be taken into account with atomic detail to determine all these parameters. Common used bio-molecular force fields are included in the MD packages CHARMM [67, 68], AMBER [69], and OPLS [70]. Some modern force fields describe small, globular proteins quite well: NMR observables computed for these proteins agree with experimental data within reasonable error [71, 72]. However, in a recent comparison of force fields for folded proteins, Amber ff99sb*-ildn and CHARMM22* were the only two force fields consistent with experimental data [73].

A variety of computational methods have been developed to study structural properties of IDPs over the last couple of decades. These methods can be broadly classified into two types: (1)

those that use experimental data to guide ensemble generation or selection, and (2) those that generate ensembles of IDPs *de novo, i.e.* without using experimental data as an input.

One example of the first type uses experimental data as restraints in the simulations. For instance, NMR chemical shift restraints and distance restraints based on paramagnetic relaxation enhancement (PRE) were used in molecular dynamics simulations of the disordered states of α-synuclein [74, 75]. The second type has been used extensively to study structural properties of IDPs *de novo*. A variety of simulation methods such as Monte Carlo simulations, meta-dynamics, or replica exchange have been used to study IDPs [76]-[82]. All these methods have different levels of representation such as coarse-grained, implicit solvent, or all-atom with explicit water. There are two main challenges in the *de novo* simulations. First, extensive simulations are needed to ensure that relevant regions of conformational space are adequately sampled. Although this requirement applies to all bio-molecular simulations, it presents a particularly formidable challenge in the case of IDPs due to their high conformational heterogeneity. Second, and more importantly, the suitability of existing force fields for IDP simulations is not well characterized.

The structure of IDPs cannot be described by a single conformation, but it has instead to be described by an ensemble of interconverting conformations [83]. Many force field modifications have been introduced to improve the accuracy of simulations of globular proteins over the last decades, but not for IDPs. Recently, several force field modifications have improved the balance of secondary structure propensities in order to be able to fold proteins of multiple structural classes, and all-atom simulations are increasingly being used to obtain ensembles of IDPs [78, 79, 84, 85].

It is difficult to draw conclusions on the accuracy of IDP simulations from the contradictory findings reported so far. On the one hand, good agreement between computed and measured experimental observables was observed for some IDP simulations [78]-[80], [82, 86, 87]. On the other hand, the structural accuracy of the unfolded state and of the IDP structural ensembles which have been obtained by using several widely-used force fields has been called into question [80, 82, 83] [88]-[93]. For example, the disordered arginine/serine (RS) peptide, which

plays an important role in RNA metabolism, RNA splicing, and RNA export, is a well-characterized IDP by NMR experiments [94, 95]. Recent studies by Rauscher *et al.* [83] showed that simulated ensembles from eight force fields of the RS peptide vary substantially (see figure 1.4.1). Ensembles obtained with different force fields exhibit marked differences in chain dimensions, hydrogen bonding, and secondary structure content. These differences are unexpectedly large: changing the force field is found to have a stronger effect on secondary structure content than changing the entire peptide sequence [83].
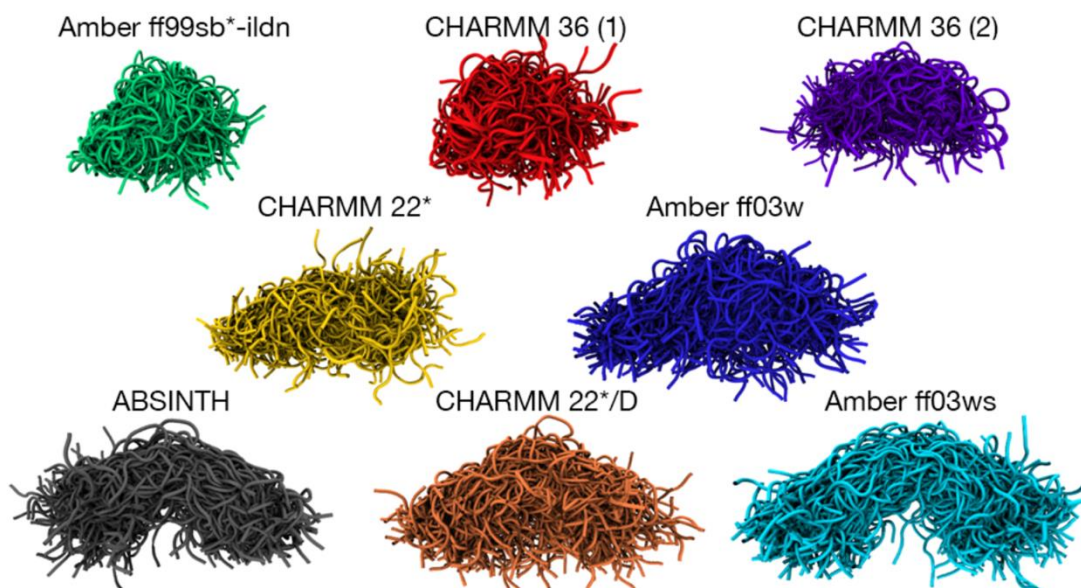
Figure 1.4.1 Ensembles of RS repeats generated by eight all-atom empirical force fields. (taken from ref. [83]) one hundred structures of RS peptide selected at random form are shown. The amino acid sequence of RS peptide is "GAMGPSYGRSRSRSRSRSRSRSRS" [83].

Overall, while these recent studies demonstrate a substantial interest in obtaining accurate ensembles for IDPs using all-atom simulations, there is currently no consensus on the best force fields to be used, or the suitability of any force fields, for this purpose. Information on the accuracy of IDP ensembles is sparse and contradictory, which may be due to a combination of multiple factors: (1) inadequate conformational sampling of IDP ensembles, (2) comparison to models derived from primary experimental data, and (3) comparison to relatively few observables.

1.5 The Nuclear Pore Complex and FG repeats

Nuclear pore complexes (NPCs) are large protein assemblies which cross the nuclear envelope. They form a selective barrier that regulates bidirectional exchange between the nucleus and the cytoplasm [96]. The proteins that compose the NPCs are known as nucleoporins. About 32 different kinds of nucleoporins (Nups) have been discovered in NPCs and can be generally classified into two groups: structural Nups and intrinsically disordered Nups [97]. The intrinsically disordered Nups contain characteristic repetitive sequence motifs consisting of the amino acids phenylalanine and glycine (F&G) [96], which form the so-called FG repeats. The FG repeats are mainly located in the central channel of NPCs and constitute flexible filaments (see figure 1.5.1). FG repeat domains are estimated to account for 12 - 20 % of the mass of NPCs [98, 99]. Generally, up to 50 FG repeats can be found in each Nup, in which a short cluster of hydrophobic residues is surrounded by a more hydrophilic spacer sequence [98].
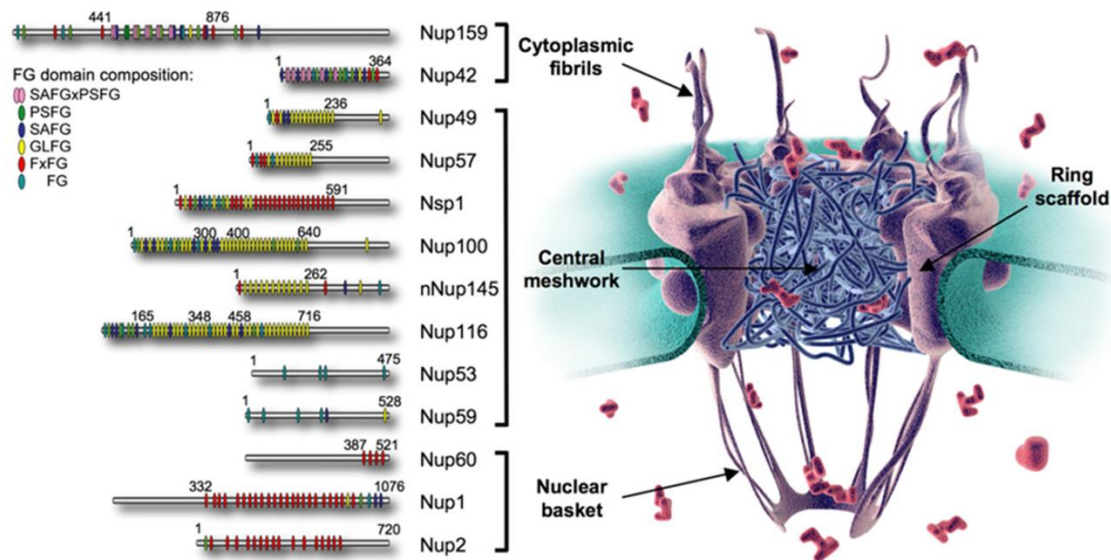
Figure 1.5.1 Schematic illustration of the nuclear pore complex (NPC) and locations of FG repeats (adapted from ref. [100]) A cross-section of the NPC is shown with ring scaffold, cytoplasmic fibrils, nuclear basket, and the meshwork of FG repeats in the central channel. Main FG Nups of *Saccharomyces cerevisiae* are shown on the left with vertical tick marks corresponding to the respective FG units. The numbers above each Nup correspond to the number of amino acid residues flanking the FG units.

1.5.1 Function of FG repeats in Nuclear Pore Complex

The nuclear pore complex allows for the free passage of small molecules, but it suppresses the passage of macromolecules (*e.g.* larger than 30 kDa). However, the NPC also permits the rapid passage of even larger cargoes if they are bound to appropriate nuclear transport receptors (NTRs), in which case the speed of passage is enhanced by factors of about 100 to >1000 fold in comparison to the passage of inert objects of similar size [101]-[103].

FG repeats are abundant in the "passageway" of NPCs and dominantly contribute to the functional barrier of NPCs [104]-[106]. How do FG repeats form this highly selective barrier?

A number of models have been proposed to answer this question. The two most prominent ones are the virtual gate model [107], and the selective phase model [108]. Schematic illustrations of the two models are shown in figure 1.5.2.

The virtual gate model assumes that the brush-like behavior of FG repeats is sufficient to repel the inert molecules away from the central channel, while the NTRs can overcome this barrier by binding to FG units. The model is mainly supported by the observation that one human FG repeat (Nup 153 anchored at the nuclear basket) shows an extended polymer brush-like behavior when probed by atomic-force microscopy [109].

The selective phase model assumes that FG repeats contain many cohesive FG units. The FG repeats which can form a hydrogel are called cohesive sequences, while others which cannot are called non-cohesive sequences. Cohesive FG units can bind to each other weakly via hydrophobic interaction, and such interactions can result in a sieve-like FG hydrogel (the selective phase) that allows for the passage of small molecules but suppresses the passage of larger ones [97]. However, NTRs can bind these hydrophobic clusters, and this binding is essential for facilitated passage of cargo-NTR complexes.
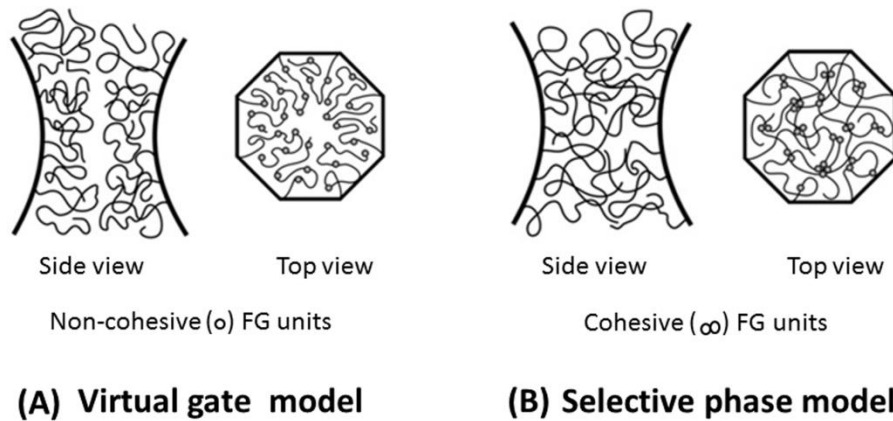
Figure 1.5.2 Models of FG repeats function as constructional elements of the highly selective barrier (adapted from ref. [100]) (A) Virtual gate model proposes that FG repeats are non-cohesive brushes that repel the entry of non-karyophilic proteins [110] into the NPC through Brownian motion. (B) Selective phase model proposes that weakly hydrophobic interaction between cohesive FG units can form a sieve-like meshwork that sieves particles by size exclusion.

The selective phase model is mainly supported by the observation that FG repeats of several yeast nucleoporins can form an elastic hydrogel *in vitro* showing fascinating properties [98, 111, 112]. For example, recent studies showed that FG hydrogel is stable even up to 95°C [111], and an *in vitro* experiments showed that cargo-NTR complexes can enter the hydrogel about 25,000 faster than inert objects [98]. A hydrogel is shown in figure 1.5.3. This homogeneous FG hydrogels can be prepared only with very high concentration, which needs to be equal or larger than some critical concentration, the so-called saturation limit. The saturation limit is supposed to be about 2.2 mM for FG repeat domains [98]. Hydrogel formation also requires an extreme chemical environment and special treatment [111].

(A) Wild-type Nsp1 repeat   (B) F→S mutated Nsp1 repeat

Figure 1.5.3 Gel formation of Nsp1 (taken from ref. [111]). (A) Wild type Nsp1 can form transparent and elastic hydrogel. Inset illustrates how interactions bet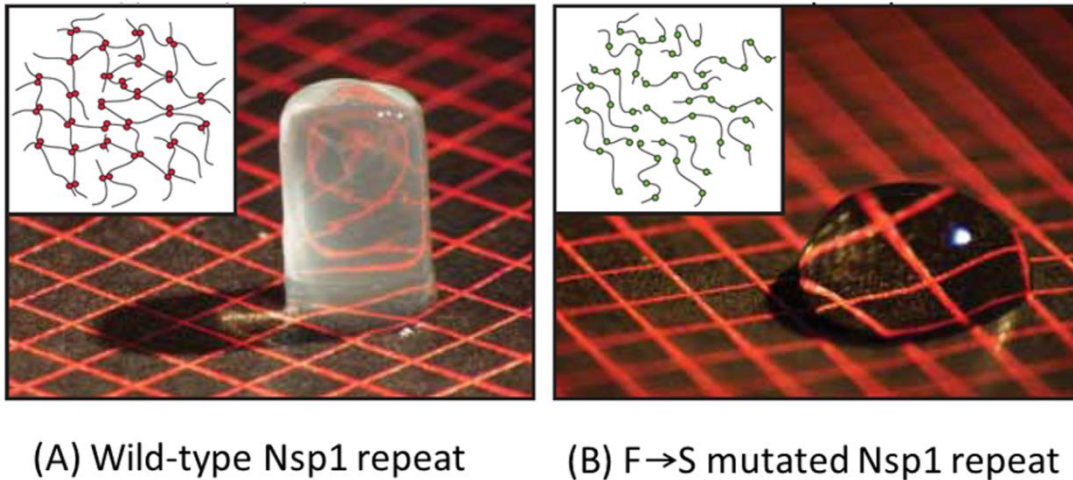ween the hydrophobic clusters (in red) cross-link the repeat domains into a hydrogel. (B) The F→S mutated repeat domains remain liquid after identical treatment. The green dots represent the mutated hydrophobic cluster, which cannot bind each other. The detailed amino acid sequences are given in the appendix.

Due to the complexity of the preparation and the non-physiological conditions for the gel formation, the selective model still remains speculative. The key difference between these two models is whether FG repeats can bind to each other or not [100]. In particular, the virtual-gate model proposes that FG repeats do not bind to each other, whereas the selective-phase model suggests that binding is necessary. In order to investigate the weak binding between FG repeats, sensitive assays were developed [100]. The general principle of these assays is shown in figure 1.5.4.

These assays were developed especially for detecting low affinity interactions in real time under equilibrium binding conditions. Firstly, Glutathione *S*-transferase (GST) and cyan fluorescent protein (CFP) labelled FG repeats were expressed and purified separately. Then, GST labelled FG repeats (GST-FG) were immobilized at high concentrations on the surface of Sepharose beads. Finally, these beads were mixed with soluble fluorescent CFP-FG repeats and examined directly under a microscope (see figure 1.5.4). When a CFP-FG repeat binds to an immobilized GST-FG repeat, the interaction is visible as a halo of fluorescence around the dark

bead. According to this study, FG repeats of many Nups do interact with each other; however, they also interact with other structural Nups which do not contain FG units. At the same time, some FG repeats can form hydrogels, but do not bind themselves or interact with others. For example, short fragment Nsp1 (1-603) can form a hydrogel, but do not bind to other FG repeats or to each other. To summarize, some data of these sensitive assays support selective-phase model, but some data do also support the virtual gate model.
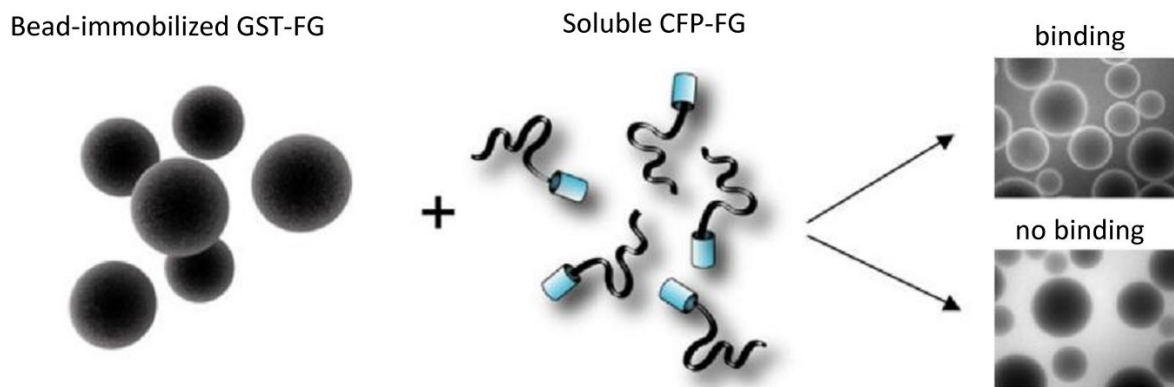
Figure 1.5.4 *In vitro* **assay that detects low affinity protein interactions of FG repeats.** (taken from ref. [100]) Soluble CFP-FG repeats fusions are mixed with bead-immobilized GST-FG repeats and the interactions are visualized in a fluorescence microscope. Fluorescence around the beads indicates binding between soluble and immobilized FG repeats, whereas dark beads indicate no binding.

Nsp1 is a well-characterized FG repeats, due to its abundance in the NPC. Nsp1 is also the only one known to form two distinct complexes: the Nic96 and Nup82 complex. The Nic96 complex, which consists of Nsp1, Nup57, Nup49 and Nic96, is located on both sides of the central gated channel. The Nup82 complex, which is formed by Nsp1, Nup82 and Nup159, is found exclusively on the cytoplasmic side of an NPC. One study proposes that a third Nsp1 complex (Nsp1 interacts with Nic96) is located at the terminal ring of the nuclear basket [113].

Some studies propose that different parts of Nsp1, which show different properties in gel formation assays, play different roles in the selective phase model. Specifically, the N-terminal part of Nsp1, which is characterized by Asn-rich inter-FG spacers, forms a tough hydrogel. The C-terminal part which is comprised of charged inter-FG spacers, shows low gelation property.

The N-terminal part is rich in FG and FXFG units, whereas the C-terminal part is rich in FSFG units (see figure 1.5.5).
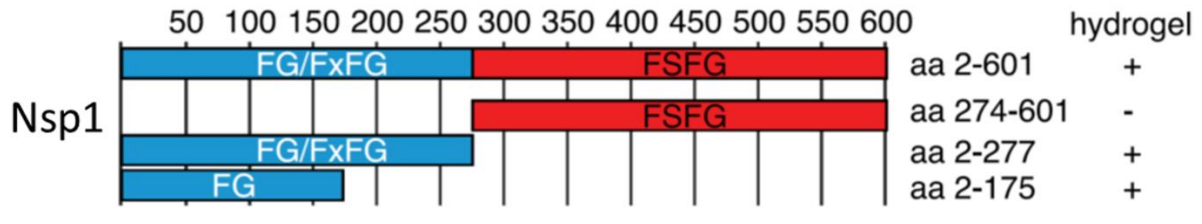
Figure 1.5.5: Bar models of Nsp1 fragments with high and low gel-forming properties. (adapted from ref. [114]). "+" and "-" represent "can" and "cannot" form hydrogel, respectively. "aa" is the abbreviation for amino acid.

To summarize, both models admit that FG repeats play an important role in the selective function of NPCs. However, the two models are contradictory with respect to how FG repeats contribute to the highly selective function of NPCs.

1.6   Goals of this study

In this thesis, the conformational dynamics of FG repeats (Nsp1) and their potential aggregation properties were investigated. In order to study the difference in the conformational dynamics between the N- and C-terminus of Nsp1, two types of Nsp1 with different numbers of FG or FSFG units in their sequence were studied, respectively.

This work was done in close cooperation with the group of Prof. Dr. Helmut Grubmüller from the Department of Theoretical and Computational Biophysics at the Max Planck Institute for Biophysical Chemistry in Göttingen. To evaluate the accuracy of existing force fields which are used in *de novo* molecular dynamics simulations, six different force fields were employed in MD simulations of the N-terminus of Nsp1, and their performance was compared. To that end, the hydrodynamic properties of Nsp1 fragments and other proteins were experimentally studied. All MD simulations were carried out in the Grubmüller group.

2. Materials & methods

2.1 Reagents and buffers

Synthetic peptides with 98% purity were purchased from GL Biochem (Shanghai, China). ProTα coupled with Atto Oxa 11 and Integrase coupled with Alexa488 were kindly provided by Prof. Dr. Ben Schuler from the Biochemistry Department of the University of Zurich. All fluorescent dyes were purchased from ATTO-TEC GmbH (Siegen, Germany). All chemicals were purchased from Sigma-Aldrich, Germany. Tissue culture chamber with a glass cover slip bottom was purchased from Sarstedt (Nümbrecht, Germany). All chromatography columns and materials used for purification were purchased from GE healthcare.

All buffers were prepared with chemicals of analytical grade. For single molecule experiments, all buffers were bleached under UV light before use. The compositions of all buffers used are listed in table 2.1.1.

Table 2.1.1 Composition of buffers used in this thesis

Name	Composition
PBS	137 mM NaCl, 3 mM KCl , 20 mM Na_2HPO4, 1.5 mM KH_2PO4
4 x PBS	548 mM NaCl, 12 mM KCl , 80 mM Na_2HPO4, 6 mM KH_2PO4
Peptide dissolving buffer	37% acetonitrile: H_2O (v/v)
Maleimide coupling buffer	4 M GdnHCl, 2 x PBS, 0.2 mM EDTA, 0.2 mM TCEP, pH=7.0
NHS ester coupling buffer	50 mM $NaHCO_3$ (pH=7.3)
His tag cleavage buffer	150 mM NaCl, 25 mM Tris-HCl, pH=8.0

2.2 Dual-Focus Fluorescence Correlation Spectroscopy (2fFCS)

2.2.1 Fluorescence

Luminescence is the emission of light by substances after absorbing light or other forms of energy. In general, luminescence can be classified into two types: fluorescence and phosphorescence, depending on the nature of the excited state (singlet vs. triplet) which are also reflected in the time scales on which they occur (nanoseconds vs. milliseconds). The excitation and emission processes can be visualized with a Jablonski diagram (see figure 2.2.1).
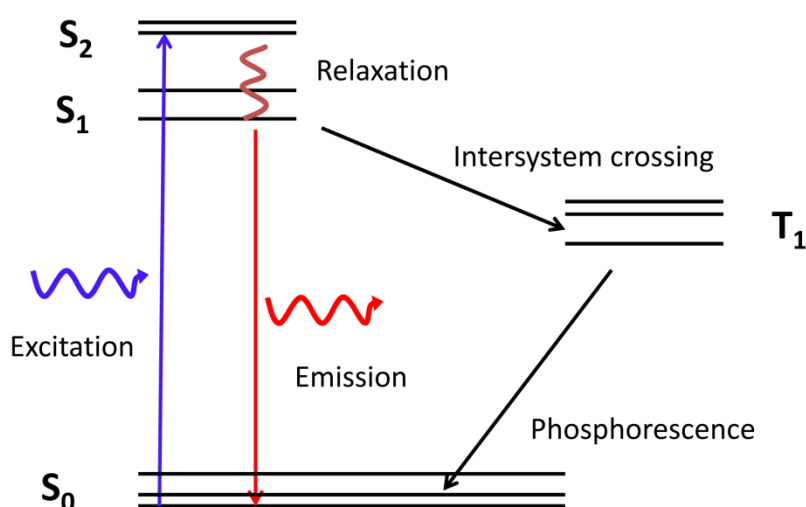
Figure 2.2.1 A typical Jablonski diagram (adapted from ref. [115]) The singlet ground, first, second and triplet electronic states are depicted as S_0, S_1, S_2 and T_1, respectively. At each energy level, a fluorophore can exist in a number of vibrational energy levels, which are depicted by the multiple lines.

After absorption of a photon, a fluorophore is excited to a higher electronic energy level $S_{n>0}$, where the fluorophore can interact with its surrounding environment and also undergoes conformational changes, which is accompanied by energy dissipation. The dissipated energy is usually given off via collisions with surrounding solvent molecules. This leads to a rapid relaxation of the exited fluorophore towards the lowest vibrational energy level of its first excited electronic S_1. This process is very fast, taking only between 10^{-14} and 10^{-11} seconds. Relaxing from the lowest excited singlet state S_1 to the ground state is accompanied by the

emission of a photon, and this is emission is called fluorescence. Compared with the vibrational relaxation in the excited state, fluorescence is a slow process which takes place on time scales on the order of 10^{-9} to 10^{-7} seconds. The emitted light has generally a lower energy than the energy of the excitation light, and has therefore a longer wavelength. The difference between excitation and emission wavelengths is called Stokes shift.

Not every excited fluorophore will emit a photon after excitation. It may also undergo intersystem crossing: the excited electron undergoes a spin flip, and the molecule switches into its first triplet state T_1. The emission from the triplet state T_1 to the ground state S_0 is called phosphorescence. Phosphorescence occurs on longer time scales (10^{-6} s), as compared to fluorescence, because the electron spin needs to flip again in order to return to the ground state S_0. Spin flip is theoretically a forbidden process (violation of angular momentum conservation) and can only occur due to weak spin-orbit coupling.

There exist also other, so called non-radiative transitions from the excited to the ground state, for which the energy of the excited state is dissipated as heat, and which reduce the chance of fluorescence emission. The ratio of the number of fluorescent transitions to the total number of absorptions is called the fluorescence quantum yield Q, which is an important and intrinsic property of fluorescent molecules. The quantum yield Q can be defined as:

$$Q = \frac{N_{em}}{N_{abs}} \tag{1}$$

where N_{em} is the number of emitted photons, and N_{abs} is the total number of absorbed photons.

The lifetime τ is another character of a fluorophore. It is defined as the average time that a molecule spends in its excited state before it returns to the ground state. Typically, this relaxation follows an exponential decay,

$$I(t) = I_0 \exp\left(-\frac{t}{\tau}\right) \tag{2}$$

where I_0 is the intensity immediately after excitation and $I(t)$ is the intensity at time t. Generally, the lifetimes of organic fluorophores are below 10 ns.

