DYNAMIC STRUCTURE

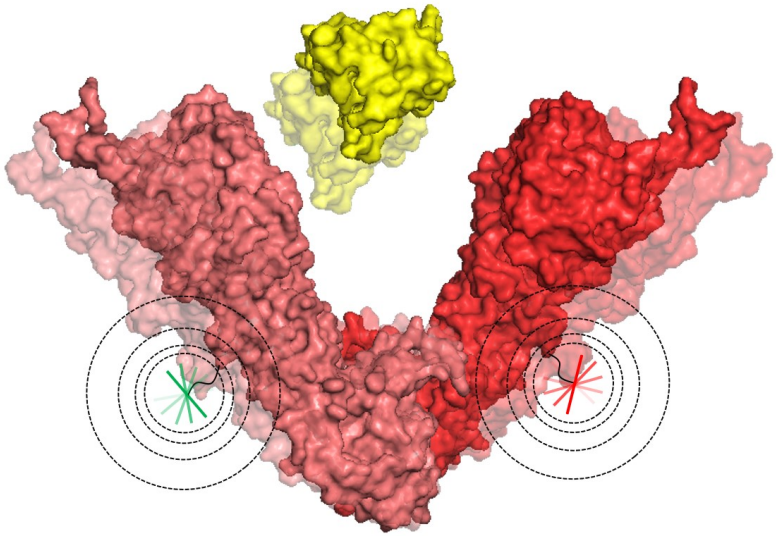
OF A MULTI-DOMAIN PROTEIN

UNCOVERED USING SELF-CONSISTENT FRET NETWORKS AND TIME-CORRELATED DISTANCE DISTRIBUTIONS

BJÖRN HELLENKAMP

Dissertation at the Physics Department E22

Technical University of Munich



Technische Universität München

Physik Department

Lehrstuhl für Biophysik

DYNAMIC STRUCTURE OF A MULTI-DOMAIN PROTEIN

UNCOVERED USING SELF-CONSISTENT FRET NETWORKS AND TIME-CORRELATED DISTANCE DISTRIBUTIONS

BJÖRN HELLENKAMP

Vollständiger Abdruck der von der Fakultät für Physik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzende(r):	Prof. Dr. M. Rief
Prüfer der Dissertation:	1. Prof. Dr. Th. Hugel
	2. Prof. Dr. M. Zacharias
	3. Prof. B. Schuler, Ph.D. (nur schriftliche Beurteilung)

Die Dissertation wurde am 01.02.2016 bei der Technischen Universität München eingereicht und durch die Fakultät für Physik am 15.02.2016 angenommen.

ACKNOWLEDGEMENTS

I would like to thank my supervisor Thorsten Hugel, as well as, Sonja Schmid, Markus Götz, Philipp Wortmann, Markus Jahn, Sara Tyrell, Heinrich Grabmayr, Felix Keber, Alexander Mehlich, Gabriel Zoldak, Johannes Stigler, Ziad Ganim, Martin Zacharias and coworkers, Matthias Rief and coworkers, Claus Seidel and coworkers and Jens Michaelis and coworkers for fruitful scientific discussions.

Many thanks go to Philipp Wortmann, Markus Götz, Markus Jahn and David Agard for biochemical support, among others, with the model client protein Delta131Delta and Aha1.

Resounding thanks go to the overall biophysics department at the TU Munich for a lively and stimulating scientific atmosphere.

Special thanks go to my mother, Sara, Sonja, Micha, Fabi, Jan, Timo, Michael and others for frequently getting me down from the ivory tower of science.

A well-answered question opens at least two new ones.

In order to progress, it needs an optimist, a critic, a theorist and a pragmatist,
whereat the theorist should be the optimist and the pragmatist should be the critic.
Or vice versa?

DYNAMIC STRUCTURE OF A MULTI-DOMAIN PROTEIN

Proteins in their native environment exist as a dynamic ensemble of diverse conformations. An important focus of this work was on the development of a novel approach that provides dynamic structural ensembles of multi-domain proteins. It is based on a global analysis of time-correlated distance distributions obtained via single-molecule Förster resonance energy transfer (FRET). Applying this hybrid approach to the heat shock protein 90 (Hsp90) enabled disentangling nested local and global conformational changes and revealed new mechanistic insights. A novel microfluidic device and an innovative optical setup are further technological achievements that were designed for detecting dynamic multi-protein complexes.

A confocal-based laser optical setup was built and established to determine time-resolved FRET efficiencies and anisotropies of site-specifically labeled single proteins in solution. For the determination of molecular distances, different FRET parameters and detection modes were compared with each other with respect to their theoretical and practical accuracy, and, different FRET dyes were spectroscopically characterized in dependence on their molecular environment. On this basis, selection criteria and photophysical correction parameter were scrutinized and a new parameter was introduced accounting for intersystem crossing. Furthermore, different FRET-based structure determination approaches were compared with each other demonstrating that homogeneously distributed dye positions and isotropic dipole averaging are reasonable assumptions when applying a threshold to the combined residual dye anisotropies.

The general applicability of the derived selection criteria was tested by an unprecedentedly large data set based on more than 100 intra-molecular FRET dye pairs spanning the entire protein. From the distance networks, a solution structure of Hsp90's closed conformation was determined and resembled the x-ray structure with an RMSD of 2.8 Å. The previously unknown average open structure was resolved with an RMSD of 3.8 Å. A third salt-dependent semi-closed conformation was found that probably resembles the structure of endoplasmic reticulum Hsp90. Further confidence in FRET-based distances was achieved through the exploration of distance redundancy and self-consistent networks of fluorescence parameters.

The state-specific distance networks were then combined with a global analysis of time-correlated distance distributions and x-ray structure information. Distance fluctuations were deconvolved from the FRET data and analyzed at the micro- and milliseconds timescale. Thanks to this highly applicable approach, a dynamic ensemble of open Hsp90 structures could be resolved. In addition