2.2.2 Fluorescence Correlation Spectroscopy (FCS)

The irregular motion of pollen seeds suspended in water was first observed under a microscope by botanist Robert Brown in 1827. Today it is well known that this so called Brownian motion is due to collisions with water molecules, which makes the particles undergo random-walk motion. The important parameter which describes Brownian motion of a particle in solution is its diffusion coefficient. For a long time, scientists had been interested in determining exact values of diffusion coefficients. For this purpose, different experimental methods have been developed. The conventional methods include the Gouy interference method, conductometric measurements, or self-diffusion measurements of isotopic solutions [116, 117]. More recent methods include dynamic light scattering (DLS), and Fluorescence Correlation Spectroscopy (FCS).

In 1906, Smoluchowski was the first to outline the relationship between Brownian motion and auto-correlation, which was later used in the statistical analysis of DLS and FCS [118]. DLS is also known as Photon Correlation Spectroscopy. Generally, when light hits small particles in solution, the particles will scatter the light into all directions. In DLS, a monochromatic and polarized laser is used to illuminate a sample solution. Then, the scattered light is passed through a polarizer and then detected by a detector. Since the scattered light can interfere either constructively or destructively, the recorded data will generate so called "speckle patterns", which appear as randomly fluctuating dark or bright spots on the detector. The intensity on the detector depends on the interference pattern created by the scattered light from all molecules in the scattering volume. Due to Brownian motion, the interference pattern will fluctuate and thus the detected intensity will change over time. Small particles diffuse rapidly and give fast fluctuations, whereas larger particles diffuse slower and give slower fluctuations. Quantitative information on the size of the particles can be obtained with a signal processing technique which is known as auto-correlation. In FCS, the signal comes from fluorescent molecules. The big advantage of fluorescence is that it allows for detecting only molecules of interest which are specifically labeled. FCS was invented in 1972 by Magde, Elson and Webb to study the kinetics of DNA-ethidium bromide interaction [119]. In 1976, a confocal microscope was first employed

for FCS by Koppel to discriminate against background fluorescence [120]. Since then, FCS has been used in many studies, such as measuring diffusion coefficients and chemical kinetics [121] as well as for monitoring conformational dynamics of proteins or DNA molecules, or to watch protein-protein interactions [122, 123].

FCS is a powerful tool to study hydrodynamic properties of biomolecules. The hydrodynamic radius R_H, which is related to the size and shape of a molecule, is related to the diffusion coefficient via the celebrated Stokes-Einstein equation [124]:

$$D = \frac{kT}{6\pi\eta R_H} \tag{3}$$

where D is the diffusion coefficient, k is the Boltzmann constant, R_H is the hydrodynamic radius, and η is the viscosity of the solvent.

The basic principle of FCS is a correlation analysis of the fluorescence intensity fluctuations in a small detection volume (on the order of a femtoliter), resulting from the diffusion of fluorescent molecules through that volume. If the concentration is low enough, the recorded fluorescent signal is strongly fluctuating when molecules diffuse in and out of the detection volume (see figure 2.2.2).
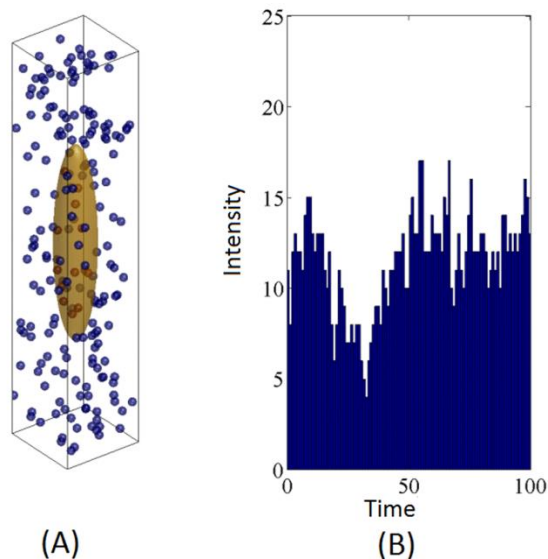
Figure 2.2.2 Scheme of fluorescence fluctuations (adapted from ref. [125]). (A) Molecules (blue) diffusing through the detection volume (yellow) are excited by a focused laser beam (red), and generate a highly fluctuating fluorescence signal. (B) Fluorescent intensity fluctuation recorded in time.

From the measured fluorescence intensity time trace, an autocorrelation function is computed. The auto-correlation function $G(\tau)$ can expressed as the temporal auto-convolution of a signal with itself and can be obtained by multiplying a measured value at a given time t and a later time $t + \tau$, normalized by the square of the average fluorescent intensity $F(t)$:

$$G(\tau) = \frac{< \delta F(t)\delta F(t + \tau) >}{< F(t) >^2} \tag{4}$$

Here, $\delta F(t)$ is the difference between the fluorescent intensity at time t and its average value $F(t)$. The angular brackets represent averaging over time t.

The auto-correlation function is a measure of the probability to detect a photon from the same molecule at a time t and at a later time $t + \tau$. The full autocorrelation contains also contributions from uncorrelated signals, *e.g.* from backscattered laser light, which do not show any lag-time (τ) dependent correlation. These uncorrelated events will contribute to a constant offset of $G(\tau)$ that is completely independent on τ. The other term contains the correlated signal, coming from photons which are emitted from one and the same molecule.

The particular expression of the correlation function which is used for data fitting depends on the shape of the detection volume. A good approximation for a confocal detection volume is a 3-D Gaussian model, which results in the following expression for the single-molecule part of the autocorrelation function [126]:

$$G(\tau) = \frac{1}{N} \left(1 + \frac{\tau}{\tau_D}\right)^{-1} \frac{1}{\sqrt{1 + \frac{\tau}{\omega^2 \tau_D}}} \tag{5}$$

Here, N is the average number of fluorescent molecules in the detection volume, τ_D is the average diffusion time which the molecules spend within the detection volume, and ω is the axial to lateral ratio of the extent of the assumedly ellipsoidal detection volume.

A typical auto-correlation curve measured for a solution of the dye Atto655 is shown on figure 2.2.3. The diffusion time τ_D is equal to several milliseconds, which is an estimate of the average time which an Atto655 molecule spends within the detection volume. As described in section 2.2.1, most fluorophores have triplet states which live for ~10^{-6} seconds, and the transition into these triplet states leads to a partial decay of the auto-correlation function on the time scale of microseconds. However, the auto-correlation curve of Atto655 is almost flat from the nanosecond to the microsecond time scale, which suggests that Atto655 has no efficient transition into the triplet state. This property is extremely important for the experimental studies presented in this thesis; the reason for this will be clarified in section 2.3.
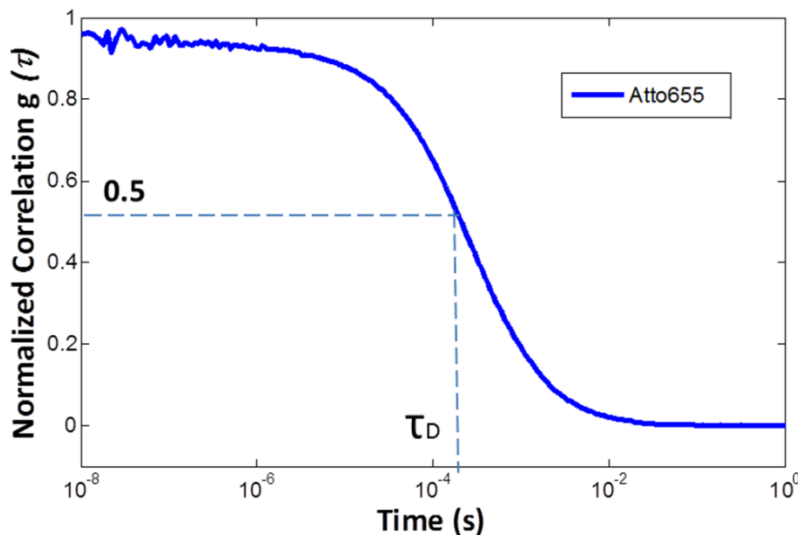
Figure 2.2.3 A typical auto-correlation curve of Atto655 measured in aqueous solution (150 mM NaCl). Auto-correlation decay on millisecond time scale is due to the diffusion of molecules into and out of the detection volume. The diffusion time τ_D of Atto655 is on a millisecond time scale.

2.2.3 FCS setup

The FCS measurement setup used in this thesis is shown in figure 2.2.4. It consists of a commercial confocal microscope (Microtime 200, PicoQuant GmbH, Berlin, Germany), equipped with a water immersion objective (UPLSAPO 60 x W, 1.2 N.A., Olympus Hamburg, Germany). Fluorescence excitation and detection are done through the same objective (epi-fluorescence set-up).

For fluorescence excitation, two pulsed red diode lasers (LDH-D-C- 640 nm, PicoQuant, Berlin, Germany) with orthogonal linear polarization are used. Both beams are coupled into a polarization-maintaining single mode fiber. At the fiber output, the light is collimated and then reflected with a dichroic mirror (FITC/TRITC, Chroma Technology, Rockingham, VT, USA) towards the objective.

Collected fluorescence is passed through the same dichroic mirror, and then focused by a tube lens through onto a 150 µm diameter confocal pinhole. After the pinhole, the light is re-

collimated, split by a 50/50 beam splitter, and focused onto two single photon avalanche diodes (tau-SPAD, PicoQuant, Berlin, Germany). Emission band-pass filters (HC692/40, Semrock, USA) are positioned in front of each detector to discriminate fluorescence against scattered light. A time-correlated single-photon counting electronics (HydraHarp 400, PicoQuant GmbH) records the detected photons of all detectors independently with an absolute temporal resolution of 4 picoseconds on a common time frame. In order to circumvent the dead time and after-pulsing effects of the APDs, the signals were recorded by two APDs.
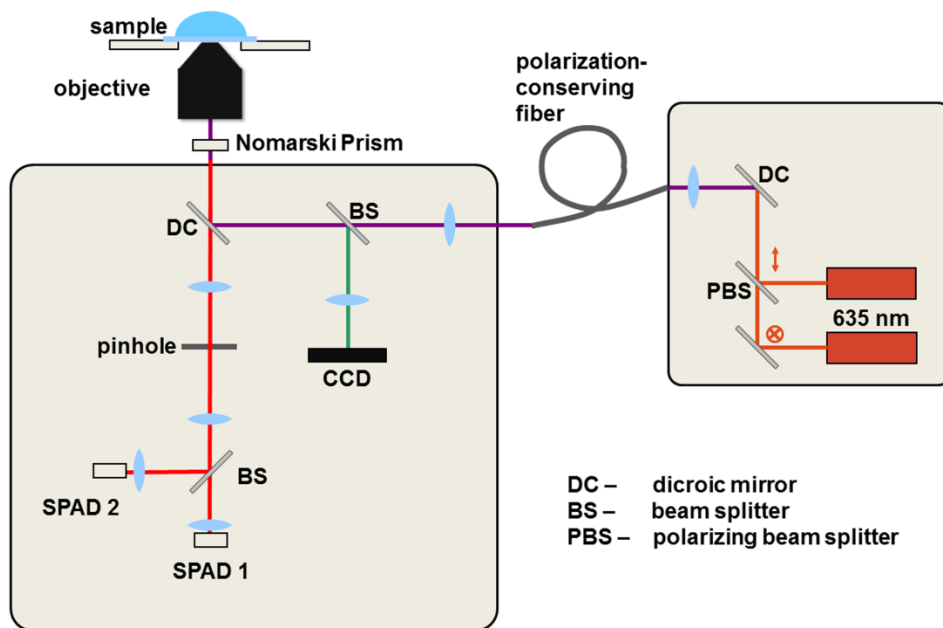
Figure 2.2.4 Scheme of FCS setup. (adapted from ref.[127])

2.2.4 Dual-focus Fluorescence Correlation Spectroscopy (2fFCS)

Conventional FCS measurements can provide absolute diffusion coefficient and concentration values only if the exact shape and size of the detection volume is known. Unfortunately, the size and shape of the detection volume is not fixed and strongly depends on many experimental parameters such as laser power, optical aberrations, or the refractive index of the sample solution [128]. Therefore, for each sample, the measurements system has to be calibrated for the specific sample conditions.

To overcome these limitations, dual-focus FCS (2fFCS) was introduced by Dertinger *et al.* [128]. In 2fFCS, a Nomarski prism is introduced into the back focal plane of the objective of a conventional FCS system. This prism deflects laser beams with different angles depending on their polarization. Two orthogonally polarized laser beams deflected through a Nomarski prism will generate two identical and overlapping foci in the focal plane of the objective (see figure 2.2.5). The distance between the two foci depends on the property of the Nomarski prism, and can be determined by performing 2fFCS with fluorescent polymer beads with known radius [129].
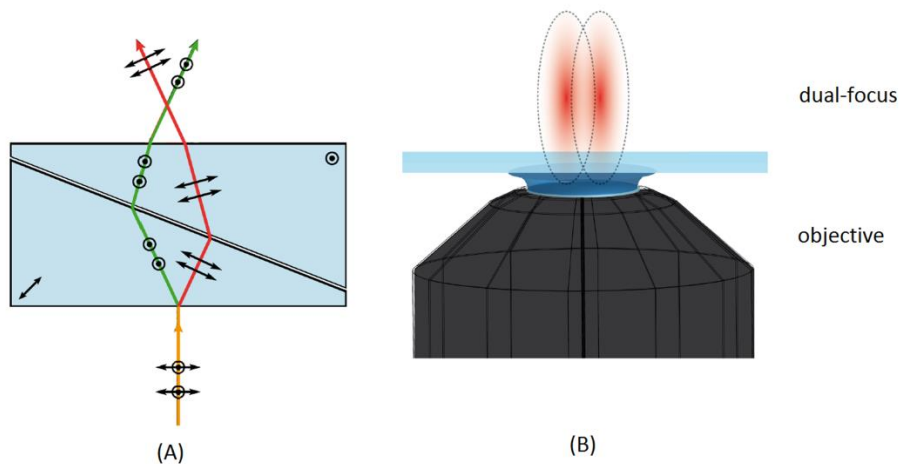
Figure 2.2.5 Generation of two overlapping foci by Nomarski prism (adapted from ref.[130]). (A) Two orthogonally polarized laser beams deflected through Nomarski prism: red is in-plane and green is out-of-plane. (B) Two foci are generated on 2fFCS setup.

Time-correlated single-photon counting (TCSPC) is a well-established technique for fluorescence lifetime measurements [131]. It is a time-measurement method which determines photon arrival times with respect to the exciting laser pluses with picosecond temporal resolution. In 2fFCS, it is used to link the detected photons to the exciting laser pulse and therefore to determine from which focus they originated.

The principle of TCSPC is shown in figure 2.2.6. In TSCPC, one detects individual photons and measures their temporal distance to the last preceding laser pulse. Usually, one uses laser excitation intensities so that the probability of detecting a photon per laser pulse is much less than one, therefore the detection of several photons per pulse can be neglected. Thus, there are many excitation periods which do not result in any photon detection event, and only few which contain one photon detection event. When a photon is detected, the time delay between its detection time and the time of the last preceding laser pulse is measured. These times are stored in a computer, and are subsequently histogrammed, which yields the so-called fluorescence decay curve.
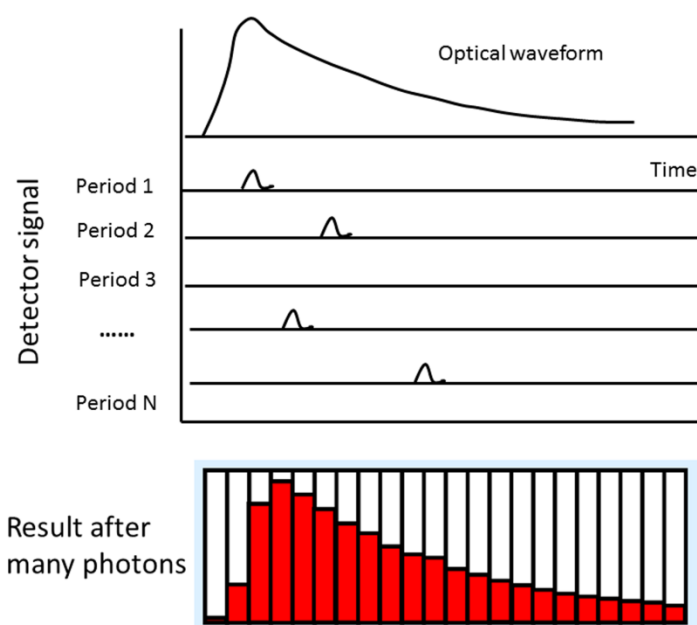
Figure 2.2.6 Principle of TCSPC measurement

A typical TCSPC histogram of a measurement on a solution of Atto655 is shown in figure 2.2.7.
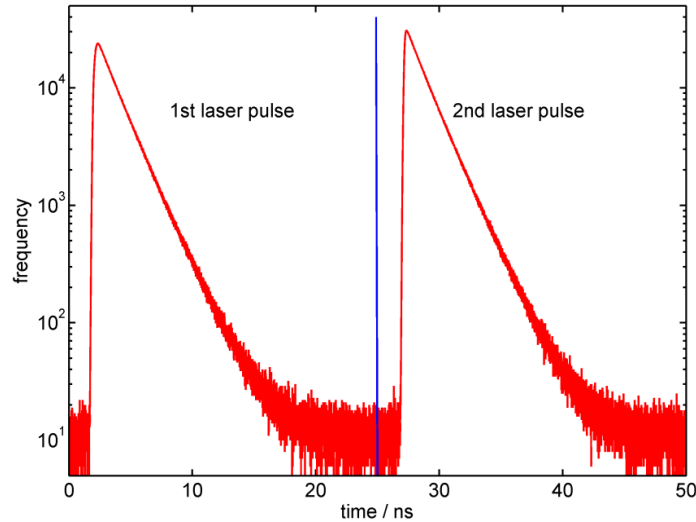
Figure 2.2.7 TCSPC histograms of Atto655 measured in aqueous solution (150mM NaCl). The photon counts in the left time window are generated by the first laser; the photon counts in the right window are generated by the second laser.

Knowing which photon comes from which focus, the auto-correlation function (ACF) for each focus and the cross-correlation function (CCF) between the two foci can be calculated. For evaluating 2fFCS measurements and to extract a correct diffusion coefficient, one needs a valid theoretical fit model. First, one has to find a realistic model for the molecule detection function (MDF) $U(\vec{r_1})$ which describes the probability to detect a photon from a molecule at a given position $\vec{r_1}$ within the detection volume [132]. Next, one needs the probability that a molecule diffuses from position $\vec{r_1}$ to $\vec{r_2}$ within time τ which is given as the fundamental solution of the diffusion equation and reads:

$$G(\vec{r}, \tau) = \frac{1}{(4\pi D\tau)^{\frac{3}{2}}} \exp\left(-\frac{|\vec{r_2} - \vec{r_1}|^2}{4D\tau}\right) \tag{6}$$

where D denotes the diffusion coefficient of the molecules. With this information, the auto-correlation $g(t)$ can be computed as the product of the probability to detect a photon from a molecule at position $\vec{r_1}$ times the probability that it diffuses from $\vec{r_1}$ to $\vec{r_2}$ within time τ, times the probability to detect a photon from a molecule at position $\vec{r_2}$. Finally, one has to average over all possible initial and final positions of the molecule. The auto-correlation $g(t)$ then reads:

$$g(t) = c\epsilon_{1,2}{}^2 \int_v d\vec{r_1} \int_v d\vec{r_2}\, U(\vec{r_2}) \frac{1}{(4\pi Dt)^{\frac{3}{2}}} \exp\left(-\frac{|\vec{r_2} - \vec{r_1}|^2}{4D\tau}\right) U(\vec{r_1}) \qquad (7)$$

where c is the molecular concentration, v the detection volume, and $\epsilon_{1,2}$ are the total detection efficiencies in the two detection volumes.

The CCF correlates photons from one focus with those detected in the other focus. Taking into account the distance between the two foci, a lateral shift parameter needs to be introduced into equation (7). Then, the CCF can be given as:

$$g(t,\delta) = c\epsilon_{1,2}{}^2 \int_v d\vec{r_1} \int_v d\vec{r_2}\, U(\vec{r_2}) \frac{1}{(4\pi Dt)^{\frac{3}{2}}} \exp\left(-\frac{|\vec{r_1} - \vec{r_2} - \hat{x}\delta|^2}{4D\tau}\right) U(\vec{r_1}) \qquad (8)$$

where $\hat{x}$ represents the unit vector along the x-axis, and δ is the lateral shift between both foci. This shift is a fixed parameter of the 2fFCS-setup and is well-known.

An appropriate model function for the MDF is crucial for accurate 2fFCS data analysis. It has been shown that a Gauss-Lorentzian function is a sufficiently accurate approximation for the MDF of a confocal microscope [128]:

$$U(r) = \frac{\kappa(z)}{\omega^2(z)} \exp\left(-\frac{2}{\omega^2(z)}(x^2 + y^2)\right) \qquad (9)$$

Here, x and y represent the transversal coordinates perpendicular to the optical axis z. In each plane perpendicular to the optical axis, the MDF is approximated by a Gaussian distribution with width $\omega(z)$ and amplitude $(z)/\omega^2(z)$. The width of the Gaussian $\omega(z)$ is given by

$$\omega(z) = \omega_0 \sqrt{\left(1 + \left(\frac{\lambda_{ex}z}{\pi\omega_0{}^2 n}\right)^2\right)}, \qquad (10)$$

and the amplitude function $\kappa(z)$ is given by

$$\kappa(z) = 1 - \exp\left(-\frac{2a^2}{R^2(z)}\right) \qquad (11)$$

where $R(z)$ is given by

$$R(z) = R_0 \sqrt{1 + \left(\frac{\lambda_{em}z}{\pi R_0^2 n}\right)}. \qquad (12)$$

Here, λ_{ex} and λ_{em} are the excitation and emission wavelengths, n is the refreactive index of the immersion medium (water), and a is the radius of the confocal aperture divided by the magnification. ω_0 and R_0 are structural parameters that need to be obtained when fitting the CCF against experimental data.

By substituting equations (9) - (12) into equation (8), and after integrating over the transverse coordinates, the CCF expression is given as:

$$
\begin{aligned}
&g(t, \delta) \\
&= \frac{c\epsilon_{1,2}{}^2}{4} \sqrt{\frac{\pi}{Dt}} \int dz_1 \int dz_2 \, \frac{\kappa(z_1)\kappa(z_2) \exp\left[-\dfrac{(z_2-z_1)^2}{4Dt} - \dfrac{2\delta^2}{8Dt + \omega^2(z_1) + \omega^2(z_2)}\right]}{8Dt + \omega^2(z_1) + \omega^2(z_2)}
\end{aligned}
\tag{13}
$$

This expression cannot be more simplified and has to be evaluated numerically. Data fitting is performed with a least-squares minimization method, with $c\epsilon_{1,2}{}^2$, the diffusion coefficient D, and a_0 and R_0 as fit parameters. The distance δ between the two foci of the 2fFCS measurements is fixed by the properties of the Nomarski prism.

A typical fit of the model against experimental data is shown in figure 2.2.8.
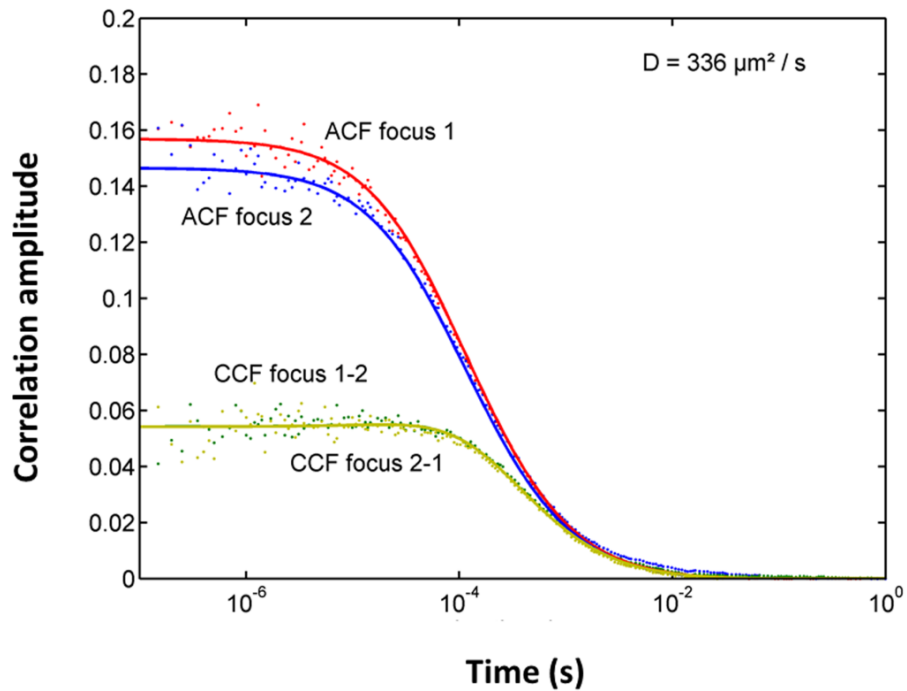
Figure 2.2.8 Typical 2fFCS curves of the dye Atto655 in aqueous solution (150 mM NaCl): The red curve represents the ACF of focus 1; the blue curve the ACF of focus 2; and the yellow curve represents the CCF between the two foci. The diffusion coefficient of Atto655 as measured by 2fFCS is 336 μm^2/s (aqueous solution with 150 mM NaCl at room temperature, 22°C).

2.3 Photo-induced Electron Transfer Fluorescence Correlation Spectroscopy (PET-FCS)

2.3.1 Fluorescence quenching mechanism

Fluorescence quenching is a process that decreases the fluorescence intensity of a fluorophore. The substance that absorbs the excitation energy from the fluorophore is called the quencher. For collisional quenching, the fluorophore and the quencher have to come into molecular contact, allowing the electron orbitals of both molecules to overlap. Since the electron clouds are strongly localized, the typical distance of quenching is therefore on the order of van-der-Waals radii. The rate of quenching depends on the extent of the electron orbitals of the fluorophore and the quencher. Since the electron density falls off very rapidly with the distance from the nuclei of the molecules (typically exponentially decreasing), the quenching rate k shows the following exponentially decaying dependence on inter-molecule distance:

$$k(r) = A \cdot \exp[-\beta(r - r_c)] \tag{14}$$

where r is the center-to-center distance, and r_c is the closest distance at molecular contact. A has a typical value of about 10^{13}/s., and values for β are typically around 1 Å^{-1}.

When the distance r is larger than several Å, this interaction becomes insignificant and quenching is negligible. Therefore, the distance β^{-1} can be regarded as a "quenching distance".

Besides collisional quenching, there exist also the following four mechanisms, which have been well studied.

(1) Intersystem crossing.

As already mentioned in section 2.2.1, the excited electron of a fluorophore may undergo a spin forbidden transition from the S_1 to the long-lived T_1 state. This transition results in the loss of fluorescence. Quenching by heavy atoms halogens and oxygen is thought to occur by intersystem crossing [115].

(2) Dexter electron transfer.

In the Dexter electron transfer, the excited electron of the fluorophore is transferred to the quencher, and then an electron of the quencher is transferred back to the fluorophore. This electron transfer leaves the fluorophore in the ground state and the quencher in an excited state.
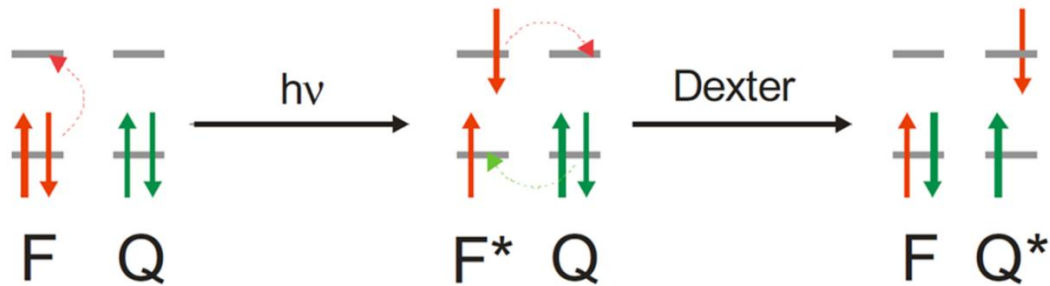
Figure 2.3.1 Dexter electron transfer mechanism of quenching (taken from ref.[133]). F and Q represent fluorophore and quencher, respectively.

(3) Fluorescence resonance energy transfer (FRET).

FRET is a process where the energy from a donor molecule is transferred to an accepter molecule through non-radiative dipole-dipole coupling. FRET can be regarded as a quenching mechanism because it results in a fluorescence loss of the donor. FRET can be used to measure intermolecular distances from 2 nm to 10 nm and, it is widely used as a "spectroscopic ruler" in the life-sciences [134]. Over the last couple of decades, FRET has become a powerful technique to quantitatively measure the distance between two different fluorophores that are conjugated to target molecules.

The FRET efficiency depends on the donor-to-acceptor separation r and is described with the following famous equation:

$$E = \frac{1}{1 + (\frac{r}{R_0})^6}$$
(15)

where R_0 is the Förster distance between donor and accepter, which is the distance where the energy transfer efficiency has fallen off to 50 % from its maximal value. The Förster distance depends on the overlap between the donor emission spectrum and the accepter absorption spectrum, their molecular orientations, and the quantum yield of the donor.

(4) Photo-induced electron transfer (PET).

In contrast to the Dexter mechanism, which does not create charged molecules, Photo-Electron Transfer (PET) creates a charged fluorophore via an electron transfer to or from a quencher. In PET, the excited fluorophore can be either the electron donor or accepter. The direction of electron transfer in the excited state depends on the redox potential between the fluorophore and the quencher. If the fluorophore is the electron donor, PET can be regarded as oxidative. If the fluorophore is the electron acceptor, PET can be regarded as reductive. In oxidative PET, after absorption of a photon, the excited electron of the fluorophore can relax to the energetically lowest unoccupied molecular orbital (LUMO) of the quencher. In this way, a positively charged fluorophore and a negatively charged quencher are generated. In reductive PET, the energy level of the quencher lies above the fluorophore. Therefore, the electron will be transferred from the quencher to the fluorophore. This process generates a negatively charged fluorophore and a positively charged quencher. In most cases, the excited state of the organic fluorophore (*e.g.* rhodamine, bodipy, or oxazine derivatives) has a lower energy than the quencher and thus acts as electron acceptor. For example, the oxazine dye Atto655 can act as an electron acceptor from the amino acid tryptophan (Trp).
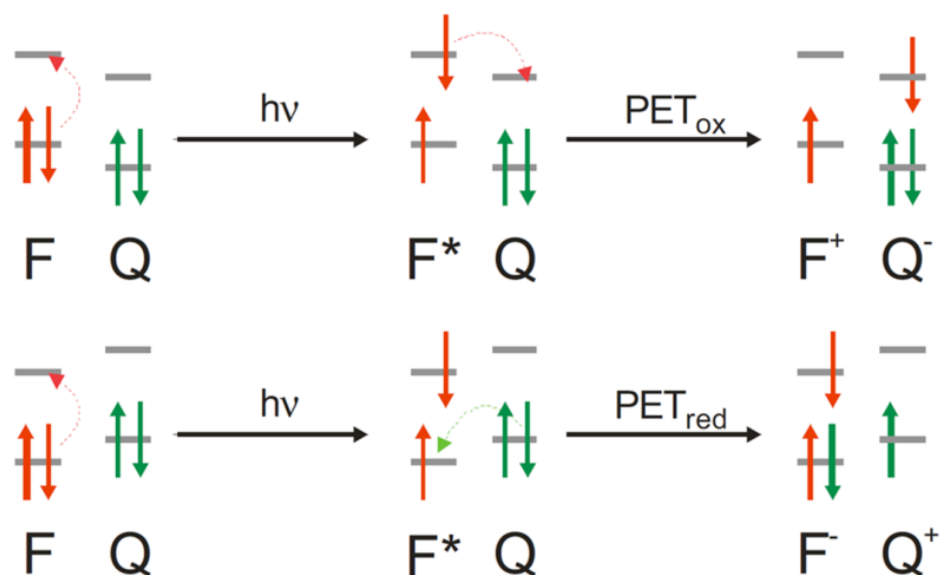
Figure 2.3.2 Principle of photo-induced electron transfer (PET) (adapted from ref.[133]). F and Q represent the fluorophore and the quencher, respectively. PET_{ox} and PET_{red} represent oxidative and reductive PET, respectively. If the redox potential of the fluorophore is higher than that of the quencher, oxidative PET can occur. In contrast, if the redox potential of the fluorophore is lower than that of the quencher, reductive PET can occur.

The PET efficiency can be estimated by the free energy change ΔG during PET. For example, the free energy change ΔG for reductive PET is given by the Rehm-Weller equation [135]:

$$\Delta G = e[E(Q/Q^+) - E(F/F^-)] - \Delta G_{00} - \frac{e^2}{\varepsilon \cdot d} \tag{16}$$

where $E(Q/Q^+)$ and $E(F/F^-)$ describe the redox potentials of fluorophore and quencher, respectively, and ΔG_{00} is the transition energy from ground state to excited state of the fluorophore. The last term of the Rehm-Weller equation is the solvent-dependent Coulombic interaction energy, where d is the distance between the two charges, and ε is the dielectric constant of the solvent. Normally, the contribution of the term $e^2/\varepsilon \cdot d$ to ΔG is negligible.

The oxidation potential of Trp is 0.6-1.0 V (vs. saturated calomel electrode, SCE) [136]; the reduction potential of Atto655 is -0.42 V (vs. SCE), and ΔG_{00} of Atto655 is 1.86 eV (vs. SCE) [137]. Therefore, the energy change ΔG of the Atto655/Trp pair is between -0.44 eV to -0.74 eV, which indicates that the PET reaction from Trp to Atto655 releases energy and can occur spontaneously.

2.3.2 PET-FCS model for conformational dynamics study

Previous studies have revealed that among all amino acids, only Tryptophan (Trp) can specifically and efficiently quench Atto655 under physiological condition [138]. Thus, Atto655 and Trp can be used to monitor the conformational dynamics of peptides and proteins.

For PET-FCS, Atto655 and Trp are introduced at two specific sites of a peptide. The quenching dynamics between Att655 and Trp can be used for investigating the conformational dynamics of the peptide. The flexible motion of the peptide's backbone will lead to a repeated contact between the dye and Trp depending, and the contact formation rate will depend on the backbone distance between the Atto655/Trp pair. The quenching distance of Atto655 and Trp is about 5.5 Å [139]. After excitation, Atto655 can either emit one photon or be quenched. Separation of Atto655 and Trp (distance > quenching distance 5.5 Å; open-state) will result in a photon emission of Atto655; contact formation between them (distance< 5.5 Å; closed-state) will result in quenching of Atto655. Therefore, the contact formation and dissociation dynamics can be modeled with a simple two-state kinetic model, with an open and a closed state, and with associated contact formation and dissociation rate constants k_+ and k_-. This is visualized in figure 2.3.3.

open state   closed state
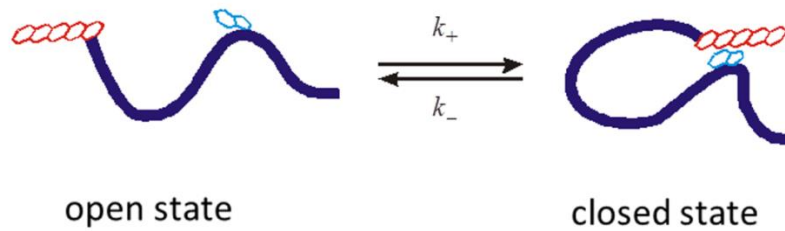
Figure 2.3.3: Illustration of a PET reporter system for studying intra-chain contact formation (adapted from ref. [140]). The quencher, here Tryptophan, is shown in blue and Atto655 in red. The rate constants between the two conformational states are indicated as k_+ and k_-. Open state (left) and closed state (right) depend on the distance between Atto655 and Trp.

The kinetic equation for the above reactions can be expressed as:

$$\frac{dF}{dt} = -k_+ F + k_- D \tag{17}$$

where *F* is the probability to find the system in the open state, and *D* the probability to find it in the open state. Obviously, one has *F* + *D* = 1.

The solution of Eq. (17) with the initial condition that the system starts in the open state, *F(0)* = 1, is found to be

$$F(t) = \frac{k_-}{k_+ + k_-} + \frac{k_+}{k_+ + k_-} \exp[-(k_+ + k_-)t] \tag{18}$$

In conventional FCS, a 3-D Gaussian model is commonly used for modeling the diffusion-related part of the auto-correlation function (see section 2.4.1). To simplify the data analysis for the kinetics of intra-chain contact formation, the diffusion-related part of the auto-correlation $G(\tau)$ will be modeled with a simple 2-D diffusion model given as:

$$G(\tau) = \frac{1}{N}\left(1 + \frac{\tau}{\tau_D}\right)^{-1} \tag{19}$$

where τ_D represents the diffusion time, and *N* is the number of fluorescent molecules in the detection volume.

Assuming that the time scale of the conformational dynamics is well separated from the time scale on which diffusion through the detection volume occurs, the full correlation function can be found as

$$G(\tau) = \frac{1}{N}\left(1 + \frac{\tau}{\tau_{\mathrm{D}}}\right)^{-1}\left[1 + K\exp\left(-\frac{\tau}{\tau_{\mathrm{R}}}\right)\right] \tag{20}$$

where the amplitude K is given equal to k_+/k_-, and the relaxation time τ_{R} is equal to $(k_+ + k_-)^{-1}$.

The contact formation rate k_+ and the dissociation rate k_- can be obtained from fitting experimental data using Equation 20. Figure 2.3.4 shows a PET-FCS correlation curve that includes diffusion and intra-chain contact formation kinetics. The time scale of intra-chain contact formation kinetics ranges from several nanoseconds to a few microseconds. Usually, one would have to consider also the triplet state dynamics of a fluorophore when fitting FCS curves. The triplet state dynamics would typically take place on the same time scale as the contact pair formation kinetics that one wants to study. Fortunately, the dye Atto655 does not show any discernable triplet state dynamics in FCS measurements, at least when excited with an excitation power of less than 2 mW (200 µW was used in this thesis) [138]. Therefore, the property that Atto655 does not show any discernable triplet state dynamics (demonstrated in section 2.2.2) makes it an ideal dye for PET-FCS studies.
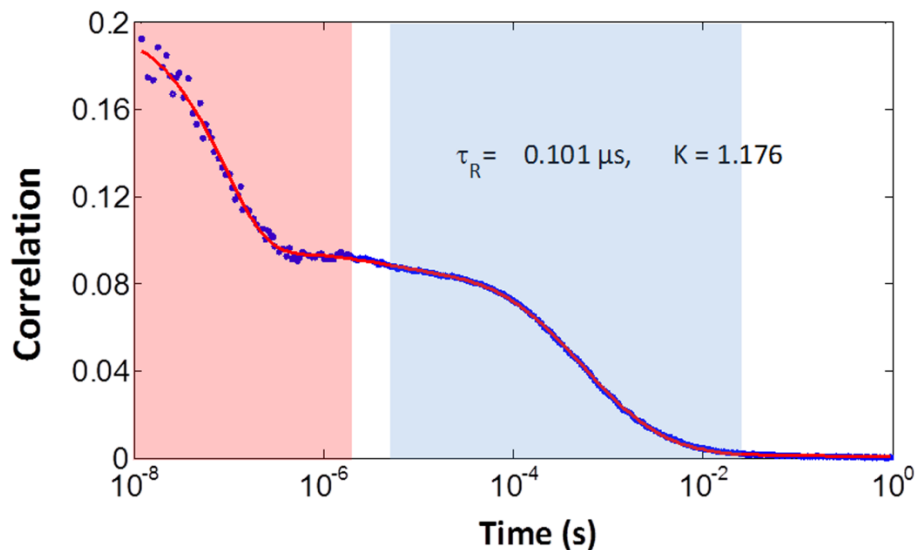
Figure 2.3.4 A typical PET-FCS curve of one 3FG repeats construct. Red background indicates time scales of PET dynamics between nanoseconds to microseconds; blue background indicates time scales of diffusion in milliseconds. Blue dots represent experimental data; red is the fitted curve. Based on the fitted decay time of τ_R = 0.101 μs and the amplitude $K = 1.176$, the association rate k_+ and dissociation rate k^- are calculated as 5.35 ×10^6 /s and 4.55×10^6 /s, respectively.

PET-FCS and Fluorescence Resonance Energy Transfer (FRET) are the commonly used single molecular fluorescence techniques to investigate protein dynamics. Compared to FRET, the big advantage of PET-FCS is that it does not require site-specific labelling of two fluorophores. Only one fluorophore is required for PET-FCS, and the quencher is an intrinsic Trp within the protein or peptide chain. PET-FCS requires contact formation at distances below the van-der-Waals radius, and yields thus only information about contact pair formation dynamics. In contrast, FRET can investigate the dynamics on length scales from 2 nm to 10 nm. Therefore, both methods are complementary to each other.

2.4 Peptides and constructs

2.4.1 FG repeats

In order to systematically map the conformational dynamics of the studied peptides, contact formation rates of different parts of peptides were investigated by PET-FCS. For doing this, a single Trp was introduced at different specific positions into the peptides. Thus, a series of constructs were produced with only one Trp in each of them.

Nsp1 contains 823 amino acids, with the molecular weight about 86 KDa. The N-terminus (FG) and C-terminus (FSFG) sequences of Nsp1 fragment as well as the mapping positions for PET-FCS measurements are shown in figure 2.4.1.

Figure 2.4.1 Sequences of Nsp1 fragments and PET-FCS mapping positions. Amino acid sequences are given in one-letter code. Amino acids marked in yellow represent FG and FSFG units. Amino acids marked in pink represent the positions that are mutated to Trp. At the end of each sequence, one Cysteine (light blue) was introduced for conjugating Atto655 via maleimide reaction chemistry.

3FG and 3FSFG contain both 50 amino acids. 3FG, 2FG and 1FG contain 3, 2 and 1 FG units in the sequences, respectively. 3FSFG, 2FSFG and 1FSFG contain 3, 2 and 1 FSFG units in the sequences, respectively. The spacers between FG units in 3FG are mainly uncharged amino acids, like Thr (T), Asn (N) and Gln (Q). Whereas, the spacers between FSFG units in 3FSFG are mainly charged amino acids, like Asp (D), Glu (E) and Lys (K).

Detailed sequences of all the Nsp1 constructs are shown in table 2.4.1.

Table 2.4.1 Amino acid sequences of Nsp1 constructs studied in this thesis

Name	Sequence
3FG_control	AGTSLFGSSSAQQTKSNGTAGGNTFGSSSLFNNSTNSNTTKLNFGGGNNTC
3FG_50	*W*AGTSLFGSSSAQQTKSNGTAGGNTFGSSSLFNNSTNSNTTKLNFGGGNNTC
3FG_40	AGTSLFGSS*W*AQQTKSNGTAGGNTFGSSSLFNNSTNSNTTKLNFGGGNNTC
3FG_30	AGTSLFGSSSAQQTKSNGT*W*GGNTFGSSSLFNNSTNSNTTKLNFGGGNNTC
3FG_20	AGTSLFGSSSAQQTKSNGTAGGNTFGSSS*W*FNNSTNSNTTKLNFGGGNNTC
3FG_13	AGTSLFGSSSAQQTKSNGTAGGNTFGSSSLFNNSTN*W*NTTKLNFGGGNNTC
3FG-10	AGTSLFGSSSAQQTKSNGTAGGNTFGSSSLFNNSTNSNT*W*KLNFGGGNNTC
2FG-31	*W*AGTSLFGSSSAQQTKSNGTAGGNTFGSSSLFC
2FG-21	AGTSLFGSS*W*AQQTKSNGTAGGNTFGSSSLFC
2FG-11	AGTSLFGSSSAQQTKSNGT*W*GGNTFGSSSLFC
1FG_12	*W*AGTSLFGSSSAQC
3FG_10_cut tail	*W*KLNFGGGNNTC
3FSFG-50	*W*KPAFSFGAKSDEKKDGDASKPAFSFGAKSDEKKDSDSSKPAFSFGTKSNEC
3FSFG-40	KPAFSFGAK*W*DEKKDGDASKPAFSFGAKSDEKKDSDSSKPAFSFGTKSNEC
3FSFG-31	KPAFSFGAKSDEKKDGDA*W*KPAFSFGAKSDEKKDSDSSKPAFSFGTKSNEC
3FSFG-21	KPAFSFGAKSDEKKDGDASKPAFSFGAK*W*DEKKDSDSSKPAFSFGTKSNEC
3FSFG-12	KPAFSFGAKSDEKKDGDASKPAFSFGAKSDEKKDSDS*W*KPAFSFGTKSNEC
2FSFG-31	*W*KPAFSFGAKSDEKKDGDASKPAFSFGAKSDEC

2FSFG-21	KPAFSFGAK***W***DEKKDGDASKPAFSFGAKSDEC
2FSFG-12	KPAFSFGAKSDEKKDGDA***W***KPAFSFGAKSDEC
1FSFG-12	***W***KPAFSFGAKSDEC

Amino acid sequences are given in one-letter code. 3FG and 3FSFG represent the sequences from N-terminus and C-terminus of Nsp1, respectively. The last two digits in the name indicate the distance between Trp (W) and Cysteine conjugated with Atto655.

2.4.2 GS repeats

Repeating units of glycine and serine (GS repeats) are used as convenient peptide linkers in most protein fusion applications which require high flexibility between the fused proteins. GS repeats are well-characterized random coil models [141]-[143]. In this thesis, the conformational dynamics of a series of artificial GS repeats was studied for comparing their dynamics with those of the FG repeats. Detailed sequences are shown in table 2.4.2.

Table 2.4.2 Amino acid sequences of GS repeats studied in this thesis

Name	Sequence
(GS)5W	GS GS GS GS GS W
(GS)7W	GS GS GS GS GS GS GS W
(GS)9W	GS GS GS GS GS GS GS GS GS W
(GS)10W	GS GS GS GS GS GS GS GS GS GS W
(GS)15W	GS GS GS GS GS GS GS GS GS GS GS GS GS GS GS W
(GS)20W	GS GS GS GS GS GS GS GS GS GS GS GS GS GS GS GS GS GS GS GS W
(GS)5W(GS)5	GS GS GS GS GS W GS GS GS GS GS

Amino acid sequences are given in one-letter code. The number in the name indicates the number of "GS" units in the sequence. For example, (GS)5W means 5 "GS" units in the sequence .

2.4.3 Prothymosin alpha and Integrase

To calibrating molecular dynamics simulations of IDPs, hydrodynamic properties of other IDPs, namely Prothymosin alpha (ProTα) and Integrase, were also investigated in this thesis.

ProTα, which lacks an amino-terminal signal peptide, is found virtually exclusively in the nuclei of all mammalian cells [144]. Recent studies indicate that ProTα is involved in many biological functions such as cell proliferation of mammalian cells, protection against apoptosis, and transcriptional regulation [145]-[150]. Structural studies suggest that ProTα adopts a random coil-like conformation with persistent curvature under physiological conditions [151]. Only one structure of a short fragment of ProTα bound to another protein was resolved by NMR [152]. ProTα is an unusual polypeptide due to its high content of acidic amino acids, its low content of hydrophobic amino acids, and the absence of aromatic and sulfur-containing amino acids. Two-thirds of the acidic residues are located in the middle part of the sequence.

Integrase is a key viral enzyme during HIV replication that catalyzes the insertion of viral DNA into the host genome [153]. Integrase is a 32 kDa protein of 288 amino acids composed of three functional domains: the N-terminal Zinc binding domain, the catalytic core domain, and the C-terminal domain. The Zinc-binding domain is disordered in the absence of Zn^{2+} ions, and adopts a secondary structure with a high alpha-helical content after binding of Zn^{2+} ions [154, 155]. In this thesis, I studied the N-terminal Zinc binding domain.

Detailed sequences of ProTα and Integrase are shown in table 2.4.3.

Table 2.4.3 Amino acid sequences of Prothymosin alpha & Integrase

Name	Sequence
Prothymosin alpha	AHHHHHHSAALEVLFQGPSDAAVDTSSEITTKDLKEKKEVVEEAENGRDAPANGNAENEENGEQEAD NEVDEE*C*EEGGEEEEEEEEGDGEEEDGDEDEEAESATGKRAAEDDEDDDVDTKKQKTDED*C*
Prothymosin alpha_cut_histag	GPSDAAVDTSSEITTKDLKEKKEVVEEAENGRDAPANGNAENEENGEQEADNEVDEE*C*EEGGEEEEEEE EGDGEEEDGDEDEEAESATGKRAAEDDEDDDVDTKKQKTDED*C*
Integrase	GSHMFLDGIDKAQEEHEKAHSNWRAMASDFNLPPVVAKEIVAS*C*DK*C*QLKGEAMHGQVD*C*

Amino acid sequences are given in one-letter code. For Prothymosin alpha, one Glutamate (E) was mutated into Cysteine (C) in the sequence and another Cysteine was introduced at the C-terminus. Both Cysteines were supposed to be not influential to the hydrodynamic property. For Integrase, two Cysteines at positions No. 44&47 that can either form a disulfide bond or bind Zn^{2+} ions, leading to the folding of the protein. Cysteine at position 60 position was labelled with Alexa 488. Single labelled peptides were chosen during the purification step.

2.5 Coupling of fluorescent dyes to synthetic peptides for single molecule spectroscopy

Despite the diversity of the 20 different amino acids which form the building blocks of proteins, only a few functional groups comprise suitable targets for chemical coupling of fluorescent dyes to proteins. The sulfhydryl group (-SH) of Cysteine, and the primary amine group (-NH_2) of the N-terminus and of the side group of Lysine (K) are the most common targets for coupling. The chemical reaction schemes for both coupling targets, maleimide and N-hydroxysuccinimide (NHS) ester, are shown in figure 2.5.1.
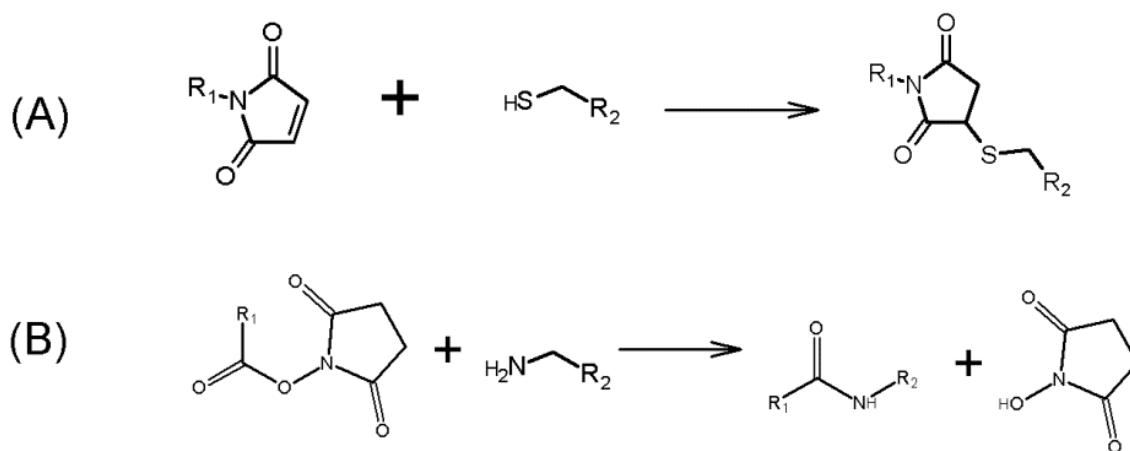
Figure 2.5.1 Reaction schemes for chemical coupling of a fluorophore to proteins or peptides. (A) Reaction scheme for maleimide coupling: a sulfhydryl group of Cysteine specifically reacts with a maleimide group to form a stable thio-ether bond. (B) Reaction scheme for N-hydroxysuccinimide (NHS) ester coupling: a primary amine of *e.g.* the N-terminus specifically reacts with a NHS ester to form a stable amide bond. R_1 represents a fluorophore (Atto655 in this thesis). R_2 represents the peptide or protein.

Sulfhydryl groups, or thiol groups, are found in the side chain of Cysteine residues. Cysteine residues are often forming disulfide bonds in proteins. The disulfide bond can be reversibly reduced into two sulfhydryl groups in the reducing environment. The maleimide-activated compounds react with sulfhydryl groups in neutral conditions (pH 6.5-7.5).

Primary amines are found at the N-terminus of each peptide or protein and in the side chain of lysine residues. These primary amines are positively charged at physiological pH, and therefore tend to be located on the outer surface of native proteins where they are readily accessible to reactive coupling reagents. The NHS ester-activated coupling compounds react with primary amines from physiological to slightly alkaline conditions (pH 7.2-9).

For labeling FG repeats, the fluorescent dye was coupled to the peptide via maleimide reaction chemistry. Therefore, a single Cysteine was introduced into the C-terminus of each peptide. First, approximately 500 µg of the peptide was dissolved in 5-10 µl peptide dissolving buffer, then 50 µl-90 µl maleimide coupling buffer was added. 50 µg lyophilized Atto655 maleimide

(Atto655md) powder was dissolved in the same reaction mixture. Coupling reactions were performed in the dark at room temperature for 2 hours.

For GS repeats, Atto655 NHS ester was used for coupling. First, approximately 500 µg peptide was dissolved in 5-10 µl peptide dissolving buffer, then 50-90 µl NHS ester coupling buffer was added. 50 µg lyophilized Atto655 NHS ester powder was dissolved in the same reaction mixture. Coupling reactions were performed in the dark at room temperature for 2 hours.

To separate uncoupled dyes from the dye-coupled peptides, the reaction mixture was purified by size-exclusion chromatography with a self-packed Sephadex G30 column and 4 x PBS as the running buffer. Peptides were freshly prepared before each measurement to minimize possible artifacts caused by oxidation or photo-oxygenation of tryptophan residues. Peptides shorter than 16 amino acids, such as $(GS)_5W$&$(GS)_7W$, were purified with a Superdex Peptide column.

To remove the His tag which was needed for affinity purification after bacterial expression, Prothymosin alpha coupled with Atto Oxa 11 was treated with HRV3C Protease. The cleavage reaction was performed overnight at 4°C in 100 µl His tag cleavage buffer with 50 µM protein concentration and 2 µl (about 2 units) HRV3C Protease. Removal of the HRV3C Protease and the cleaved His tag from Prothymosin alpha was done by size-exclusion chromatography with a self-packed Sephadex G30 column and 4 x PBS as running buffer.

2.6 Fluorescence Correlation Spectroscopy measurements of IDPs

2.6.1 2fFCS and PET-FCS setups

Dual-focus FCS measurements were performed using the 2fFCS setup which had been described in section 2.2.2. For fluorescence excitation, two pulsed diode lasers with linear polarization were used, with 647 nm wavelength, pulse duration of 50 picoseconds (FWHM), and 40 MHz repetition rate. The laser power was adjusted to 7 µW each.

PET-FCS measurements were also performed on this FCS setup. For fluorescence excitation in PET-FCS, diode lasers in continuous wave mode were used, with 647 nm wavelength. The laser power was adjusted to 100 µW each. For these measurements, the Nomarski prism was removed for maximizing excitation intensity in one single focus.

2.6.2 Sample preparation for single molecule spectroscopy measurements

In order to characterize the conformational dynamics of intrinsically disordered proteins, association and dissociation rates of intra-chain contact formation of different peptides were studied by PET-FCS. The hydrodynamic radii of intrinsically disordered proteins were measured with 2fFCS.

All measurements were performed in 8-well tissue culture chamber with a glass cover slip bottom. For PET-FCS, each sample was measured 5 times with a measurement time of 5 min. For 2fFCS, each sample was measured 3 times with a measurement time of 15 min. The diffusion coefficient of Atto655md was taken as reference for all 2fFCS measurements. All measurements were done at room temperature.

For FG repeats and GS repeats, both PET-FCS and 2fFCS measurements were done in 150 mM NaCl, containing 0.3 mg/ml of bovine serum albumin (BSA) and 0.05 % (v/v) Tween-20 to suppress surface interactions.

Integrase was measured in 150 mM NaCl with 0.005% (v/v) Tween-20, 0.1 mM EDTA, and 1 mM DTT. Prothymosin alpha was measured in 150 mM NaCl, 0.005 % (v/v) Tween-20, and 1 mM DTT.

2.6.3 Aggregation study of FG repeats

In order to study potential aggregation property of FG repeats, hydrodynamic radii and the conformational dynamics at a series of concentrations were investigated to monitor whether aggregation, such as dimer or trimer formation can be observed.

First, 200 µM unlabeled FG peptide was dissolved in 150 mM NaCl, containing 0.005 % (v/v) Tween-20 and 37 % acetonitrile, as the stock solution. Then, a serial dilution method (dilution factor 1:5 or 1:10) was used to prepare diluted solutions of unlabeled FG peptides.

According to the average number N in the Equation (3) of FCS, the concentration of each labelled peptide can be estimated and adjusted to about 10 molecules, on average, in the detection volume. Then, 50 µl concentration-adjusted labelled peptides and 50 µl unlabeled peptides were mixed in the dark at room temperature. After 1-2 hours, the mixture was transferred into 8-well tissue culture chamber for measurements. A series of concentrations of unlabeled peptides mixed with labelled peptides were measured by both 2fFCS and PET-FCS as previously described (see section 2.6.2).

2.7 Molecular dynamic simulations

All MD simulations of 3FG constructs were carried out by Dr. Sarah Rauscher from the Department of Theoretical and Computational Biophysics of the Max Planck Institute for Biophysical Chemistry in Göttingen.

The simulation placed a peptide into a rhombic dodecahedral box with water molecules and 0.15 M NaCl, with a total of ~42000 atoms. GROMACS version 4.5.4 [156] was used for all simulations. Prior to the simulation runs, energy minimization with a steepest descent algorithm was performed. The bond lengths with hydrogen atoms were constrained using the LINCS algorithm [157]. An integration time step of 2 fs was used. A cutoff of 0.95 nm was used

for the Lennard-Jones interactions and short-range electrostatic interactions. Long-range electrostatic interactions were calculated by particle-mesh Ewald summation with a grid spacing of 0.12 nm and a fourth order interpolation [158]. The velocity rescaling thermostat was used for all simulations [159]. Equilibration was performed at 298 K for 1 ns using Berendsen pressure coupling [160], followed by 5 ns of simulation in the NPT ensemble using the Parrinello-Rahman algorithm [161]. The configuration from this simulation with a volume closest to the average volume was then used for subsequent replica exchange (RE) [162] simulations in the canonical ensemble. Simulations with the TIP4P-D water model [81] were carried out without replica exchange, consistent with Piana *et al.* [81] as conformational sampling in this force field is fast, and RE was not needed to obtain sufficient sampling, comparable to the other force fields. Visual molecular dynamics (VMD) was used for all molecular visualizations [163].

The GROMACS utility *g_gyrate* was used to calculate the radius of gyration R_g. Standard error of the mean was computed using a blocking procedure [164]. For each simulation, an equilibration period was delineated on the basis of R_g; this initial collapse of the peptide to more compact conformations was excluded from analysis.

3.   Results & Discussion

3.1   Hydrodynamic Radii of FG and GS repeats

I performed 2fFCS measurements of the diffusion coefficients of different peptides for determining their hydrodynamic radii. The goal of these measurements was twofold: Firstly, I wanted to check how the fluorescent labeling influences the dynamic properties of the labeled peptides. Secondly, by comparing the obtained hydrodynamic radii with MD simulations using different force fields, we (Sarah Rauscher and myself) wanted to find out which force field is best for IDP simulations. Values of hydrodynamic radii R_H were calculated from measured diffusion coefficients according to equation (3). The diffusion coefficient of Atto655md under the same experimental condition was used as reference. The hydrodynamic radius of Atto655md is approximately 6 Å [125].

All protein hydrodynamic data are summarized in table 3.1.1. The results show that the hydrodynamic radius R_H of different constructs with Trp at different positions from the same sequence is constant. The variation between different constructs of 3FG and 3FSFG is less than 1 Å. This result suggests that the fluorescence modification and also the introduction of Trp at different positions do not affect the overall structure or diffusion behavior.

Previous theoretical and experimental studies indicate that random coils prefer to be highly compact [165]-[168]. Thus, more flexible random coils tend to be more compact and therefore have smaller hydrodynamic radii. The hydrodynamic radii R_H of FG repeats are larger than those of GS repeats with the same number of amino acids. For example, the hydrodynamics radius of 2FG is found to be 13 Å, whereas that of (GS)15W is 9.7 Å. This suggests that FG repeats are less compact than GS repeats. This result is in agreement with previous studies which suggest that GS repeats are random coils with very high flexibility [141, 142].

When comparing 3FG with 3FSFG, it is found that the hydrodynamic radius of 3FG is smaller than that of 3FSFG, which indicates that 3FG is more compact than 3FSFG. In contrast, the hydrodynamics radius of 2FG is slightly larger than that of 2FSFG, which indicates that 2FG is

less compact than 2FSFG. But how can a short FG sequence be less compact than a longer FG sequence? As shown in the Materials & Methods section, 2FG has the same sequence as the 31 amino acids of the N-terminus of 3FG ($3FG_{1-31}$), and 2FSFG has the same sequence as the 31 amino acids of the N-terminus of 3FSFG ($3FSFG_{1-31}$). Therefore, one explanation for this result could be that the fragment of $3FG_{32-50}$ is more flexible than $3FSFG_{32-50}$, which makes the full sequence of 3FG more compact than 3FSFG. Another possible explanation is that the hydrophobic interaction between $3FG_{1-31}$ and $3FG_{32-50}$ is stronger than that between $3FSFG_{1-31}$ and $3FSFG_{32-50}$, which makes 3FG more compact, because 3FG contains more uncharged amino acids such as Thr(T), Asn (N) and Gln (Q), while 3FSFG contains more charged amino acids such as Asp (D), Glu (E), or Lys (K).

Table 3.1.1: Hydrodynamic radii (R_H) of FG, FSFG and GS repeats

Peptides	R_H (Å) ±SEM	Peptides	R_H (Å) ±SEM
3FG_10	14.8±0.07	3FSFG_12	15.9±0.08
3FG_20	14.5±0.11	3FSFG_21	16.7±0.12
3FG_30	14.6±0.05	3FSFG_31	16.7±0.06
3FG_40	15.2±0.05	3FSFG_40	16.4±0.03
3FG_50	15.2±0.03	3FSFG_50	16.4±0.17
2FG_11	13.6±0.1	2FSFG_12	13.0±0.06
2FG_21	13.2±0.22	2FSFG_21	13.0±0.04
2FG_31	13.7±0.02	2FSFG_31	12.5±0.08
1FG_12	10.3±0	1FSFG_12	9.5±0.02
(GS)5W	8±0.06	(GS)10W	9±0.05
(GS)7W	7.9±0.07	(GS)15W	9.7±0.02
(GS)9W	9.2±0.09	(GS)20W	10.4±0.07

In a previous study by Anastasia Loman [125], it has been shown that there is a close relation between the diffusion coefficient and the molecular size of random peptide coils. A simple

approximation for the diffusion coefficient of random coils can be approximately deduced from its molecular weight, as also shown in figure 3.1.1. by

$$D = 11 \times 10^{-5} \times MW^{-0.51} \qquad (21)$$

where D is the diffusion coefficient, and MW is the molecular weight of a peptide.
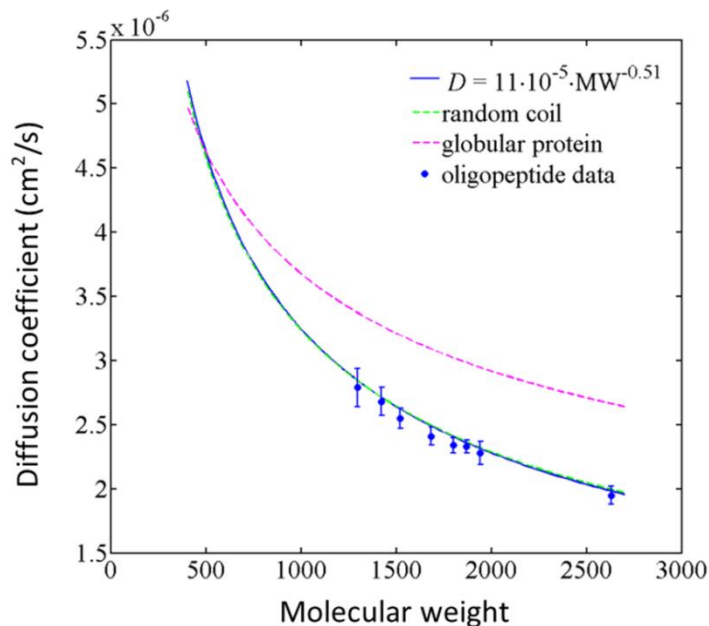
Figure 3.1.1 Relationship between molecular size and diffusion coefficient of different types of peptides (taken from ref. [125]).

Based on this prediction, the hydrodynamic radii of all peptides which were studied in this thesis were calculated and are summarized in table 3.1.2.

Table 3.1.2 Hydrodynamic data of FG and GS repeats

Peptides	MW (KDa)	Diffusion coefficient ($\times 10^{-6}$ cm^2/s)	Calculated R_H (Å)
3FG*	5.0	1.43	16.8
2FG*	3.2	1.79	13.4
1FG	1.4	2.73	8.8
3FSFG*	5.5	1.36	17.6
2FSFG*	3.5	1.71	14
1FSFG	1.6	2.55	9.4
(GS)5W	0.9	3.43	7
(GS)7W	1.2	2.96	8.1
(GS)9W	1.5	2.64	9.1
(GS)10W	1.6	2.55	9.4
(GS)15W	2.4	2.08	11.6
(GS)20W	3.0	1.85	12.9

MW, molecular weight; * molecular weights are given as the average of all different constructs with Trp at different positions. Atto655md is used as reference. Diffusion coefficient of Atto655md in water at 25 °C is determined as 4×10^{-6} cm^2/s . R_H of Atto655md is 6 Å.

In summary, R_H values measured by 2fFCS are in good agreement with values calculated from the relationship between molecular size and diffusion coefficient obtained by Anastasia Loman. This suggests that the proteins in this thesis are intrinsically disordered proteins.

3.2 Molecular dynamic simulation to study peptides

In order to evaluate the accuracy of various force fields used for IDPs simulation, six all-atom empirical force fields of *de novo* molecular simulations were investigated. All MD data are provided by Dr. Sarah Rauscher. Detailed information of the used force fields is shown in table 3.2.1.

Amber ff99sb*-ildn and CHARMM 22* are two modern force fields, which were supposed to generate IDP and unfolded state ensembles that are on average too compact [89]. Amber force fields were also suggested to systematically underestimate chain dimensions of IDPs and unfolded states [92]. Another force field is OPLS-AA/L, which performed poorly in two other force field benchmark studies [71, 73]. Different water models were also studied.

Table 3.2.1: Force fields included in the comparison

Force Field (Abbreviation)	Peptide Force Field	Water Model
Amber ff99sb*-ildn (a99sb)	Amber ff99sb*-ildn [169]	TIP3P [170]
Amber ff03w (a03w)	Amber ff03w [171]	TIP4P-2005 [172]
CHARMM 22* (c22*)	CHARMM 22* [169]	Charm-modified TIP3P [173]
CHARMM 22* (c22*/D)	CHARMM 22* [169]	TIP4P-D [81]
CHARMM 36 (c36)	CHARMM 36 [174]	TIP3P [170]
OPLS-AA/L (opls)	OPLS-AA/L [175]	TIP3P [170]

Simulations of 3FG peptides were carried out using six different force fields. Significant differences of the obtained structural ensembles between force fields can be observed. The data is shown in figure 3.2.1. The Amber ff99sb*-ildn and OPLS-AA/L results are very compact, resembling collapsed globule-like ensembles. The most expanded ensemble is obtained with ChARMM 22*/D. The other three force fields generated a broad spectrum of chain dimensions that fall between these two extreme cases. The ensembles of 3FG repeats obtained using the ChARMM 22* force field with TIP4P-D water model exhibit a much broader distribution of chain

dimensions than those obtained with the other force fields. These differences are consistent with those reported by Mercadante *et al.* [93]. The R_H of 3FG calculated from the diffusion coefficient is 16 ±1 Å (see table 3.1.1). Comparison with experimental data of 3FG repeats indicates that the CHARMM 22* ensemble yields the best agreement between MD simulations and experimental data.

MD simulations of Prothymosin alpha and Integrase obtained by ChARMM 22* were also in best agreement with experimental data. The hydrodynamic radii of Prothymosin alpha and Integrase measured by 2fFCS in this thesis are presented in the appendix. The simulation data of these two proteins is not shown.

In summary, the goal of this section was the study of the accuracy of conformation ensembles obtained from *de novo* simulations. The results demonstrate an unexpectedly high sensitivity of IDP conformational ensembles to differences between the force fields. For example, major differences in chain dimensions were found: ensembles span the entire range from collapsed globule-like to highly expanded chain. A key finding of this study is that the conformation ensembles obtained using CHARMM 22* with charm-modified TIP3P water agrees best with the available experimental data. This is in agreement with the recent studies of RS peptide [83].
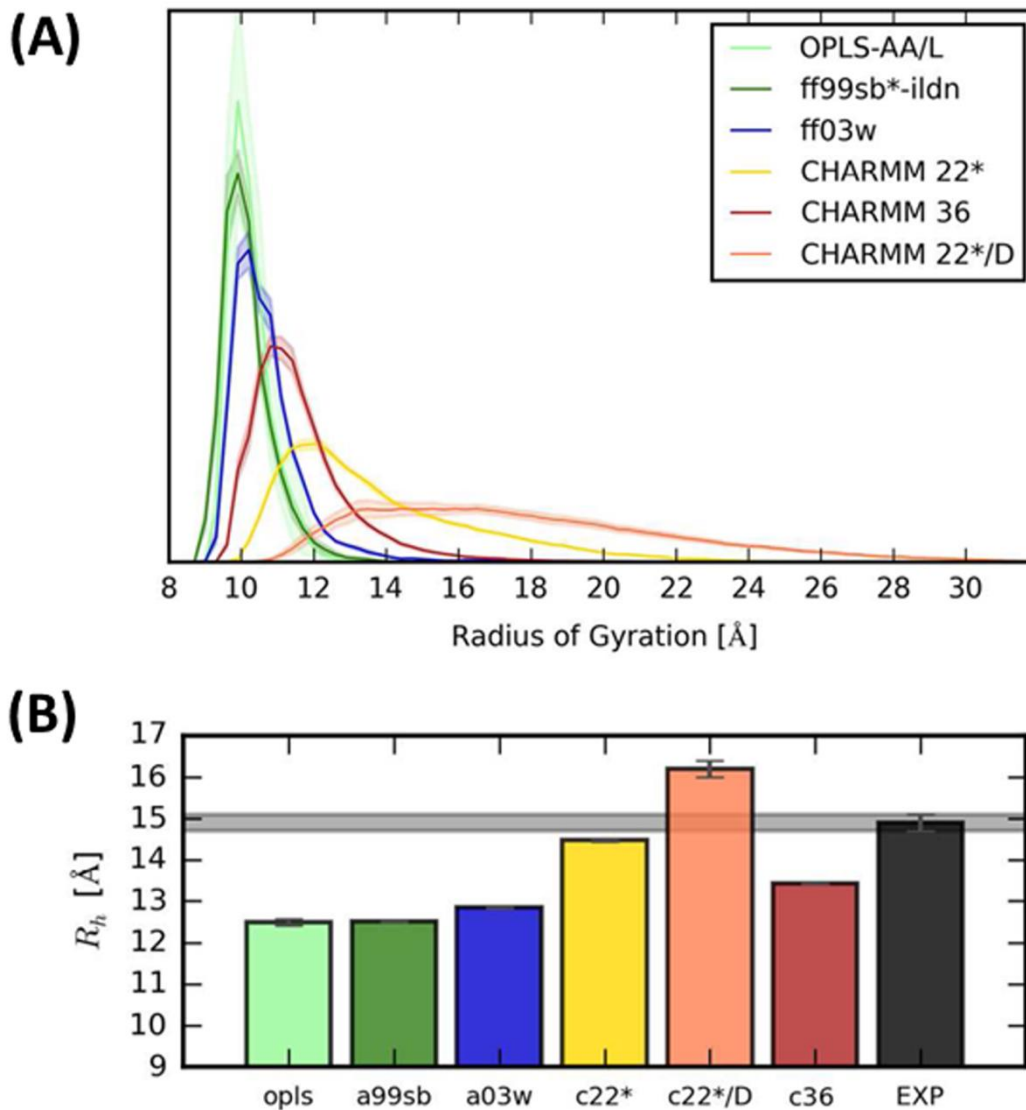
Figure 3.2.1 Comparison to chain dimensions measured by FCS for 3FG repeats (figure from Dr. Sarah Rauscher) (A) Histograms of the radius of gyration, R_g, for structural ensembles obtained with different force fields. (B) The hydrodynamic radius is shown for each force field and compared to the experimentally measured R_H, which is the mean value of the five R_H values of 3FG constructs. In each case, the error bars indicate the standard error of the mean (in some cases smaller than the line width). For better comparison with simulation data, the mean R_H value of 3FG constructs together with its standard error is shown by the gray bar.

3.3 Conformational dynamics of FG repeats

3.3.1 Optimizing conditions for single molecule spectroscopy measurements

During single molecule spectroscopy measurements which were performed in this study, a remarkable decrease of the photon count rate was observed. Figure 3.3.1 (A) shows the time course of the count rate, which is proportional to the molecular concentration, and which decreases almost 50% within a 5 min measurement. This decrease is due to the non-specific absorption of peptides and proteins on the glass surface. This non-specific absorption causes a considerable and continuous reduction of molecules within the detection volume. This reduction can affect the quality of the PET measurement, as shown in figure 3.3.1 (B). The data points of the auto-correlation curve show a large scatter at lag times below 10^{-5} s.

(A)   (B)
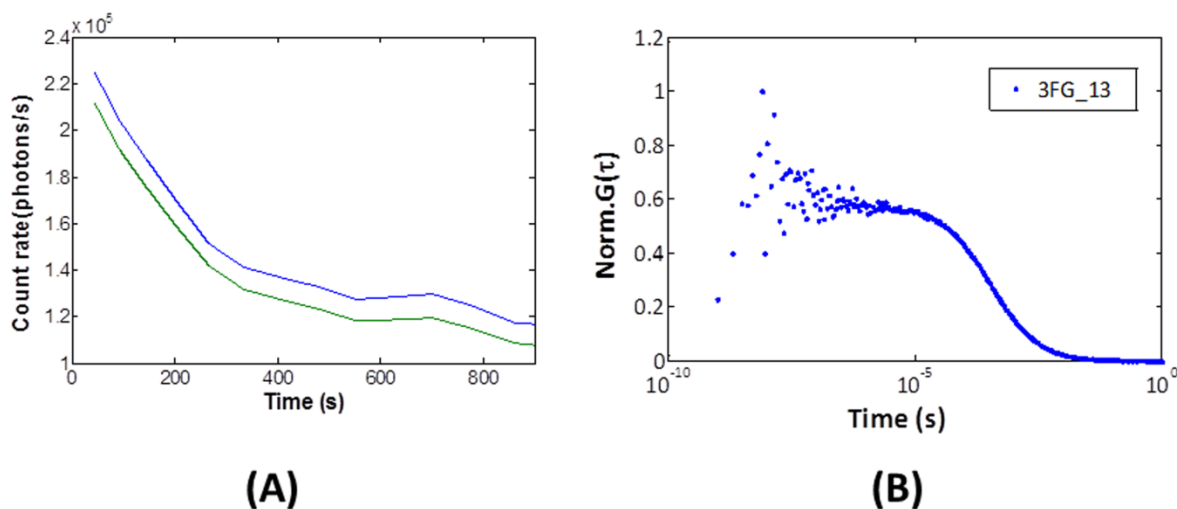
Figure 3.3.1 (A) Time course of the count rates [photons/s] of the two detectors during a single molecule spectroscopy measurement. The count rates decrease rapidly. (B) Normalized auto-correlation function of 3FG_13.

In order to reduce the glass-surface interaction, one can add specific reagents to the buffer. After consulting the literature [176]-[178], bovine serum albumin (BSA) and polyoxyethylene (20) sorbitan monolaurate (Tween-20, a detergent) occurred to be most suitable additives. BSA is a large globular protein with a molecular weight of 66 kDa, and is commonly used in many

biological assays to prevent adhesion of enzymes to the reaction tube. Micelles can be formed when the concentration of Tween-20 is above the so-called critical micelle concentration (CMC) [179]. Both BSA and micelles can affect the conformational dynamics of the peptide.

To check the impact of BSA and Tween addition on the properties of a labeled IDP, the peptide 3FG_13 was measured at different concentrations of BSA and Tween-20 to investigate the potential influence of these additions on the conformational dynamics of the peptide. The data is shown in figure 3.3.2. The practically constant correlation curves for different concentrations imply that the effect of BSA and Tween-20 on the conformational dynamics of the peptide is negligibly small. The association and dissociation rate constants from curve fitting do also not show any significant changes (data not shown).

The data of the auto-correlation curves show also a large scatter when experiments are performed at very low concentrations of Tween-20 or BSA (see blue curves in figure 3.3.2). This is due to the absorption of protein to the glass walls in the absence of Tween-20 and/or BSA. Considering the different hydrophobic properties of peptides which result in different absorption affinities to the glass surface, we found that the optimal concentration of BSA and Tween-20 for efficient suppression of unspecific absorption were 0.3 mg/ml BSA and 0.05 % Tween-20, without affecting the conformational dynamics of the studied peptides.
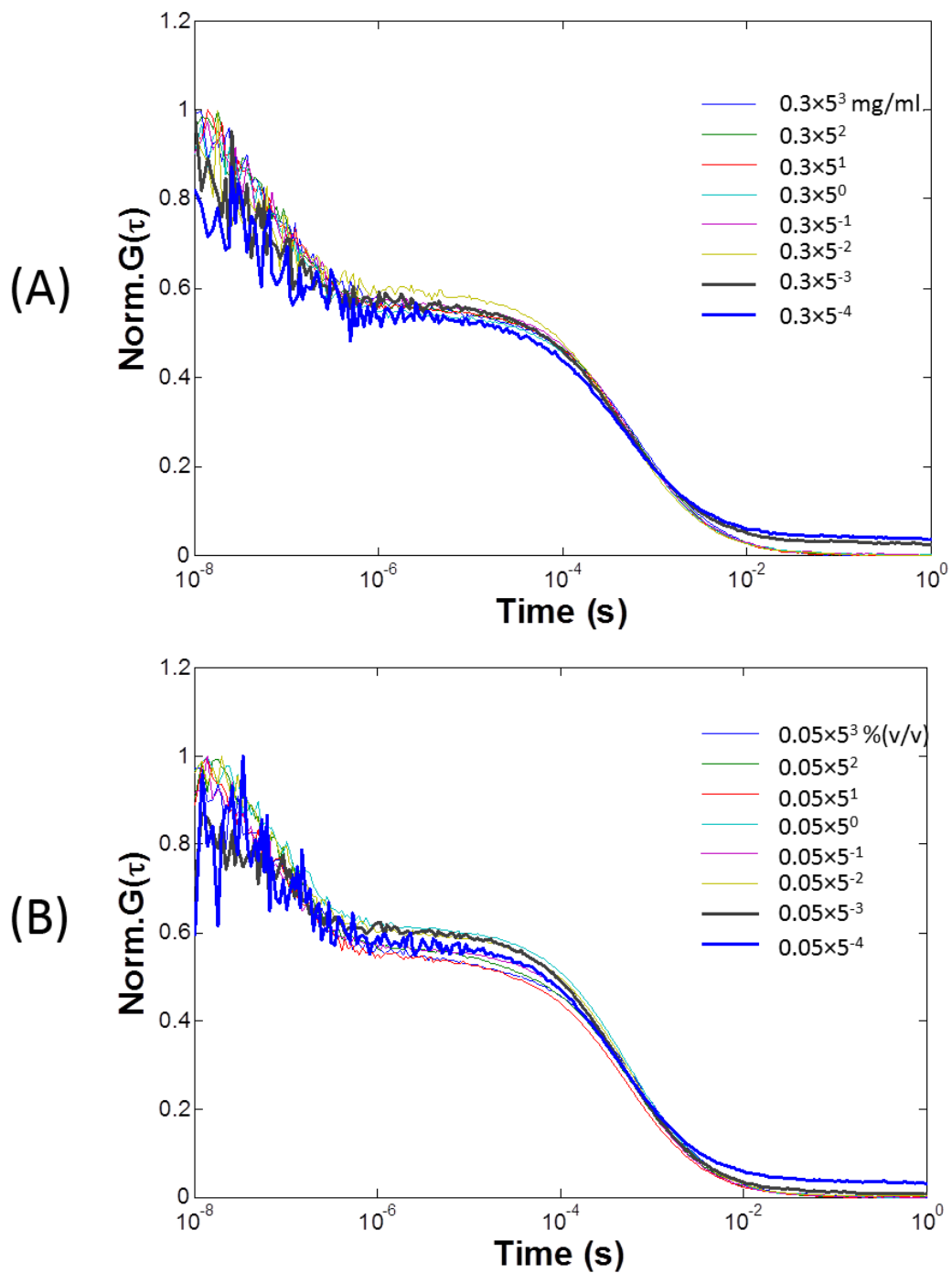
Figure 3.3.2 Influence of BSA and Tween-20 to the conformational dynamics of 3FG_13. Auto-correlation functions of 3FG_13 measured at various BSA (A) and Tween-20 (B) concentrations. The concentrations were prepared with a dilution factor of 1:5.

3.3.2 Conformational dynamics of the N- and C-terminus of Nsp1

The N- and C-terminus of Nsp1 show different features in gel formation experiments as mentioned before. Therefore, it is interesting to investigate the difference in conformational dynamics between the N- and C-terminus. In this thesis, N- and C-terminus sequences of Nsp1 are abbreviated as FG and FSFG, respectively.

Two control experiments were performed for PET-FCS. One was measuring the free dye (Atto655md) alone. The other was measuring 3FG without any Trp (3FG_control). The brightness of the dye can be estimated by measuring the count rate per molecule during the measurement. As it occurred, the brightness of the dye (Atto655) was the same before and after coupling to peptides (data not shown), which indicates that this coupling does not influence the photophysical properties of Atto655. The normalized correlation curves of Atto655md and the 3FG_control are shown in figure 3.3.3. The curve of the 3FG_control shows no decay between ten nanoseconds and a few dozen microseconds, which indicates that the non-Trp amino acids do not quench Atto655.

Normalized auto-correlation curves for different 3FG and 3FSFG constructs are shown in figure 3.3.3. There is a systematic decrease of the PET decay amplitude from shorter to longer distances between Atto655 and Trp. Because the PET decay amplitude is proportional to the efficiency of quenching, this suggests that the longer the distance between Atto655 and Trp, the lower is the chance for Atto655 and Trp to come in contact with each other.
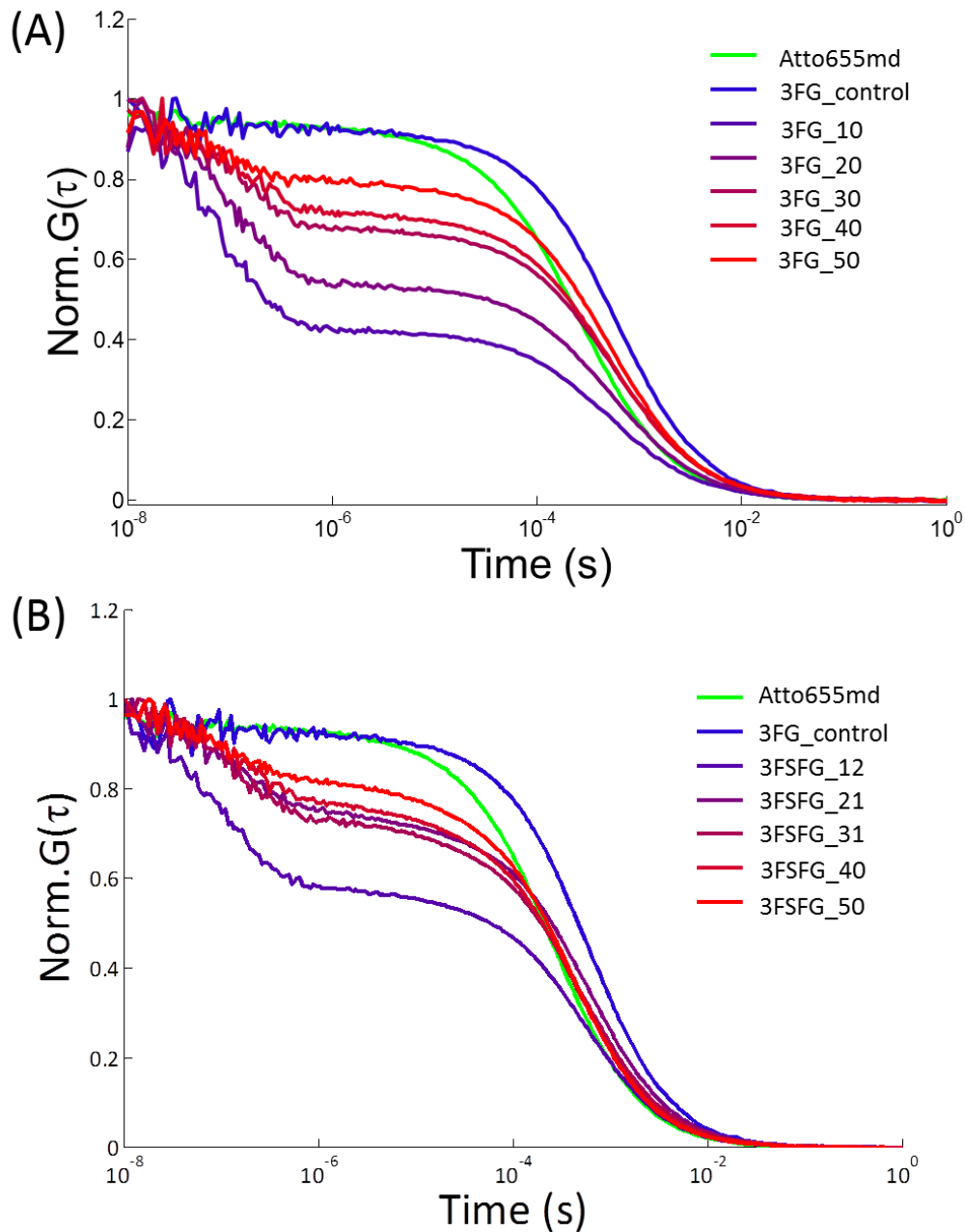
Figure 3.3.3 Normalized auto-correlation functions of different (A) 3FG repeats and (B) 3FSFG repeats.
For comparison, auto-correlation data recorded from Atto655md is shown in green and 3FG_control is shown in blue. Auto-correlation data recorded from different constructs are shown in a red-to-blue color scale (from blue to red, the distance between Atto655 and Trp becomes larger).

The obtained correlation curves could be fitted well with the PET-FCS model as described in the theory section (see section 2.5). The association rates for the FG and FSFG constructs are plotted against the distance between Atto665 and Trp (see figure 3.3.4.)
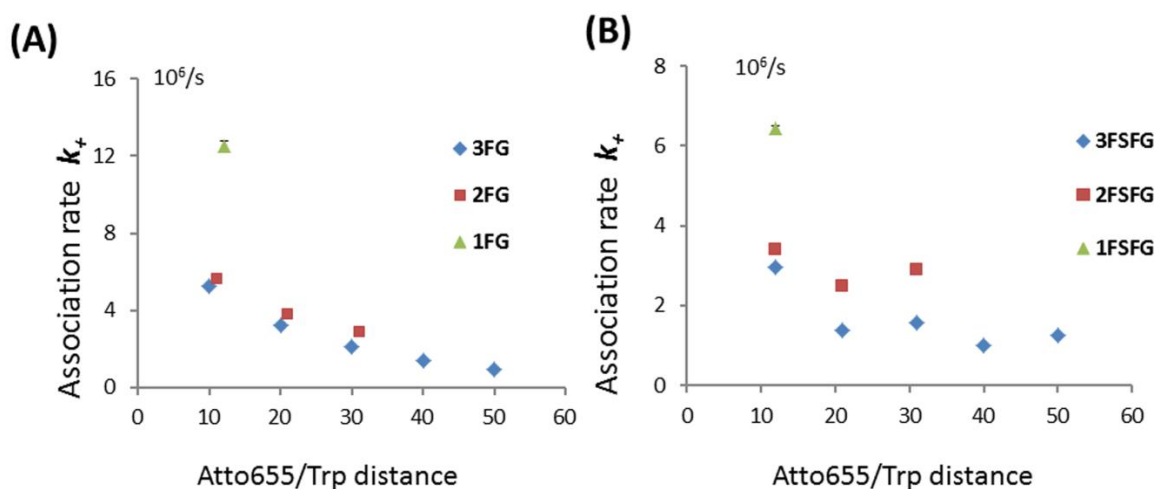
Figure 3.3.4 Dependence of the association rates of different FG (A) and FSFG (B) constructs. Green triangles represent 1FG and 1FSFG, respectively. Red squares represent 2FG and 2FSFG, respectively. Blue squares represent 3FG and 3FSFG, respectively. The length of standard error bar is comparable to or smaller than the size of the corresponding markers.

The following points can be observed in figure 3.3.4.

(1) For 3FG, the association rate constant k_+ decreases from 5.29×10^6/s of 3FG_10 to 0.92×10^6/s of 3FG_50 while the distance between Atto655 and Trp changes from 10 aa to 50 aa. The association rate constant in 3FG (blue squares in figure 3.3.4 A) steadily decreases from short to longer distances between Atto655 and Trp. This is in agreement with previous studies and with polymer chain theory which claims that the association rate k_+ should follow a power law as a function of chain length [143, 180].

(2) For 3FSFG, the association rate k_+ decreases from 2.97×10^6/s for 3FSFG_12 to 1.25×10^6/s for 3FSFG_50, with the distance between Atto655 and Trp increasing from 10 aa to 50 aa. The distance dependence of the association rate of 3FSFG (blue squares in figure 3.3.4 B) does,

however, not decrease. The k_+ for 3FSFG_31 is even bigger than that for 3FSFG_21, with 1.56 ×10^6/s compared to 1.38×10^6/s. This implies that there might be some residual secondary structure within the 3FSFG sequence. A previous study using NMR found abundant β-sheet structures in Nsp1 formed hydrogel [114]. From these results I conclude that, although both FG and FSFG sequences are dominant in Nsp1, only FSFG will contribute to the formation of β-sheet structures while FG is intrinsically disordered.

(3) If one compares the k_+ values for the same sequence, the association rate k_+ varies dramatically from short to long constructs, keeping the distances between Atto655 and Trp the same. For example: k_+ of 1FG_12 is two times larger than 3FG_10 and k_+ of 1FSFG_12 is also two times larger than 3FSFG_12 (see table 3.3.1). In principle, the k_+ rates should not vary significantly when the distance between Atto655 and Trp is kept constant. The only difference between all these constructs is the dangling amino acid "tail" on the other side of the Trp, and this "tail" might affect the conformational dynamics of peptide.

Table 3.3.1: Conformational dynamics data of FG and FSFG constructs

Peptides	K_+ (×10^6/s) ±SEM	Peptides	K_+ (×10^6/s) ±SEM
3FG_10	5.29±0.07	3FSFG_12	2.97±0.04
1FG_12	12.55±0.24	1FSFG_12	6.42±0.08

(4) If one compares the k_+ values of the different constructs of 3FSFG and 3FG which have similar distance values between Atto655 and Trp, one observes that the k_+ values for 3FSFG are systematically smaller. For example: the k_+ value for 3FSFG_12 is smaller than that for 3FG_10, and that for 3FSFG_21 is smaller than that for 3FG_20, and so on (see table 3.3.2). The observed smaller k_+ values for 3FSFG constructs imply that 3FSFG constructs are stiffer than 3FG constructs.

Table 3.3.2: Conformational dynamics data of 3FG and 3FSFG

Peptides	K_+ ($\times 10^6$/s) ±SEM	Peptides	K_+ ($\times 10^6$/s) ±SEM
3FG_10	5.29±0.07	3FSFG_12	2.97±0.04
3FG_20	3.24±0.05	3FSFG_21	1.38±0.02
3FG_30	2.13±0.07	3FSFG_31	1.57±0.02
3FG_40	1.43±0.06	3FSFG_40	1.00±0.02
3FG_50	0.92±0.01	3FSFG_50	1.25±0.02

Previous studies suggest that each FG repeat contains a small structured domain that serves as the anchor site for the rest of the NPC, and a larger unstructured region that are largely exposed on the inner spoke surface of the NPC [96] (see figure 1.5.1). As mentioned in the introduction, Nsp1 is abundant in the NPC and interacts with many other Nups to form two complexes: the Nic96 and the Nup82 complex. Some studies suggest that the C-terminus of Nsp1 contains four coiled regions [181]. These coiled regions are more structured than disordered and have important functions. Mutations in coil-1 lead to a strong nuclear mRNA export defect. Coil-2 region is proposed to be the platform to accommodate both the Nic96 and Nup82 complex, which means that this region contributes to the interaction with other Nups. Mutations in coil-3 and coil-4 lead to defects in nuclear protein import [181].

The 3FG sequence contains residues 99-153, and the 3FSFG sequence contains residues 452-501 from Nsp1. The relative positions of FG, FSFG, and the four coils in the Nsp1 sequence are shown in figure 3.3.5.
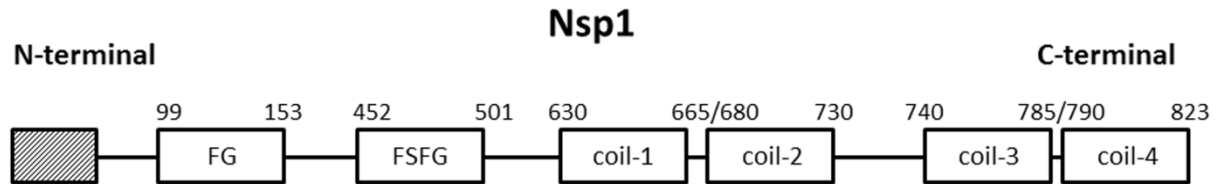
Figure 3.3.5 Schematic diagram displaying positions of Nsp1 fragments. Amino acid numbers are indicated. FG is located at the N-terminus (aa 99-153), whereas FSFG is located at the center of Nsp1 (aa 452-501). Coil-1, Coil-2, Coil-3 and Coil-4 locate at C-terminus of Nsp1.

3FSFG is closer to the four coiled regions than 3FG. Since coil-2 region interacts with other Nups, FSFG might be involved in that interaction. However, the N-terminus, where 3FG is located, is far away from the four coiled regions. Therefore, the N-terminus might remain flexible and reach into the channel.

To summarize, the N-terminus of Nsp1 (3FG) is more flexible than the C-terminus (3FSFG). N-terminus might remain disordered and be exposed on the inner spoke surface of the NPC.

3.3.3 "Tail" influence to the conformational dynamics

The dissociation rates of 3FG and 3FSFG are also plotted as a function of the distance between Atto655 and Trp. The data is shown in figure 3.3.6.
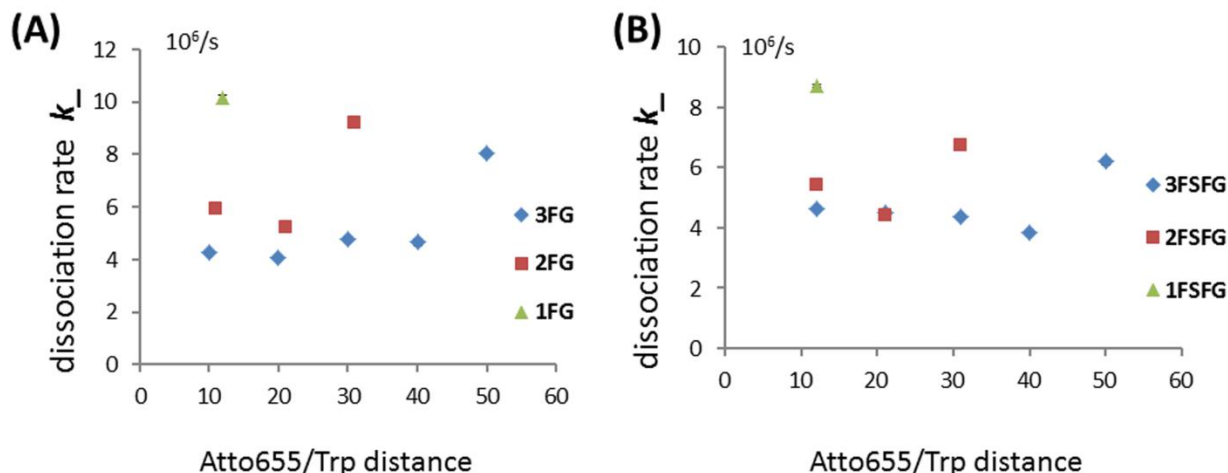
Figure 3.3.6 Dependence of the dissociation rates of different FG (A) and FSFG (B) constructs. Green triangles represent 1FG and 1FSFG, respectively. Red squares represent 2FG and 2FSFG, respectively. Blue squares represent 3FG and 3FSFG, respectively. The length of standard error bar is comparable to or smaller than the size of the corresponding markers.

Previous studies [143, 140] have shown that the dissociation rate k_- depends on the hydrophobic stacking interaction energies between the aromatic moieties of Atto655 and Trp, rather than on the peptide chain mobility. Therefore, k_- should remain constant for different distances between dye and quencher. Surprisingly, k_- rates are not constant; one can even see that they vary significantly for the different constructs used in our study. Moreover, as shown above, the association rate k_+ of short and long constructs but same distance between Atto655 and Trp also varies substantially. The difference between all these constructs is the dangling amino acids "tail" on the other side of the Trp. One assumption is that these varying k_+ and k_- values can be attributed to the hydrodynamic influence of the "tail" on the peptide's conformational dynamics.

To test whether the "tail" influences the association and dissociation rates, the exact same construct sequence but without any "tail" was compared with the one with "tail". The sequences and labelling positions are shown in figure 3.3.7.

3FG_10 : AGTSL**FG**SSSAQQTKSNGTAGGNT**FG**SSSLFNNSTNSNT**W**KLN**FG**GGNNT**C**

1FG_12: **W**AGTSLFGSSSAQ**C**

3FG_10_cut tail : **W**KLN**FG**GGNNT**C**

■: Trp(W) replaced position

F G: FG repeats

C: Cys (labelled with Atto655md)
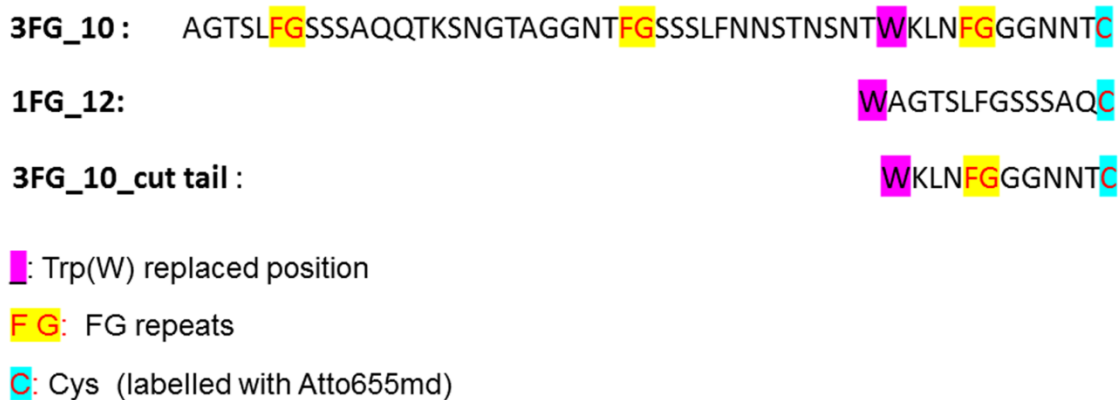
Figure 3.3.7 amino acid sequences of three constructs and labelling positons. 3FG_10_cut tail contains only the last 10 amino acids of the construct 3FG_10. The difference of amino acid sequences between 3FG_10 and 1FG_12 is shown.

PET-FCS measurements were performed for the two constructs, 3FG_10 and 3FG_10_cut tail. The normalized auto-correlation curves are shown in figure 3.3.8.
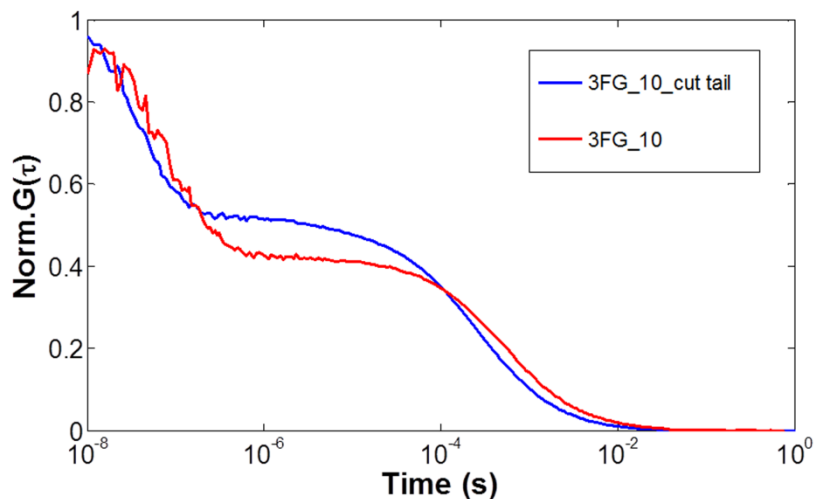
Figure 3.3.8 Normalized auto-correlation functions of constructs with (red) and without (blue) "tail".
The PET relaxation time τ_R of 3FG_10_cut tail (46 ns) is clearly shorter than3FG_10 (100 ns).

The obtained values for the k_+ and k_- rates of 3FG_10 and 3FG_10_cut_tail are presented in figure 3.3.9. The k_+ rate for 3FG_10_cut_tail ($1.12×10^7$/s) is almost twice as large as that for 3FG_10 ($5.28×10^6$/s). Also, the k_- rate for 3FG_10_cut_tail ($1.05×10^7$/s) is about 2.5 times larger than that for 3FG_10 ($4.25×10^6$/s). However, the k_+ and k_- rates of 3FG_10_cut_tail ($1.12×10^7$/s and $1.05×10^7$/s) are practically the same as those for 1FG_12 ($1.25×10^7$/s and $1.01×10^7$/s). The smaller values of k_+ and k_- rates seen for the constructs with "tail" clearly confirm that the "tail" indeed decreases the association and dissociation rates.
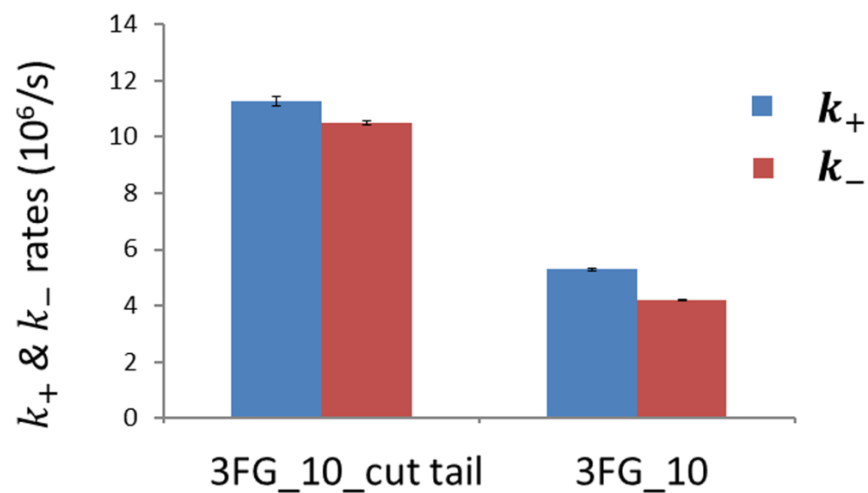
Figure 3.3.9 Association rate k_+ and dissociation rate k_- of two FG constructs with and without "tail".

Additionally, the "tail" effect was also tested on GS repeats. The construct (GS)5W(GS)5 has a tail (GS)5 at the other side of Trp (W), in contrast to the construct (GS)5W. The k_+ and k_- rates for the construct (GS)5W(GS)5 with "tail" are smaller than those of the construct (GS)5W without tail, with k_+-values from $2.37×10^7$/s to $3.64 ×10^7$/s and k_--values from $0.92 ×10^7$/s to $1.28×10^7$/s. The comparison is shown in figure 3.3.10.
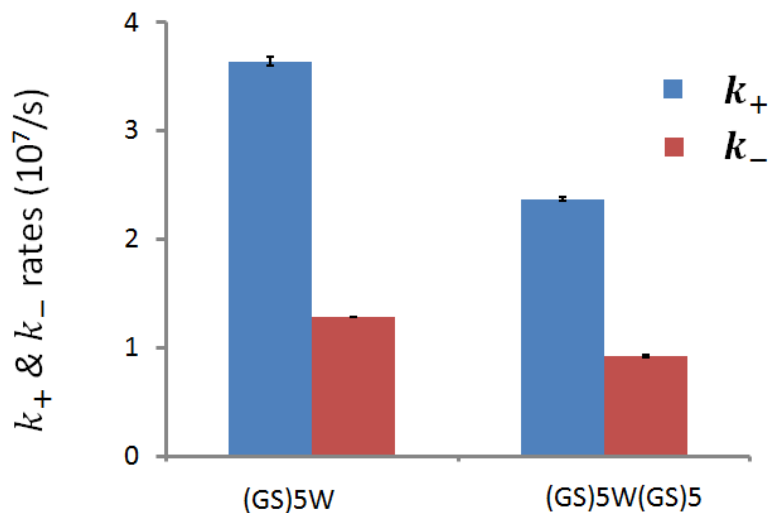
Figure 3.3.10 Association rate k_+ and dissociation rate k_- of two GS constructs with and without "tail".

How does the "tail" decrease the conformational dynamics of FG repeats and GS repeats?

Previous molecular dynamics simulations indicate that a stacked configuration is essential for the efficient quenching of Atto655 by Trp [139]. This means that the formation of a non-fluorescent complex requires a well-defined face-to-face stacking between the aromatic moieties of Atto655 and Trp. Experimental data also suggest that, on average, three collisional encounters between Atto655 and Trp are needed for one successful formation of a non-fluorescent complex [143], which implies that two thirds of the collisional encounters do not lead to the right interaction geometry. Therefore, two explanations are possible for the influence of a "tail" on the conformational dynamics: (1) The "tail" does not decrease the conformational dynamics of the peptide, but it only hinders the correct face-to-face orientation between Atto655 and Trp, which decreases the probability of non-fluorescent complex formation. (2) The "tail" does decrease the conformational dynamics of the peptide.

Following Kramer's theory, transitions between to conformations can be modeled as a diffusive process over a barrier [182]-[184]. For a diffusive process, the transition rate is proportional to the effective diffusion constant and thus inversely proportional to the solvent viscosity η. To investigate the role of solvent viscosity in the conformational dynamics of a peptide, Ansari and co-workers proposed that two sources of friction must be considered [185]: solvent friction η,

and internal friction σ. Solvent friction hinders the motion of atoms on the surface of the protein. Internal friction of the peptide slows down the motion of peptide atoms relative to each other. The relation between the rate of conformational transitions and the viscosity can be written as:

$$k_+ = \frac{C}{\eta + \sigma} \exp(-E_0/RT) \tag{22}$$

where η is the solvent viscosity, σ is interpreted as the "internal viscosity" of the peptide, R is the gas constant, T is the temperature, E_0 is the average height of the potential energy barrier separating the two protein conformations, and C is an adjustable parameter.

Three regions of viscosity which influence the rate of conformational transitions differently were observed: (1) Below ~1cP (the viscosity of water at 20°C), the solvent friction contributes only little to the decreasing of the rate; (2) Between 1 and 15 cP, both the protein friction and the solvent friction significantly contribute to the decreasing of the rate constant. (3) Above 15 cP, solvent friction dominates.

The solvent viscosity is ~1cP in this thesis; therefore, according to Ansari's assumption, the contribution of solvent friction to the rate of conformational transitions is negligible. Also, the introduction of a "tail" does not change the solvent viscosity. Instead, it may increase the "internal viscosity" of a peptide which could significantly slow down a protein's conformational dynamics.

3.3.4 Conformational dynamics of GS repeats

I performed also measurements on GS repeats, to compare data for a perfectly flexible IDP (GS repeat) with those obtained for different FG repeats. Normalized auto-correlation curves for different GS constructs are shown in figure 3.3.11. There, one can observe a systematic decrease of the PET decay amplitude with increasing length of the GS repeat. Because longer GS repeats have a larger distance between Atto655 and Trp, the systematic decrease of the PET decay amplitude is in good agreement with that observed for the FG repeats.
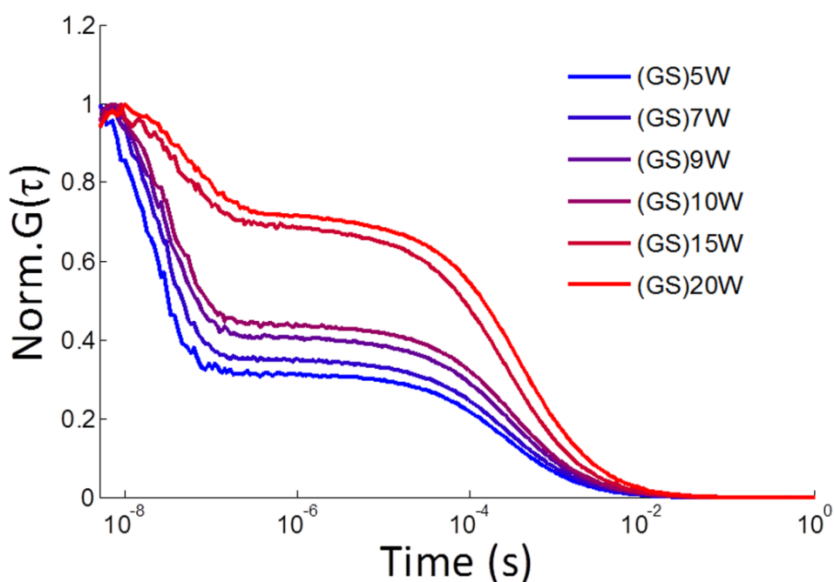
Figure 3.3.11 Normalized auto-correlation functions of GS repeats. Auto-correlation data recorded from different constructs are shown in a red-to-blue color scale (from blue to red represents short to long GS repeats.)

To further study the dependence of association rates, I fitted the data with an exponential and with a power-law, both for the data for the GS repeats and the FG repeats. The results are shown in figure 3.3.12.
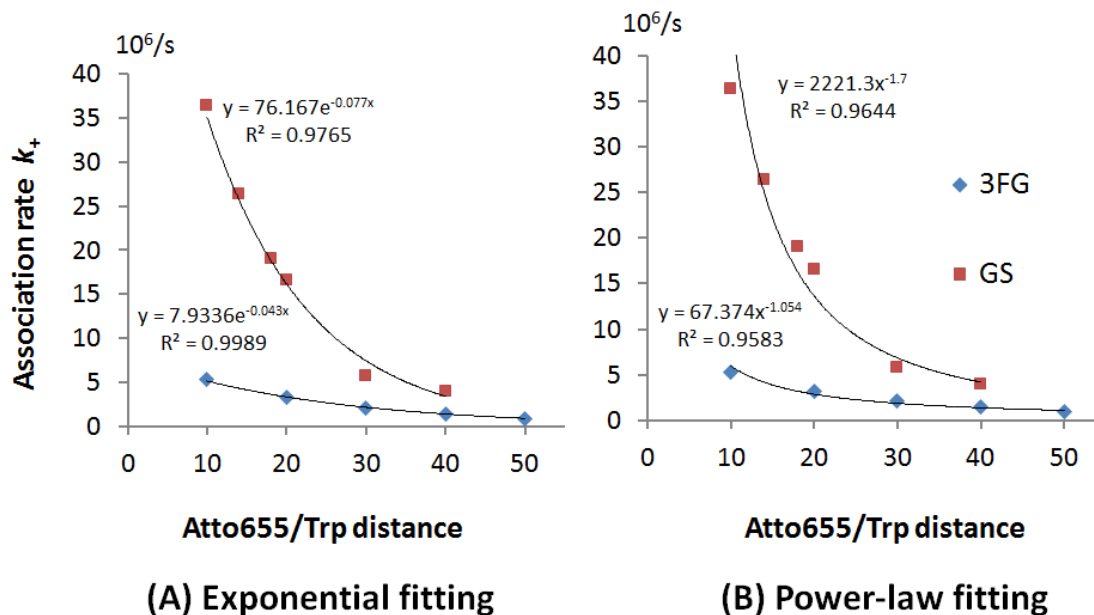
Figure 3.3.12 Dependence of the association rates of different 3FG and GS constructs fitted with an exponential decay (A) and with a power law (B). Blue and red squares represent 3FG and GS constructs, respectively. The length of standard error bar is comparable to or smaller than the size of the corresponding markers.

At a first glance, the exponential fitting seems to be better than the power-law fitting. However, this is not in good agreement with previous studies of the same GS repeats by Neuweiler *et al.* [143], where the association rates of long GS repeats showed a power-law dependence as a function of peptide length down to $N = 10$ (N is number of peptide bonds). The found scaling law for the on-rate was $k_+ \sim N^{-1.4 \pm 0.1}$. Moreover, the values for the k_+ rate in that study is slightly bigger than that obtained in this thesis. For example, Neuweiler *et al.* determined a value k_+ of (GS)5W of ~4.7×10^7/s, as compared to ~3.6×10^7/s in this thesis.

In the studies of Neuweiler *et al.*, measurements were performed in 10 mM sodium phosphate buffer (pH 7.0). The buffer in my case (150 mM NaCl) has a higher ion concentration, and therefore has a higher solvent viscosity. However, the variation of viscosity is less than 5 % when the concentration of NaCl is below ~500 mM (0.5 mol/kg salt molality) [186]. Therefore, the variation of association rates due to solvent viscosity is at maximum 5 %, as the viscosity of 10 mM sodium phosphate and 150 mM NaCl are ~1cP. Besides that, it is not fully clear why we

see such a rather big discrepancy between the on-rates reported by Neuweiler *et al.*, and the values measured in this thesis. ₊

3.4 **FG repeat aggregation**

How FG repeats form highly selective barriers in NPCs remains unknown. To address this question, two different models have been proposed, which differ in whether FG repeats can bind to each other or not [100]. In particular, the selective-phase model proposes that FG units bind to each other, whereas the virtual-gate model does not. To test the validity of these two models, potential aggregation properties of FG repeats under different concentrations were investigated with PET-FCS and 2fFCS.

In principle, aggregation begins with dimer or trimer formation as the concentration of the target protein increases. Dimer or trimer formation will significantly change the hydrodynamic radius and the conformational dynamics. Therefore, the hydrodynamic radius as measured by 2fFCS, and the conformational dynamics as measured with PET-FCS, can be used to monitor dimer or trimer formation.

Measurements were performed on mixed solutions of labelled peptides (nM concentration) with unlabeled peptides at different concentrations (up to 100 μM). Normalized auto-correlation curves for 3FG_10 at different concentrations are shown in figure 3.4.1.

Figure 3.4.1 Normalized auto-correlation functions for 3FG_10 at different peptide concentrations. Auto-correlation data recorded from Atto655md is shown in green. Auto-correlation data recorded at different peptide concentrations are shown with different colors (from blue to red: high to low concentration).

The obtained correlation curves were well fitted with the PET-FCS model as described above. The values of k_+, k_-, and of the relaxation time τ_R are given in table 3.4.1. The diffusion coefficients measured with 2fFCS at different concentrations are also included.

Table 3.4.1 Conformational dynamics data of 3FG_10 at different concentrations

Concentration	k_+ ($\times 10^6$/s)	k_- ($\times 10^6$/s)	Relaxation time τ_R (ns)	Diffusion coefficient ($\times 10^{-6}$ cm^2/s)
100 μM	4.21±0.04	6.17±0.02	96±0.7	1.55±0.02
10 μM	4.29±0.12	6.48±0.08	93±1.7	1.50±0.03
1 μM	5.18±0.02	5.62±0.02	92±0.4	1.56±0.02
100 nM	5.07±0.06	5.73±0.05	93±1.0	1.63±0.05
10 nM	4.90±0.10	5.51±0.10	96±2.0	1.55±0.01

Samples were prepared several days in advance and stored at −80 °C. Therefore, the measured values of k_+ and k_- are slightly different from those measured for freshly prepared samples in section 3.3.2.

The values of k_+, k_-, the relaxation time τ_R, and the diffusion coefficients are practically constant for all studied concentrations. One observes only slightly bigger values of k_+ and slightly lower values of k_- at 10 nM, 100 nM and 1 μM. However, as it is expected that any dimer or trimer formation has to significantly change the k_+ and k_- values, the practically constant relaxation time indicates that one finds the same conformational dynamics for different concentrations. Also, the practically constant diffusion coefficient suggests that no dimer or trimer formation occurs. The slightly varying values of association and dissociation rates might be attributed to the high concentration of acetonitrile (37%, v/v) in the peptide dissolving buffer of the stock solution. When the diluted solutions of unlabeled FG peptides are prepared, the concentration of acetonitrile in the buffer is also diluted. Thus, high concentrations of acetonitrile in 100 μM and 10 μM solutions of unlabeled peptides might slightly effect the conformational dynamics.

To investigate the potential aggregation dynamics more precisely, PET-FCS and 2fFCS measurements were performed also for other 3FG construct at serial concentrations using a dilution factor of 1:5. The normalized auto-correlation curves of 3FG_20 at different concentrations are shown in figure 3.4.2. One observes that the curves do practically not change for the different studied concentrations.
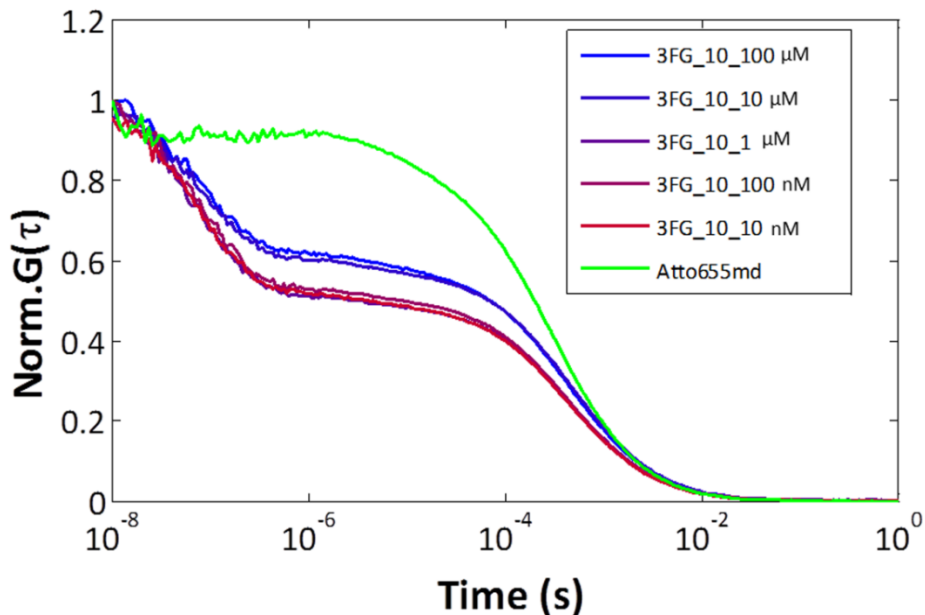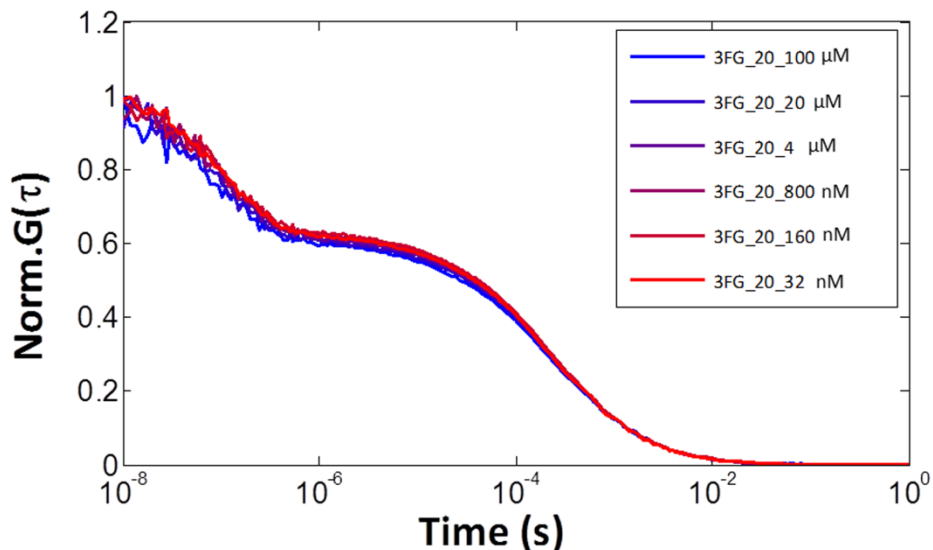
Figure 3.4.2 Normalized auto-correlation curves for 3FG_20 at different peptide concentrations. Auto-correlation data recorded at different peptide concentrations are shown in different colors (from blue to red: high to low concentration).

The values of k_+, k_-, the relaxation time τ_R, and the diffusion coefficient of 3FG_20 are given in table 3.4.2.

Table 3.4.1 Conformational dynamics data of 3FG_20 for different concentrations

Concentration	k_+ ($\times 10^6$/s)	k_- ($\times 10^6$/s)	Relaxation time τ_R (ns)	Diffusion coefficient ($\times 10^{-6}$ cm^2/s)
100 μM	2.95±0.05	5.26±0.03	121±1.3	1.53±0.004
20 μM	2.94±0.16	5.46±0.16	119±3.1	1.49±0.01
4 μM	2.43±0.25	5.30±0.13	129±3.6	1.52±0.01
800 nM	2.49±0.18	5.59±0.21	123±0.7	1.55±0.002
160 nM	2.39±0.15	5.57±0.25	125±2.3	1.55±0.007
32 nM	3.01±0.09	5.02±0.08	124±1.5	1.50±0.005

It occurs that the variation of peptide concentration does not lead to a systematic change of the conformational dynamics of the peptides. The k_+ and k_- values remain practically constant upon variation of peptide concentration, up to 100 µM, which indicates that no dimer or trimer formation takes place. Furthermore, no aggregation was observed in PBS buffer during the purification steps of all FG repeats. The peptide concentration upon purification is in the µM range. The ion concentrations of PBS are close to those in living cells. Thus, it can be concluded that short FG repeats (up to 50 amino acids) at low concentration (up to 100 µM) do not tend to aggregate under physiological condition.

Nsp1 is supposed to be able to form a hydrogel [111, 114]. However, the gel formation requires a very high concentration (200mg/ml) and very strict initiation buffer (not physiological buffer), conditions which are unlikely to be found within a living cell. Previous studies of the same Nsp1 protein (2-601 aa) with CD spectroscopy at concentration of 30 mg/ml did also not observe any hydrogel formation [187, 188]. Also, other studies did not observe any interaction of FG domains of Nsp1 with itself or with other FG domains, even at high concentration (about 500 µM) [100]. These results are in good agreement with the data obtained in this thesis.

It is also very interesting to estimate the actual concentration of FG repeats in the NPC. The diameter of the whole NPC is ~ 100 nm. The diameter of the central channel of the NPC is ~ 38 nm, and the height of the NPC is ~ 37 nm [96]. Therefore, the volume of NPC can be calculated as ~ 2.9 ×10^{-19} L, and the volume of the central channel is ~ 4.2 ×10^{-20} L. A fundamental symmetry unit of the NPC is the spoke. NPC contains eight spokes arranged radially around a central channel (see figure 1.5.1). Four copies of Nsp1 can be found in each spoke of NPC. Thus, the concentration of Nsp1 can be estimated as ~ 1.3 mM in the central channel.

Based on the data above, the actual concentration of Nsp1 in the NPC is 10 times higher than the concentration studied in this thesis (100 µM) and 2-3 times higher than that used for gel formation assays (500 µM). This implies that some gel-like structure might be formed in the NPC. However, as mentioned above, Nsp1 interacts with many other Nups to form two complexes in the NPC. Some studies suggest that Nsp1 forms a platform for the biding of many

NPC constituents. This indicates that the two Nsp1 molecules within one of the eight NPC spokes do not interact with each other. Thus, the high selectivity of the NPC barrier is probably due to the interaction between many Nups, or at most to the interaction of Nsp1 molecules between spokes, or with other Nups.

4. Conclusion & Outlook

In this thesis, the conformational structure and dynamics of intrinsically disordered proteins (IDPs), namely the FG repeats form the nuclear pore complex of eukaryotic cells, was investigated with PET-FCS and with 2fFCS. In the PET-FCS measurements, one dye (Atto655) and one Trp amino acid were introduced at specific positions of the peptides to generate a PET dye/quencher pair. In order to test the influence of the fluorescent labeling on the properties of the peptides, I measured the diffusion coefficients of all constructs with 2fFCS. From these measurements, I calculated values of the hydrodynamic radius. I found that these values of hydrodynamic radius for different constructs were basically the same and did not depend on the exact position of the Trp or the dye. Moreover, the calculated values of hydrodynamic radius follow nicely a relationship between molecular size and diffusion coefficient as previously found by Anastasia Loman. This again suggests that the proteins studied in this thesis are intrinsically disordered proteins, and that the fluorescent labelling does not sensibly influence their diffusion behavior.

Next, MD simulations were used to investigate the structure and dynamics of IDPs. The simulation data demonstrates an unexpectedly high sensitivity of computed IDP conformations on the employed force field. For comparison of MD simulations with experiment, I measured the hydrodynamic radii of various IDPs: FG repeats, GS repeats, Prothymosin alpha, and Integrase. The method I used was 2fFCS, which yields precise and absolute values of diffusion coefficients. The comparison of experimental data with the theoretical predictions showed that the conformations obtained with the force field CHARMM 22* and a charm-modified TIP3P water agrees best with the actual experimental data.

Furthermore, I studied the difference in conformational dynamics between the N- and C-terminus of Nsp1. The obtained experimental data suggests substantially different behaviors of the N- and the C-terminus of Nsp1. In particular, the N-terminus, which shows random coil behavior, tends to be more flexible than the C-terminus. During the PET-FCS measurements, I

investigated systematically which additional buffer reagents can be used for an efficient suppression of unspecific absorption of proteins to the glass surface.

Finally, I investigated the aggregation behavior of the N- terminus of Nsp1, again using PET-FCS and 2fFCS. My data were taken at physiological peptide concentrations and under physiological buffer conditions. It indicates that short FG repeats (up to 50 amino acids) at low concentration (up to 100 µM) do not tend to aggregate under physiological condition.

For future work, it would be very interesting to investigate the conformational dynamics of full length Nsp1. One possibility is that the gel formation might depend not only on the hydrophobic interaction between different FG units, but also on the numbers of FG units in one protein, similar to the wild type Nsp1 fragment which was used for gel formation experiments, and which contains 601 amino acids and more than 30 FG units (see Appendix). It would also be advisable to mix short peptides with full length Nsp1 to investigate how the conformational dynamics of short Nsp1 fragments is affected by the crowding of full length Nsp1. It would be also interesting to investigate the conformational dynamics of Nsp1 in the interaction with other Nups such as Nup49, Nup57, Nic96, Nup82 and Nup159. I*n vivo* studies could also be used for investigating the conformational dynamics of FG repeats in the context of cellular NPCs.

Appendix

The algorithms and *Matlab* scripts used for PET-FCS analysis in this work were developed and written by Dr. Ingo Gregor and Prof. Dr. Jörg Enderlein from University of Göttingen.

(A)   (B)
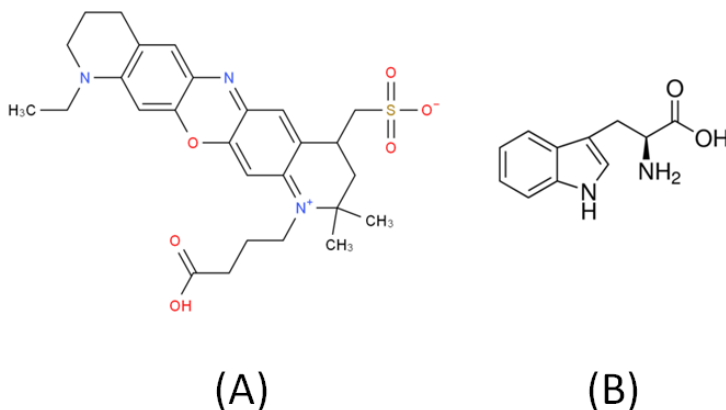
Figure A.1. Molecular structures: (A) Atto655 and (B) Trp (taken from http://en.chembase.cn/

Table A.1: Hydrodynamic radii (R_H) of Prothymosin alpha and Integrase

Samples	R_H (Å) ±SEM
Atto655md	0.60
Prot_A_attoOXA11	2.80±0.01
Cut_his_Prot_A_attoOXA11	2.84±0.02
Integrase_alexa488	1.90±0.02

ProTα coupled with Atto Oxa 11 and Integrase coupled with Alexa488. Atto655md is used as reference. Diffusion coefficient of Atto655md in water at 25 °C is determined as 4× 10^{-6} cm^2/s. R_H of Atto655md is 6 Å.

Gel formation protocol (adapted form ref. [111])

Starting point for gelation had been salt- and solvent-free Nsp1-repeats. Therefore, purified repeats were dialyzed against 6M deionized urea and precipitated with 4 volumes 98% ethanol, 40 mM potassium acetate. The precipitate was washed twice with 98% ethanol, 40 mM potassium acetate and once with 100% diethyl ether and dried under vacuum.

Gelation was initiated by dissolving dried Nsp1 repeat-domains at 500 µM in 50 mM KOH, 1 mM EDTA yielding a pH of ~11, followed by quick neutralization with 1/4 volume neutralization buffer (250 mM K2HPO4/KH2PO4 pH 7.5; 50 mM KOH; 250 mM acetic acid; 1 mM EDTA). Resulting gels were analyzed 20-72h later.

wild-type
Nsp1 repeat-domain

MNFNTPQQNK
TPFSFGTANNNSNTTNQNSST
GAGAFGTGQ
STFGFNNSAPNNTNNANSS
ITPAFGSNNT
GNTAFGNSNP
TSNVFGSNNS
TTNTFGSNSA
GTSLFGSSSAQQTKSNGTA
GGNTFGSSSLFNNSTNSNT
TKPAFG
  GLNFGGGNNTTPSSTGNANT
SNNLFGATANANK
PAFSFGATTNDDKKTEPDK
PAFSFNSSVGNKTDAQAPT
TGFSFGSQLGGNKTVNEAAK
PSLSFGSGSAGANPAGASQPEPTTNEPAK
PALSFGTATSDNKTTNTT
PSFSFGAKSDENKAGATSK
PAFSFGAKPEEKKDDNSSK
PAFSFGAKSNEDKQDGTAK
PAFSFGAKPAEKNNNETSK
PAFSFGAKSDEKKDGDASK
PAFSFGAKPDENKASATSK
PAFSFGAKPEEKKDDNSSK
PAFSFGAKSNEDKQDGTAK
PAFSFGAKPAEKNNNETSK
PAFSFGAKSDEKKDGDASK
PAFSFGAKSDEKKDSDSSK
PAFSFGTKSNEKKDSGSSK
PAFSFGAKPDEKKNDEVSK
PAFSFGAKANEKKESDESK
SAFSFGSKPTGKEEGDGAK
AAISFGAKPEEQKSSDTSK
PAFTFGAQKDNEKKTE

mutations
Phe→Ser

F→S mutated
Nsp1 repeat-domain

MNFNTPQQNK
TPSSSGTANNNSNTTNQNSST
GAGASGTGQ
STSGSNNSAPNNTNNANSS
ITPASGSNNT
GNTASGNSNP
TSNVSGSNNS
TTNTSGSNSA
GTSLSGSSSAQQTKSNGTA
GGNTSGSSSLFNNSTNSNT
TKPASG
  GLNSGGGNNTTPSSTGNANT
SNNLSGATANANK
PASSSGATTNDDKKTEPDK
PASSSNSSVGNKTDAQAPT
TGSSSGSQLGGNKTVNEAAK
PSLSSGSGSAGANPAGASQPEPTTNEPAK
PALSSGTATSDNKTTNTT
PSSSSGAKSDENKAGATSK
PASSSGAKPEEKKDDNSSK
PASSSGAKSNEDKQDGTAK
PASSSGAKPAEKNNNETSK
PASSSGAKSDEKKDGDASK
PASSSGAKPDENKASATSK
PASSSGAKPEEKKDDNSSK
PASSSGAKSNEDKQDGTAK
PASSSGAKPAEKNNNETSK
PASSSGAKSDEKKDGDASK
PASSSGAKSDEKKDSDSSK
PASSSGTKSNEKKDSGSSK
PASSSGAKPDEKKNDEVSK
PASSSGAKANEKKESDESK
SASSSGSKPTGKEEGDGAK
AAISSGAKPEEQKSSDTSK
PASTSGAQKDNEKKTE

Figure A.2. Amino acid sequences of wild-type Nsp1 used for gel formation and F→S mutated Nsp1 that cannot form hydrogel (taken from ref.[111])**.**

Left: Amino acid sequences (residues 1-601 in one-letter code) of the wild-type FG repeat-domain of Nsp1. Right: the F→S mutated repeat domain of Nsp1, in which Phenylalanine (F) had been mutated into Serine (S).
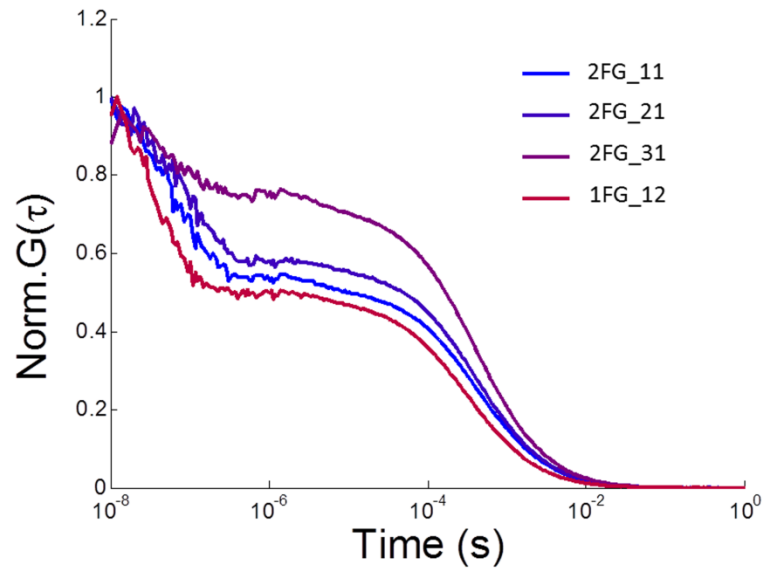
Figure A.3. Normalized auto-correlation functions of 1FG and 2FG.
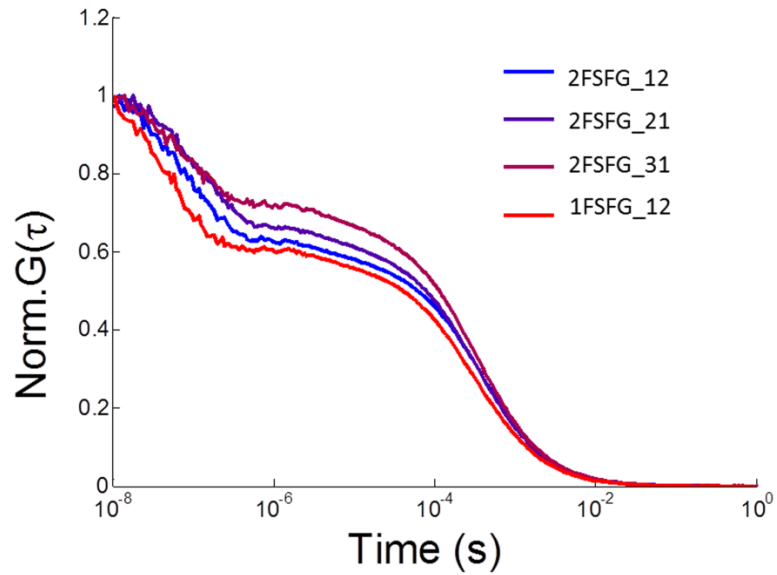
Figure A.4.Normalized auto-correlation functions of 1FSFG and 2FSFG.

Table A2: Conformational dynamics data of FG, FSFG and GS repeats

Peptides	$K_+ (\times10^6/s)$ ±SEM	$K_- (\times10^6/s)$ ±SEM	Peptides	$K_+ (\times10^6/s)$ ±SEM	$K_- (\times10^6/s)$ ±SEM
3FG_10	5.29±0.07	4.25±0.04	3FSFG_12	2.97±0.04	4.64±0.05
3FG_20	3.24±0.05	4.07±0.03	3FSFG_21	1.38±0.02	4.50±0.04
3FG_30	2.13±0.07	4.78±0.11	3FSFG_31	1.57±0.02	4.38±0.03
3FG_40	1.43±0.06	4.67±0.13	3FSFG_40	1.00±0.02	3.83±0.05
3FG_50	0.92±0.01	8.07±0.08	3FSFG_50	1.25±0.02	6.20±0.11
2FG_11	2.90±0.08	9.19±0.18	2FSFG_12	3.39±0.04	5.42±0.04
2FG_21	3.81±0.03	5.25±0.03	2FSFG_21	2.48±0.02	4.41±0.04
2FG_31	5.62±0.08	5.92±0.06	2FSFG_31	2.91±0.03	6.74±0.05
1FG_12	12.55±0.24	10.16±0.10	1FSFG_12	6.42±0.08	8.72±0.04
(GS)5W	36.4±0.4	12.8±0	(GS)10W	16.6±0.2	11.0±0
(GS)7W	26.4±0.2	11.4±0.1	(GS)15W	5.8±0.1	12.4±0.1
(GS)9W	19.1±0.2	11.2±0.1	(GS)20W	4.0±0.0	9.9±0.0

Bibliography

[1]   A. K. Dunker, J. D. Lawson, C. J. Brown, R. M. Williams, P. Romero, J. S. Oh, C. J. Oldfield, A. M. Campen, C. M. Ratliff, K. W. Hipps, J. Ausio, M. S. Nissen, R. Reeves, C. Kang, C. R. Kissinger, R. W. Bailey, M. D. Griswold, W. Chiu, E. C. Garner, and Z. Obradovic, "Intrinsically disordered protein," *J. Mol. Graph. Model.*, vol. 19, no. 1, pp. 26–59, Feb. 2001.

[2]   E. Fischer, "Einfluss der Configuration auf die Wirkung der Enzyme," *Berichte der Dtsch. Chem. Gesellschaft*, vol. 27, no. 3, pp. 2985–2993, 1894.

[3]   P. E. Wright and H. J. Dyson, "Intrinsically disordered proteins in cellular signalling and regulation," *Nat. Rev. Mol. Cell Biol.*, vol. 16, no. 1, pp. 18–29, Dec. 2015.

[4]   A. K. Dunker, C. J. Brown, J. D. Lawson, L. M. Iakoucheva, and Z. Obradovic, "Intrinsic disorder and protein function," *Biochemistry*, vol. 41, no. 21, 2002.

[5]   P. E. Wright and H. J. Dyson, "Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm," *J. Mol. Biol.*, vol. 293, no. 2, pp. 321–331, Oct. 1999.

[6]   M. Sickmeier, J. A. Hamilton, T. LeGall, V. Vacic, M. S. Cortese, A. Tantos, B. Szabo, P. Tompa, J. Chen, V. N. Uversky, Z. Obradovic, and A. K. Dunker, "DisProt: the database of disordered proteins" *Nucleic Acids Res.*, vol. 35, no. Database issue, pp. D786–93, Jan. 2007.

[7]   T. Chouard, "Breaking the protein rules" *Nat. news*, vol. 471, pp. 151–153, 2011.

[8]   H. J. Dyson and P. E. Wright, "Coupling of folding and binding for unstructured proteins" *Curr. Opin. Struct. Biol.*, vol. 12, no. 1, pp. 54–60, Feb. 2002.

[9]   H. J. Dyson and P. E. Wright, "Intrinsically unstructured proteins and their functions" *Nat. Rev. Mol. Cell Biol.*, vol. 6, no. 3, pp. 197–208, Mar. 2005.

[10]   A. L. Fink, "Natively unfolded proteins" *Curr. Opin. Struct. Biol.*, vol. 15, no. 1, pp. 35–41, Feb. 2005.

[11] I. Carlberg, M. Hansson, T. Kieselbach, W. P. Schröder, B. Andersson, and A. V. Vener, "A novel plant protein undergoing light-induced phosphorylation and release from the photosynthetic thylakoid membranes" *Proc. Natl. Acad. Sci. U. S. A.*, vol. 100, no. 2, pp. 757–762, 2003.

[12] J. Song, M. S. Lee, I. Carlberg, A. V. Vener, and J. L. Markley, "Micelle-induced folding of spinach thylakoid soluble phosphoprotein of 9 kDa and its functional implications" *Biochemistry*, vol. 45, no. 51, pp. 15633–15643, Dec. 2006.

[13] P. Bayer, A. Arndt, S. Metzger, R. Mahajan, F. Melchior, R. Jaenicke, and J. Becker, "Structure determination of the small ubiquitin-related modifier SUMO-1" *J. Mol. Biol.*, vol. 280, no. 2, pp. 275–286, 1998.

[14] M. J. Matunis, E. Coutavas, and G. Blobel, "A novel ubiquitin-like modification modulates the partitioning of the Ran-GTPase-activating protein RanGAP1 between the cytosol and the nuclear pore complex" *J. Cell Biol.*, vol. 135, no. 6, pp. 1457–1470, 1996.

[15] R. Mahajan, C. Delphin, T. Guan, L. Gerace, and F. Melchior, "A small ubiquitin-related polypeptide involved in targeting RanGAP1 to nuclear pore complex protein RanBP2" *Cell*, vol. 88, no. 1, pp. 97–107, 1997.

[16] K. Majorek, L. Kozlowski, M. Jakalski, and J. M. Bujnicki, "Prediction of protein structures, functions, and interactions " Book, (2008): 39-62

[17] H. Frauenfelder, S. G. Sligar, and P. G. Wolynes, "The energy landscapes and motions of proteins" *Science (80-.).*, vol. 254, pp. 1598–1603, 1991.

[18] R. H. Austin, K. W. Beeson, L. Eisenstein, H. Frauenfelder, and I. C. Gunsalus, "Dynamics of ligand binding to myoglobin" *Biochemistry*, vol. 14, no. 24, pp. 5355–5373, 1975.

[19] W. Doster, "Brownian oscillator analysis of molecular motions in biomolecules" in *Neutron Scattering in Biology*, 2006, pp. 461–483.

[20] S. Cusack and W. Doster, "Temperature dependence of the low frequency dynamics of myoglobin. Measurement of the vibrational frequency distribution by inelastic neutron scattering" *Biophys. J.*, vol. 58, no. 1, pp. 243–251, 1990.

[21]   W. Doster and M. Settles, "Protein-water displacement distributions" *Biochim. Biophys. Acta - Proteins Proteomics*, vol. 1749, no. 2, pp. 173–186, 2005.

[22]   S. Khodadadi and A. P. Sokolov, "Protein dynamics: from rattling in a cage to structural relaxation" *Soft Matter*, vol. 11, no. 25, pp. 4984–98, 2015.

[23]   D. E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood, J. A. Bank, J. M. Jumper, J. K. Salmon, Y. Shan, and W. Wriggers, "Atomic-level characterization of the structural dynamics of proteins" *Science,* vol. 330, no.6002 pp. 341–346, 2010.

[24]   H. Frauenfelder, G. Chen, J. Berendzen, P. W. Fenimore, H. Jansson, B. H. McMahon, I. R. Stroe, J. Swenson, and R. D. Young, "A unified model of protein dynamics" *Proc. Natl. Acad. Sci. U. S. A.*, vol. 106, no. 13, pp. 5129–5134, 2009.

[25]   F. Massi, M. J. Grey, and A. G. Palmer, "Microsecond timescale backbone conformational dynamics in ubiquitin studied with NMR R1rho relaxation experiments" *Protein Sci.*, vol. 14, no. 3, pp. 735–742, 2005.

[26]   A. C. M. Ferreon, C. R. Moran, Y. Gambin, and A. A. Deniz, "Single-molecule fluorescence studies of intrinsically disordered proteins" *Methods Enzymol.*, vol. 472, no. 10, pp. 179–204, Jan. 2010.

[27]   S. Mukhopadhyay, R. Krishnan, E. A. Lemke, S. Lindquist, and A. A. Deniz, "A natively unfolded yeast prion monomer adopts an ensemble of collapsed and rapidly fluctuating structures" *Proc. Natl. Acad. Sci. U. S. A.*, vol. 104, no. 8, pp. 2649–2654, 2007.

[28]   X. Michalet, S. Weiss, and M. Jäger, "Single-molecule fluorescence studies of protein folding and conformational dynamics" *Chem. Rev.*, vol. 106, pp. 1785–1813, 2006.

[29]   R. Van Der Lee, M. Buljan, B. Lang, R. J. Weatheritt, G. W. Daughdrill, A. K. Dunker, M. Fuxreiter, J. Gough, J. Gsponer, D. T. Jones, P. M. Kim, R. W. Kriwacki, C. J. Old, R. V Pappu, P. Tompa, V. N. Uversky, P. E. Wright, and M. M. Babu, "Classification of intrinsically disordered regions and proteins" *Chem. Rev.*, vol. 114, pp. 6589–6631, 2014.

[30]   P. Tompa, "The interplay between structure and function in intrinsically unstructured

proteins" *FEBS Lett.*, vol. 579, no. 15, pp. 3346–3354, Oct. 2005.

[31]   V. N. Uversky, "The most important thing is the tail: Multitudinous functionalities of intrinsically disordered protein termini" *FEBS Lett.*, vol. 587, no. 13, pp. 1891–1901, 2013.

[32]   G. W. Daughdrill, P. Narayanaswami, S. H. Gilmore, A. Belczyk, and C. J. Brown, "Dynamic behavior of an intrinsically unstructured linker domain is conserved in the face of negligible amino acid sequence conservation" *J. Mol. Evol.*, vol. 65, no. 3, pp. 277–288, 2007.

[33]   T. Kouzarides, "Chromatin modifications and their function" *Cell*, vol. 128, no. 4, pp. 693–705, 2007.

[34]   C. A. Galea, A. Nourse, Y. Wang, S. G. Sivakolundu, W. T. Heller, and R. W. Kriwacki, "Role of intrinsic flexibility in signal transduction mediated by the cell cycle regulator, p27 Kip1" *J. Mol. Biol.*, vol. 376, no. 3, pp. 827–38, 2008.

[35]   A. M. Bode and Z. Dong, "Post-translational modification of p53 in tumorigenesis" *Nat. Rev. Cancer*, vol. 4, no. 10, pp. 793–805, 2004.

[36]   L. Rajkowitsch, K. Semrad, O. Mayer, and R. Schroeder, "Assays for the RNA chaperone activity of proteins" *Biochem. Soc. Trans.*, vol. 33, no. 3, pp. 450–456, 2005.

[37]   J. C. Young, V. R. Agashe, K. Siegers, and F. U. Hartl, "Pathways of chaperone-mediated protein folding in the cytosol" *Nat. Rev. Mol. Cell Biol.*, vol. 5, no. 10, pp. 781–791, 2004.

[38]   P. Tompa and P. Csermely, "The role of structural disorder in the function of RNA and protein chaperones" *FASEB J.*, vol. 18, no. 11, pp. 1169–1175, 2004.

[39]   R. Ivanyi-Nagy, L. Davidovic, E. W. Khandjian, and J.-L. Darlix, "Disordered RNA chaperone proteins: from functions to disease" *Cell. Mol. Life Sci.*, vol. 62, no. 13, pp. 1409–17, 2005.

[40]   P. E. Wright and H. J. Dyson, "Linking folding and binding" *Curr. Opin. Struct. Biol.*, vol. 19, no. 1, pp. 31–38, 2009.

[41]   K. Sugase, H. J. Dyson, and P. E. Wright, "Mechanism of coupled folding and binding of an

intrinsically disordered protein" *Nature*, vol. 447, no. 7147, pp. 1021–5, Jun. 2007.

[42] C. A. Galea, Y. Wang, S. G. Sivakolundu, and R. W. Kriwacki, "Regulation of cell division by intrinsically unstructured proteins: Intrinsic flexibility, modularity, and signaling conduits" *Biochemistry*, vol. 47, no. 29, pp. 7598–7609, 2008.

[43] H. Wu, "Higher-order assemblies in a new paradigm of signal transduction" *Cell*, vol. 153, no. 2, pp. 287–292, 2013.

[44] A. Cumberworth, G. Lamour, M. M. Babu, and J. Gsponer, "Promiscuity as a functional trait: intrinsically disordered regions as central players of interactomes" *Biochem. J.*, vol. 454, no. 3, pp. 361–9, 2013.

[45] Z. Peng, C. J. Oldfield, B. Xue, M. J. Mizianty, A. K. Dunker, L. Kurgan, and V. N. Uversky, "A creature with a hundred waggly tails: Intrinsically disordered proteins in the ribosome" *Cell. Mol. Life Sci.*, vol. 71, no. 8, pp. 1477–1504, 2014.

[46] Jixi Li, T. McQuade, A. B. Siemer, J. Napetschnig, K. Moriwaki, Yu-Shan Hsiao, E. Damko, D. Moquin, T. Walz, A. McDermott, F. K. M. Chan, and Hao. Wu, "The RIP1/RIP3 necrosome forms a functional amyloid signaling complex required for programmed necrosis" *Cell*, vol. 150, no. 2, pp. 339–350, 2012.

[47] M. Fuxreiter, P. Tompa, I. Simon, V. N. Uversky, J. C. Hansen, and F. J. Asturias, "Malleable machines take shape in eukaryotic transcriptional regulation" *Nat. Chem. Biol.*, vol. 4, no. 12, pp. 728–737, 2008.

[48] K. Gunasekaran, C.-J. Tsai, S. Kumar, D. Zanuy, and R. Nussinov, "Extended disordered proteins: targeting function with less scaffold" *Trends Biochem. Sci.*, vol. 28, no. 2, pp. 81–85, 2003.

[49] A. J. Daniels, R. J. P. Williams, and P. E. Wright, "The character of the stored molecules in chromaffin granules of the adrenal medulla: A nuclear magnetic resonance study" *Neuroscience*, vol. 3, no. 6, pp. 573–585, 1978.

[50] M. M. Babu, R. van der Lee, N. S. de Groot, and J. Gsponer, "Intrinsically disordered proteins: regulation and disease" *Curr. Opin. Struct. Biol.*, vol. 21, no. 3, pp. 432–40, Jul. 2011.

[51] V. N. Uversky, C. J. Oldfield, and A. K. Dunker, "Intrinsically disordered proteins in human diseases: introducing the D2 concept" *Annu. Rev. Biophys.*, vol. 37, pp. 215–46, Jan. 2008.

[52] S. Raychaudhuri, S. Dey, N. P. Bhattacharyya, and D. Mukhopadhyay, "The role of intrinsically unstructured proteins in neurodegenerative diseases" *PLoS One*, vol. 4, no. 5, p. e5566, May 2009.

[53] M. Wells, H. Tidow, T. J. Rutherford, P. Markwick, M. R. Jensen, E. Mylonas, D. I. Svergun, M. Blackledge, and a. R. Fersht, "Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain" *Proc. Natl. Acad. Sci.*, vol. 105, no. 15, pp. 5762–5767, Apr. 2008.

[54] A. R. Venkitaraman, "Cancer susceptibility and the functions of BRCA1 and BRCA2" *cell Rev.*, vol. 108, pp. 171–182, 2002.

[55] P. H. Weinreb, W. Zhen, A. W. Poon, K. A. Conway, and P. T. Lansbury, "NACP , A protein implicated in Alzheimer ' s disease and learning , is natively unfolded" *Biochemistry*, vol. 35, no. 43, pp. 13709–13715, 1996.

[56] M. D. Mukrasch, S. Bibow, J. Korukottu, S. Jeganathan, J. Biernat, C. Griesinger, E. Mandelkow, and M. Zweckstetter, "Structural polymorphism of 441-residue Tau at single residue resolution" *PLoS Biol.*, vol. 7, no. 2, p. e1000034, Feb. 2009.

[57] E. T. Jaikaran, C. E. Higham, L. C. Serpell, J. Zurdo, M. Gross, A. Clark, and P. E. Fraser, "Identification of a novel human islet amyloid polypeptide beta-sheet domain and factors influencing fibrillogenesis" *J. Mol. Biol.*, vol. 308, no. 3, pp. 515–25, 2001.

[58] L. E. Kay, "Protein dynamics from NMR 1" *Cell*, vol. 76, pp. 145–152, 1998.

[59] C. Bracken, "NMR spin relaxation methods for characterization of disorder and folding in proteins" *J. Mol. Graph. Model.*, vol. 19, no. 1, pp. 3–12, 2001.

[60] J. Gu and V. J. Hilser, "The significance and impacts of protein disorder and conformational variants" *Struct. Bioinforma.*, pp. 937–960, 2009.

[61] D. A. Jacques and J. Trewhella, "Small-angle scattering for structural biology - Expanding the frontier while avoiding the pitfalls" *Protein Sci.*, vol. 19, no. 4, pp. 642–657, 2010.

[62]   L. Pollack and S. Doniach, "Time-resolved X-ray scattering and RNA folding" *Methods Enzymol.*, vol. 469, pp. 253–268, 2009.

[63]   P. Bernadó and D. I. Svergun, "Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering" *Mol. BioSyst.*, vol. 8, no. 1, pp. 151–167, 2012.

[64]   S. J. Hubbard, F. Eisenmenger, and J. M. Thornton, "Modeling studies of the change in conformation required for cleavage of limited proteolytic sites" *Protein Sci.*, vol. 3, no. 5, pp. 757–68, 1994.

[65]   A. Rahman, "Correlations in the motion of atoms in liquid silicon" *Phys. Rev.*, vol. 136, no. 2A, p. A405, 1964.

[66]   B. J. Alder and T. E. Wainwright, "Studies in molecular dynamics. I. General method" *J. Chem. Phys.*, vol. 2, no. 31, pp. 459–466., 1959.

[67]   A. D. MacKerell, N. Banavali, and N. Foloppe, "Development and current status of the CHARMM force field for nucleic acids" *Biopolymers*, vol. 56, no. 4, pp. 257–65, 2001.

[68]   B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, "CHARMM: A program for macromolecular energy, minimization, and dynamics calculations" *J. Comput. Chem.*, vol. 4, no. 2, pp. 187–217, 1983.

[69]   W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, "A second generation force field for the simulation of proteins, nucleic acids, and organic molecules" *J. Am. Chem. Soc.*, vol. 117, no. 19, pp. 5179–5197, 1995.

[70]   W. L. Jorgensen and J. Tirado-Rives, "The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin" *J. Am. Chem. Soc.*, vol. 110, no. 6, pp. 1657–1666, 1988.

[71]   K. A. Beauchamp, Y. S. Lin, R. Das, and V. S. Pande, "Are protein force fields getting better? A systematic benchmark on 524 diverse NMR measurements" *J. Chem. Theory Comput.*, vol. 8, no. 4, pp. 1409–1414, 2012.

[72]   O. F. Lange, D. van der Spoel, and B. L. de Groot, "Scrutinizing molecular mechanics force fields on the submicrosecond timescale with NMR data" *Biophys. J.*, vol. 99, no. 2, pp. 647–55, 2010.

[73]   K. Lindorff-Larsen, P. Maragakis, S. Piana, M. P. Eastwood, R. O. Dror, and D. E. Shaw, "Systematic validation of protein force fields against experimental data" *PLoS One*, vol. 7, no. 2, pp. 1–6, 2012.

[74]   M. M. Dedmon, K. Lindorff-Larsen, J. Christodoulou, M. Vendruscolo, and C. M. Dobson, "Mapping long-range interactions in α-synuclein using spin-label NMR and ensemble molecular dynamics simulations" *J. Am. Chem. Soc.*, vol. 127, no. 2, pp. 476–477, 2005.

[75]   J. R. Allison, P. Varnai, C. M. Dobson, and M. Vendruscolo, "Determination of the free energy landscape of alpha-synuclein using spin label nuclear magnetic resonance measurements" *J. Am. Chem. Soc.*, vol. 131, no. 51, pp. 18314–18326, 2009.

[76]   A. Vitalis and R. V. Pappu, "ABSINTH: A new continuum solvation model for simulations of polypeptides in aqueous solutions" *J. Comput. Chem.*, vol. 30, pp. 673–699, 2009.

[77]   W. W. Smith, C. F. Schreck, N. Hashem, S. Soltani, A. Nath, E. Rhoades, and C. S. O'Hern, "Molecular simulations of the fluctuating conformational dynamics of intrinsically disordered proteins" *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.*, vol. 86, pp. 1–9, 2012.

[78]   E. Yedvabny, P. S. Nerenberg, C. So, and T. Head-Gordon, "Disordered structural ensembles of vasopressin and oxytocin and their mutants" *J. Phys. Chem. B*, vol. 119, no. 3, pp. 896–905, 2014.

[79]   N. Stanley, S. Esteban-Martín, and G. de Fabritiis, "Kinetic modulation of a disordered protein domain by phosphorylation" *Nat. Commun.*, vol. 5, no. May, p. 5272, 2014.

[80]   S. R. Gerben, J. A. Lemkul, A. M. Brown, and D. R. Bevan, "Comparing atomistic molecular mechanics force fields for a difficult target: a case study on the Alzheimer's amyloid β-peptide" *J. Biomol. Struct. Dyn.*, vol. 1102, no. February 2014, pp. 37–41, 2013.

[81]   S. Piana, A. G. Donchev, P. Robustelli, and D. E. Shaw, "Water dispersion interactions strongly influence simulated structural properties of disordered protein states" *J. Phys. Chem. B*, vol. 119, no. 16, pp. 5113–5123, 2015.

[82]   Wei Ye, Dingjue Ji, Wei Wang, Ray Luo, and Haifeng Chen, "Test and evaluation of ff99IDPs force field for intrinsically disordered proteins" *J. Chem. Inf. Model.*, p. 150428112228000, 2015.

[83]   S. Rauscher, V. Gapsys, M. J. Gajda, M. Zweckstetter, B. L. de Groot, and H. Grubmüller, "Structural ensembles of intrinsically disordered proteins depend strongly on force field: A comparison to experiment" *J. Chem. Theory Comput.*, p. 151009180807001, 2015.

[84]   Z. A. Levine, L. Larini, N. E. LaPointe, S. C. Feinstein, and J.-E. Shea, "Regulation and aggregation of intrinsically disordered peptides" *Proc. Natl. Acad. Sci.*, vol. 112, no. 9, pp. 2758–2763, 2015.

[85]   A. Sethi, J. Tian, D. M. Vu, and S. Gnanakaran, "Identification of minimally interacting modules in an intrinsically disordered protein" *Biophys. J.*, vol. 103, no. 4, pp. 748–57, 2012.

[86]   D. J. Rosenman, C. R. Connors, Wen Chen, Chunyu Wang, and A. E. García, "Aβ monomers transiently sample oligomer and fibril-like configurations: Ensemble characterization using a combined MD/NMR Approach" *J. Mol. Biol.*, vol. 425, no. 18, pp. 3338–3359, 2013.

[87]   A. H. Mao, S. L. Crick, A. Vitalis, C. L. Chicoine, and R. V. Pappu, "Net charge per residue modulates conformational ensembles of intrinsically disordered proteins" *Proc. Natl. Acad. Sci.*, vol. 107, no. 18, pp. 8183–8188, 2010.

[88]   R. B. Best and J. Mittal, "Free-energy landscape of the GB1 hairpin in all-atom explicit solvent simulations with different force fields: Similarities and differences" *Proteins Struct. Funct. Bioinforma.*, vol. 79, no. 4, pp. 1318–1328, 2011.

[89]   S. Piana, J. L. Klepeis, and D. E. Shaw, "Assessing the accuracy of physical models used in protein-folding simulations: quantitative evidence from long molecular dynamics simulations" *Curr. Opin. Struct. Biol.*, vol. 24, pp. 98–105, 2014.

[90]   J. J. Skinner, W. Yu, E. K. Gichana, M. C. Baxa, J. R. Hinshaw, K. F. Freed, and T. R. Sosnick, "Benchmarking all-atom simulations using hydrogen exchange" *Proc. Natl. Acad. Sci. U. S. A.*, vol. 111, no. 45, pp. 15975–80, 2014.

[91]   K. Lindorff-Larsen, N. Trbovic, P. Maragakis, S. Piana, and D. E. Shaw, "Structure and dynamics of an unfolded protein examined by molecular dynamics simulation" *J. Am. Chem. Soc.*, vol. 134, pp. 3787–3791, 2012.

[92]   R. B. Best, W. Zheng, and J. Mittal, "Balanced protein – water interactions improve properties of disordered proteins and non-speci fi c protein association" *J. Chem. Theory Comput.*, vol. 10, pp. 5113–5124, 2014.

[93]   D. Mercadante, S. Milles, G. Fuertes, D. I. Svergun, E. A. Lemke, and F. Gräter, "Kirkwood-buff approach rescues overcollapse of a disordered protein in canonical protein force fields" *J. Phys. Chem. B*, vol. 119, no. 25, pp. 7975–7984, 2015.

[94]   Xiangyang Zhong, Pingping Wang, J. Han, M. G. Rosenfeld, and Xiangdong Fu, "SR proteins in vertical integration of gene expression from transcription to RNA processing to translation" *Mol. Cell*, vol. 35, no. 1, pp. 1–10, 2009.

[95]   Shengrong Lin and F. Xiangdong Fu, "SR proteins and related factors in alternative splicing" *Adv. Exp. Med. Biol.*, vol. 623, pp. 107–122, 2006.

[96]   F. Alber, S. Dokudovskaya, L. M. Veenhoff, W. Zhang, J. Kipper, D. Devos, A. Suprapto, O. Karni-Schmidt, R. Williams, B. T. Chait, A. Sali, and M. P. Rout, "The molecular architecture of the nuclear pore complex" *Nature*, vol. 450, no. 7170, pp. 695–701, Nov. 2007.

[97]   B. B. Hülsmann, A. A. Labokha, and D. Görlich, "The permeability of reconstituted nuclear pores provides direct evidence for the selective phase model" *Cell*, vol. 150, no. 4, pp. 738–51, Aug. 2012.

[98]   S. Frey and D. Görlich, "A saturated FG-repeat hydrogel can reproduce the permeability properties of nuclear pore complexes" *Cell*, vol. 130, no. 3, pp. 512–23, Aug. 2007.

[99]   J. M. Cronshaw, "Proteomic analysis of the mammalian nuclear pore complex" *J. Cell Biol.*, vol. 158, no. 5, pp. 915–927, Aug. 2002.

[100] S. S. Patel, B. J. Belmont, J. M. Sante, and M. F. Rexach, "Natively unfolded nucleoporins gate protein diffusion across the nuclear pore complex" *Cell*, vol. 128, no. 1, pp. 83–96, Apr. 2007.

[101] I. G. Macara, "Transport into and out of the nucleus" *Microbiol. Mol. Biol. Rev.*, vol. 65, no. 4, pp. 570–594, Dec. 2001.

[102] H. Fried and U. Kutay, "Nucleocytoplasmic transport: taking an inventory" *Cell. Mol. Life Sci.*, vol. 60, no. 8, pp. 1659–1688, Aug. 2003.

[103] G. Dirk and U. Kutay, "transport between the cell nucleus and the cytoplasm" *Annu. Rev. Cell Dev. Biol.*, vol. 15, pp. 607–660, 1999.

[104] F. Stutz, E. Izaurralde, I. W. Mattaj, and M. Rosbash, "A role for nucleoporin FG repeat domains in export of human immunodeficiency virus type 1 Rev protein and RNA from the nucleus" *Mol. Cell. Biol.*, vol. 16, no. 12, pp. 7144–7150, Dec. 1996.

[105] R. Bayliss, T. Littlewood, and M. Stewart, "Structural basis for the interaction between FxFG nucleoporin repeats and importin-β in nuclear trafficking" *Cell*, vol. 102, no. 1, pp. 99–108, Jul. 2000.

[106] R. Bayliss, S. W. Leung, R. P. Baker, B. B. Quimby, A. H. Corbett, and M. Stewart, "Structural basis for the interaction between NTF2 and nucleoporin FxFG repeats" *Eur. Mol. Biol. Organ.*, vol. 21, no. 12, pp. 2843–2853, 2002.

[107] M. P. Rout, J. D. Aitchison, A. Suprapto, K. Hjertaas, Y. Zhao, and B. T. Chait, "The yeast nuclear pore complex : Composition , architecture , and transport mechanism" *J. Cell Biol.*, vol. 148, no. 4, pp. 635–651, 2000.

[108] K. Ribbeck and D. Görlich, "Kinetic analysis of translocation through nuclear pore complexes" *EMBO J.*, vol. 20, no. 6, 2001.

[109] R. Y. H. Lim, Ningping Huang, J. Köser, Jie Deng, K. H. A. Lau, K. Schwarz-Herion, B. Fahrenkrog, and U. Aebi, "Flexible phenylalanine-glycine nucleoporins as entropic barriers to nucleocytoplasmic transport" *Proc. Natl. Acad. Sci.*, vol. 103, no. 25, pp. 9512–9517, 2006.

[110] M.C. Dabauvalle and W. W. Franke, "Karyophilic proteins : Polypeptides synthesized in vitro accumulate in the nucleus on microinjection into the cytoplasm of amphibian oocytes" *Proc. Natl. Acad. Sci.*, vol. 79, no. September, pp. 5302–5306, 1982.

[111] S. Frey, R. P. Richter, D. Görlich, D. G. S. Frey, and R. P. Richter, "FG-rich repeats of nuclear pore proteins form a three dimensional meshwork with hydrogel-like properties" *Science* vol. 314, no. 5800, pp. 815–817, 2006.

[112] S. Frey and D. Görlich, "FG/FxFG as well as GLFG repeats form a selective permeability barrier with self-healing properties" *EMBO J.*, vol. 28, no. 17, pp. 2554–67, Sep. 2009.

[113] B. Fahrenkrog, E. C. Hurt, U. Aebi, and N. Panté, "Molecular architecture of the yeast nuclear pore complex: Localization of Nsp1p subcomplexes" *J. Cell Biol.*, vol. 143, no. 3, pp. 577–588, 1998.

[114] C. Ader, S. Frey, W. Maas, H. B. Schmidt, D. Görlich, and M. Baldus, "Amyloid-like interactions within nucleoporin FG hydrogels" *Proc. Natl. Acad. Sci. U. S. A.*, vol. 107, no. 14, pp. 6281–5, Apr. 2010.

[115] J. R. Lakowicz, "Principles of fluorescence spectroscopy" Book,2006.

[116] L. J. Gosting and L. Onsager, "A general theory for the gouy diffusion method" *J. Am. Chem. Soc.*, vol. 74, pp. 6066–6074, 1952.

[117] G. Fate and D. G. Lynn, "Molecular diffusion coefficients: Experimental determination and demonstration" *J. Chem. Educ.*, vol. 67, no. 6, pp. 536–538, 1990.

[118] M. von Smoluchowski, "Zur kinetischen Theorie der Brownschen Molekularbewegung und der Suspensionen" *Ann. Phys.*, vol. 21, pp. 756–780, 1906.

[119] D. Magde, E. Elson, and W.W.Webb, "Thermodynamic fluctuations in a reacting system-measurement by fluorescence correlation spectroscopy" *Phys. Rev. Lett.*, vol. 29, no. September, pp. 705–708, 1972.

[120] D. E. Koppel, D. Axelrod, J. Schlessinger, E. L. Elson, and W. W. Webb, "dynamics of fluorescence marker concentration as a probe of mobility" vol. 16, pp. 1315–1329, 1976.

[121] M. Eigen and R. Rigler, "Sorting single molecules: application to diagnostics and evolutionary biotechnology" *Proc. Natl. Acad. Sci. U. S. A.*, vol. 91, no. 13, pp. 5740–7, 1994.

[122] R. M. Dickson, A. B. Cubitt, R. Y. Tsien, and W. E. Moerner, "On/off blinking and switching behaviour of single molecules of green fluorescent protein" *Nature*, vol. 388, no. 6640, pp. 355–358, 1997.

[123] O. Krichevsky and G. Bonnet, "Fluorescence correlation spectroscopy: the technique and its applications" *Reports Prog. Phys.*, vol. 65, pp. 251–297, 2002.

[124] A. Einstein, "Investigations on the theory of the brownian movement" Courier Corporation, 1905.

[125] A. Loman, "Molecular Sizing using Fluorescence Correlation Spectroscopy" Ph.D. Thesis, University of Göttingen, 2010.

[126] J. Ries and P. Schwille, "Fluorescence correlation spectroscopy" *Bioessays*, vol. 34, no. 5, pp. 361–8, May 2012.

[127] K. Weiß, "Quantifying the Diffusion of Membrane Proteins and Peptides" Ph.D. Thesis, University of Göttingen, 2013.

[128] T. Dertinger, V. Pacheco, I. von der Hocht, R. Hartmann, I. Gregor, and J. Enderlein, "Two-focus fluorescence correlation spectroscopy: a new tool for accurate and absolute diffusion measurements" *Chemphyschem*, vol. 8, no. 3, pp. 433–43, Feb. 2007.

[129] C. B. Müller, K. Weiß, W. Richtering, A. Loman, and J. Enderlein, "Calibrating differential interference contrast microscopy with dual-focus fluorescence correlation spectroscopy" *Opt. Express*, vol. 16, no. 6, pp. 4322–4329, 2008.

[130] C. M. Pieper, "Diffusion and flow on microscopic length scales studied with fluorescence correlation spectroscopy" Ph.D. Thesis, University of Göttingen, 2012.

[131] W. Becker, "Advanced time-correlated single photon counting techniques" Book, *Springer Science & Business Media*, Vol. 81, 2005.

[132] J. Enderlein and I. Gregor, "Art and artefacts of fluorescence correlation spectroscopy" *Curr. Pharm. Biotechnol.*, vol. 5, pp. 155–161, 2004.

[133] H. Barsch, "Investigation of biopolymer properties with combined fluorescence correlation spectroscopic and computational methods" Ph.D. Thesis, Bielefeld University, 2008.

[134] B. Schuler, E. A. Lipman, P. J. Steinbach, M. Kumke, and W. A. Eaton, "Polyproline and the 'spectroscopic ruler' revisited with single-molecule fluorescence" *Proc. Natl. Acad. Sci. U. S. A.*, vol. 102, no. 8, pp. 2754–2759, 2005.

[135] D. Rehm and A. Weller, "Kinetics of fluorescence quenching by electron and H-atom transfer" *Isr. J. Chem.*, vol. 2, no. 8, pp. 259–271., 1970.

[136] M. R. DeFelippis, C. P. Murthy, M. Faraggi, and M. H. Klapper, "Pulse radiolytic measurement of redox potentials: the tyrosine and tryptophan radicals" *Biochemistry*, vol. 28, pp. 4847–4853, 1989.

[137] J. Vogelsang, T. Cordes, C. Forthmann, C. Steinhauer, and P. Tinnefeld, "Controlling the fluorescence of ordinary oxazine dyes for single-molecule switching and superresolution microscopy" *Proc. Natl. Acad. Sci. U. S. A.*, vol. 106, no. 20, pp. 8107–12, 2009.

[138] S. Doose, H. Neuweiler, and M. Sauer, "A close look at fluorescence quenching of organic dyes by tryptophan" *Chemphyschem*, vol. 6, no. 11, pp. 2277–85, Nov. 2005.

[139] A. C. Vaiana, H. Neuweiler, A. Schulz, J. Wolfrum, M. Sauer, and J. C. Smith, "Fluorescence quenching of dyes by tryptophan: interactions at atomic detail from combination of experiment and computer simulation" *J. Am. Chem. Soc.*, vol. 125, no. 47, pp. 14564–72, Nov. 2003.

[140] H. Neuweiler, A. Schulz, M. Böhmer, J. Enderlein, and M. Sauer "Measurement of submicrosecond intramolecular contact formation in peptides at the single-molecule level" *J. Am. Chem. Soc.*, vol. 125, no. 18, pp. 5324–5330, 2003.

[141] O. Bieri, J. Wirz, B. Hellrung, M. Schutkowski, M. Drewello, and T. Kiefhaber, "The speed limit for protein folding measured by triplet-triplet energy transfer" *Proc. Natl. Acad. Sci. U. S. A.*, vol. 96, no. 17, pp. 9597–9601, 1999.

[142] T. H. Evers, E. M. W. M. van Dongen, A. C. Faesen, E. W. Meijer, and M. Merkx, "Quantitative understanding of the energy transfer between fluorescent proteins connected via flexible peptide linkers" *Biochemistry*, vol. 45, no. 44, pp. 13183–13192,

2006.

[143] H. Neuweiler, M. Löllmann, S. Doose, and M. Sauer, "Dynamics of unfolded polypeptide chains in crowded environment studied by fluorescence correlation spectroscopy" *J. Mol. Biol.*, vol. 365, no. 3, pp. 856–69, Jan. 2007.

[144] R. E. Manrow, A. R. Sburlati, J. A. Hanover, and S. L. Berger, "Nuclear targeting of prothymosin a" *J Biol Chem*, vol. 266, no. 6, pp. 3916–3924, 1991.

[145] R. N. Karapetian, A. G. Evstafieva, I. S. Abaeva, N. V Chichkova, G. S. Filonov, Y. P. Rubtsov, E. A. Sukhacheva, S. V. Melnikov, U. Schneider, E. E. Wanker, and A. B. Vartapetian, "Nuclear oncoprotein prothymosin alpha is a partner of Keap1: implications for expression of oxidative stress-protecting genes" *Mol. Cell. Biol.*, vol. 25, no. 3, pp. 1089–99, 2005.

[146] A. G. Evstafieva, G. A. Belov, Y. P. Rubtsov, M. Kalkum, B. Josephd, N. V Chichkovaa, E. A Sukhachevae, A.A Bogdanova, R. F Petterssond, V. I Agola, and A. B Vartapetiana "Apoptosis-related fragmentation, translocation, and properties of human prothymosin alpha" *Exp. Cell Res.*, vol. 284, no. 2, pp. 209–221, 2003.

[147] J. Gómez-Márquez, F. Segade,M. Dosil, J. G. Pichel, X. R. Bustelo, and M. Freire, "The Exmession of prothsmosin alpha gene in T lymphocytes and leukemic lymphoid cells is tied to lymphocyte proliferation" *Communication*, vol. 264, no. 15, pp. 8451–8454, 1989.

[148] Xuejun Jiang, Hyun-Eui Kim, Hongjun Shu, Yingming Zhao, Haichao Zhang, J. Kofron, J. Donnelly, D. Burns, Shi-Chung Ng, S. Rosenberg, and Xiaodong Wang "Distinctive roles of PHAP proteins and prothymosin-α in a death regulatory pathway" *Science*, vol. 299, no. 5604, pp. 223–226, 2003.

[149] Z. Karetsou, A. Kretsovali, C. Murphy, O. Tsolas, and T. Papamarcaki, "Prothymosin alpha interacts with the CREB-binding protein and potentiates transcription" *EMBO Rep.*, vol. 3, no. 4, pp. 361–6, 2002.

[150] P. G.V. Martini, R. Delage-Mourroux, D. M. Kraichely, and B. S. Katzenellenbogen, "'Prothymosin alpha selectively enhances estrogen receptor transcriptional activity by interacting with a repressor of estrogen receptor activity.,'" *Mol. Cell. Biol.*, vol. 20, no. 17, pp. 6224–6232, 2000.

[151] K. Gast, H. Damaschun, K. Eckert, K. Schulze-Forster, H. R. Maurer, M. Müller-Frohne, D. Zirwer, J. Czarnecki, and G. Damaschun, "Prothymosin alpha: a biologically active protein with random coil conformation" *Biochemistry*, vol. 34, pp. 13211–13218, 1995.

[152] B. Padmanabhan, Y. Nakamura, and S. Yokoyama, "Structural analysis of the complex of Keap1 with a prothymosin alpha peptide" *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.*, vol. 64, no. 4, pp. 233–238, 2008.

[153] F. Esposito and E. Tramontano, "Past and future. Current drugs targeting HIV-1 integrase and reverse transcriptase-associated ribonuclease H activity: single and dual active site inhibitors" *Antivir. Chem. Chemother.*, vol. 23, no. 4, pp. 129–144, 2013.

[154] Mengli Cai, Ronglan Zheng, M. Caffrey, R. Craigie, G. M. Clore, and A. M. Gronenborn, "Solution structure of the N-terminal zinc binding domain of HIV-1 integrase" *Nature structural biology,* vol. 4, no. 7, pp. 567-577,1997.

[155] Ronglan Zheng, T. M. Jenkins, and R. Craigie, "Zinc folds the N-terminal domain of HIV-1 integrase, promotes multimerization, and enhances catalytic activity" *Proc. Natl. Acad. Sci. U. S. A.*, vol. 93, no. 24, pp. 13659–64, 1996.

[156] S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, B. Hess, and E. Lindahl, "GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit" *Bioinformatics*, vol. 29, no. 7, pp. 845–854, 2013.

[157] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl, "GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation" *J. Chem. Theory Comput.*, vol. 4, no. 3, pp. 435–447, 2008.

[158] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, "A smooth particle mesh Ewald method" *J Chem Phys*, vol. 103, no. 1995, pp. 8577–8593, 1995.

[159] G. Bussi, D. Donadio, and M. Parrinello, "Canonical sampling through velocity rescaling" *J. Chem. Phys.*, vol. 126, no. 1, pp. 1–7, 2007.

[160] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, "Molecular dynamics with coupling to an external bath" *J. Chem. Phys.*, vol. 81, no. 8, pp. 3684–3690, 1984.

[161] M. Parrinello and A. Rahman, "Polymorphic transitions in single crystals: A new molecular dynamics method" *J. Appl. Phys.*, vol. 52, no. 12, p. 7182, 1981.

[162] Y. M. Rhee, and V. S. Pande "Multiplexed-replica exchange molecular dynamics method for protein folding simulation" *Biophys. J.,*vol.84, no. 2, pp. 775-786, 2003.

[163] W. Humphrey, A. Dalke, and K. Schulten, "VMD: Visual molecular dynamics" *J. Mol. Graph.*, vol. 14, no. 1, pp. 33–38, 1996.

[164] H. Flyvbjerg and H. G. Petersen, "Error estimates on averages of correlated data" *J. Chem. Phys.*, vol. 91, no. 1, p. 461-466, 1989.

[165] L. J. Smith, K. M. Fiebig, H. Schwalbe, and C. M. Dobson, "The concept of a random coil. Residual structure in peptides and denatured proteins" *Fold. Des.*, vol. 1, no. 5, pp. R95–R106, 1996.

[166] C. M. Dobson, "Unfolded proteins, compact states and molten globules" *Curr. Opin. Struct. Biol.*, vol. 2, no. 1, pp. 6–12, 1992.

[167] K. A Dill, S. Bromberg, Kaizhi Yue, Hue Sun Chan,K. M. Ftebig, D. P. Yee, and P. D. ThomasK "Principles of protein folding-a perspective from simple exact models" *Protein Sci.*, vol. 4, no. 4, pp. 561–602, 1995.

[168] D. P.Yee, H. S. Chan, T. F. Havel, and K. A. Dill, "Does compactness induce secondary structure in proteins? A study of poly-alanine chains computed by distance geometry" *J. Mol. Biol.*, vol. 241, pp. 557–573, 1994.

[169] S. Piana, K. Lindorff-Larsen, and D. E. Shaw, "How Robust Are Protein Folding Simulations with Respect to Force Field Parameterization?" *Biophys. J.*, vol. 100, no. 9, pp. L47–L49, 2011.

[170] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, "Comparison of simple potential functions for simulating liquid water" *J. Chem. Phys.*, vol. 79, no. 2, p. 926, 1983.

[171] R. B. Best and J. Mittal, "Protein simulations with an optimized water model: Cooperative helix formation and temperature-induced unfolded state collapse" *J. Phys. Chem. B*, vol. 114, no. 46, pp. 14916–14923, 2010.

[172] J. L. F. Abascal and C. Vega, "A general purpose model for the condensed phases of water: TIP4P/2005" *J. Chem. Phys.*, vol. 123, no. 23, p. 234505, 2005.

[173] A. D. Mackerell, Jr.D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, III, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, and M. Karplus, "All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins" *J Phys Chem B*, vol. 102, no. 97, pp. 3586–3616, 1998.

[174] R. B. Best, Xiao Zhu, J. Shim, P. E. M. Lopes, J. Mittal, M. Feig, and A. D. MacKerell, "Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ, ψ and side-chain χ 1 and χ 2 dihedral angles" *J. Chem. Theory Comput.*, vol. 8, no. 9, pp. 3257–3273, 2012.

[175] G. A. Kaminski, R. A. Friesner, J. Tirado-Rives, and W. L. Jorgensen, "Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides" *J. Phys. Chem. B*, vol. 105, no. 28, pp. 6474–6487, 2001.

[176] H. Neuweiler, S. Doose, and M. Sauer, "A microscopic view of miniprotein folding: enhanced folding efficiency through formation of an intermediate" *Proc. Natl. Acad. Sci. U. S. A.*, vol. 102, no. 46, pp. 16650–5, Nov. 2005.

[177] S. Doose, H. Neuweiler, H. Barsch, and M. Sauer, "Probing polyproline structure and dynamics by photoinduced electron transfer provides evidence for deviations from a regular polyproline type II helix" *Proc. Natl. Acad. Sci. U. S. A.*, vol. 104, no. 44, pp. 17400–5, Oct. 2007.

[178] H. Neuweiler, C. M. Johnson, and A. R. Fersht, "Direct observation of ultrafast folding and denatured state dynamics in single protein molecules" *Proc. Natl. Acad. Sci. U. S. A.*, vol. 106, no. 44, pp. 18569–74, Nov. 2009.

[179] C. Carnero-Ruiz, J. Molina-Bolívar, J. Aguiar, G. MacIsaac, S. Moroze, and R. Palepu,

"Effect of ethylene glycol on the thermodynamic and micellar properties of Tween 20" *Colloid Polym. Sci.*, vol. 281, no. 6, pp. 531–541, 2003.

[180] F. Krieger, B. Fierz, O. Bieri, M. Drewello, and T. Kiefhaber, "Dynamics of unfolded polypeptide chains as model for the earliest steps in protein folding" *J. Mol. Biol.*, vol. 332, no. 1, pp. 265–274, Sep. 2003.

[181] S. M. Bailer, C. Balduf, and E. Hurt, "The Nsp1p carboxy-terminal domain is organized into functionally distinct coiled-coil regions required for assembly of nucleoporin subcomplexes and nucleocytoplasmic transport" *Mol. Cell. Biol.*, vol. 21, no. 23, pp. 7944–7955, 2001.

[182] D. K. Klimov and D. Thirumalai, "Viscosity dependence of the folding rates of proteins" *Phys. Rev. Lett.*, vol. 79, no. 2, p. 14, 1997.

[183] P. Hänggi, P. Talkner, and M. Borkovec, "Reaction-rate theory: fifty years after Kramers" *Rev. Mod. Phys.*, vol. 62, no. 2, pp. 251–341, 1990.

[184] H. A. Kramers, "Brownian motion in a field of force and the diffusion model of chemical reactions" *Phys. (The Hague)*, vol. 7, pp. 284–304, 1940.

[185] A. Ansari, C. M. Jones, E. R. Henry, J. Hofrichter, and W. A. Eaton, "The role of solvent viscosity in the dynamics of protein conformational changes" *Science*, vol. 256, no. 5065, pp. 1796–1798, 1992.

[186] A. Ruiz-Llamas and R. Macías-Salinas, "Modeling the Dynamic Viscosity of Ionic Solutions" *Ind. Eng. Chem. Res.*, vol. 54, no. 28, pp. 7169–7179, 2015.

[187] D. P. Denning, S. S. Patel, V. Uversky, A. L. Fink, and M. Rexach, "Disorder in the nuclear pore complex: the FG repeat regions of nucleoporins are natively unfolded" *Proc. Natl. Acad. Sci. U. S. A.*, vol. 100, no. 5, pp. 2450–5, Mar. 2003.

[188] M. Petri, S. Frey, A. Menzel, D. Görlich, and S. Techert, "Structural characterization of nanoscale meshworks within a nucleoporin FG hydrogel" *Biomacromolecules*, vol. 13, no. 6, pp. 1882–9, Jun. 2012.

Acronyms

IDPs: Intrinsically disordered proteins

NPCs: Nuclear pore complexes

Nups: Nucleoporins

ProTα: Prothymosin alpha

Atto655md: Atto655 Maleimide

BSA: Bovine serum albumin

Tween-20: Polyoxyethylene (20) sorbitan monolaurate

GdnHCl: Guanidine hydrochloride

PBS: Phosphate-buffered saline

EDTA: Ethylenediaminetetraacetic acid

TCEP: Tris(2-chloroethyl) phosphate

FCS: Fluorescence Correlation Spectroscopy

2fFCS: Dual-focus Fluorescence Correlation Spectroscopy

PET-FCS: Photo-induced electron transfer fluorescence correlation spectroscopy

TCSPC: Time Correlated Single Photon Counting

FRET: Fluorescence Resonance Energy Transfer

MD simulation: Molecular dynamics simulation

Acknowledgements

This thesis would not have been possible without the help and support of several people who constantly encouraged me to keep proceeding.

First of all, I would like to thank my supervisor, Jörg Enderlein, for the opportunity to work with you. Thank you for engaging me in new ideas, giving me intellectual freedom in my project, and supporting my attendance at various conferences. I do appreciate that you are not only a teacher but also a friend to me during these four years.

I would like to thank my thesis committee members, Prof. Holger Stark and Prof. Heinz Neumann. Thank you for discussing the progress of my PhD work. I do appreciate your encouragement and advice.

Qui Van, I am grateful for your help. You are the first person who helped me adapt to a foreign country, helped me manage the setups in the lab, helped me with my project, and even helped me practice my English. You are like an older brother to me.

Ingo Gregor, thank you for helping me adjust the setups, discussing about all questions and writing scripts for my data analysis.

Stefanie, thank you for your contagious smile. "Morgen" and "Feierabend und bis morgen" that put me in a good mood every day.

Thank Alexey, Anna, and your little son: Maxim. His pure laugh brought me much happiness.

I would like to thank all the group members: Hao, Weixing, Melanie, Sebastian, Dirk, Anja, Daja, Simon, Jan, Narain and Sid.

I want to thank Elke Zech, Nicole Rehbein and Ursula Hahn-Wörgötter for their help with thousands of papers for the Rathaus. I also want to thank Thomas Geiling for the IT support.

I would like to thank Antje Erdmann and the coordinators of GGNB program and the financial support from the SFB 755 (DFG).

Last but not least, I am grateful to my family for the support they provided me through my entire life and in particular, I must acknowledge my wife and best friend, Ting Wang, without your love and encouragement, I would not have finished this thesis.

Curriculum Vitae

Man Zhou

Date of birth: 14. May 1986

Place of birth: Hubei, China

EDUCATION

2012-2016	Ph.D. Georg August University, Göttingen, Germany Third Institute of Physics - Biophysics Prof. Dr. Jörg Enderlein
2008-2011	Master Wuhan Institute of Virology, Chinese Academy of Sciences Major: Biochemistry and Molecular biology
2004-2008	Bachelor Wuhan University, China Major: Life Science and Biotechnology

PAPERS & PUBLICATIONS

Man Zhou, *et al.* "Rapid colorimetric testing for pyrazinamide susceptibility of M. tuberculosis by a PCR-based in-vitro synthesized pyrazinamidase method." *PloS one* 6.11 (2011): e27654.

Li Huang, Xiumei Hu, **Man Zhou** *et al.* "Rapid detection of new delhi metallo-β-lactamase gene and variants coding for carbapenemases with different activities by use of a PCR-based in vitro protein expression method." *Journal of clinical microbiology,* June 2014 vol. 52 no. 6 1947-1953

Heng Li, Jun Chen, **Man Zhou** *et al.* "Rapid Detection of Mycobacterium tuberculosis and Pyrazinamide Susceptibility Related to pncA Mutations in Sputum Specimens through an Integrated Gene-to-Protein Function Approach." *Journal of clinical microbiology*, January 2014 vol. 52 no. 1260-267

Joerg Enderlein, **Man Zhou,** *et al*. "Study conformational dynamics of intrinsically disordered protein by PET-FCS" (Conference Presentation). Proc. SPIE 9714, Single Molecule Spectroscopy and Superresolution Imaging IX, 971404 (April 27, 2016); doi:10.1117/12.2212247.

PATENTS

Publication No. WO2012116578 A1

Method for detecting pyrazinamide resistance of mycobacterium tuberculosis . HP Wei, **M Zhou**, XE Zhang, ZP Zhang, DB Wang. Wuhan Institute of Virology, Chinese Academy of Sciences.

Publication No. CN102220433 B

Method for rapid detection of drug resistant gene New Delhi metallo-beta-lactamase in bacteria. HP Wei, **M Zhou**, JF Zhao, ZP Zhang, XE Zhang: Wuhan Institute of Virology, Chinese Academy of Sciences.

Publication No. CN102925554 A

Joint detection method for drug resistance of mycobacterium tuberculosis and pyrazinamide in clinical sample. HP Wei, H Li, **M Zhou**, XL Geng, DB Wang, JP Yu, XE Zhang: Wuhan Institute of Virology, Chinese Academy of Sciences.

AWARDS & ACHIEVEMENTS

Outstanding graduate of Chinese Academy of Sciences (Jun.2011)

Di'ao scholarship for outstanding master student (May. 2011)

Triple-A student of Chinese Academy of Sciences (2010-2011)

Bachelor's Scholarship Prize (Dec.2004)

CONFERENCES AND WORKSHOPS

The 18th International Workshop on Single Molecule Spectroscopy, Berlin, September 2012

The 1st International Symposium on Nanoscale Photonic Imaging, Göttingen, April 2013

The 9th EBSA European Biophysics Congress, Lisbon, Portugal, July 2013

The 20th International Workshop on Single Molecule Spectroscopy and Ultrasensitive Analysis in the Life Sciences, Berlin, September 2014

Annual conference of Chinese-German Association of Life Science and Biotechnology, Essen, September 2014

SFB 755 Winter School 2015, Nanoscale Photonic Imaging, Teistungen, Germany, January 2015 (talk)

SPIE conference: Single Molecule Spectroscopy and Superresolution Imaging IX, San Francisco, February 13th-18th 2016 (Talk: Study conformational dynamics of intrinsically disordered protein by PET-FCS)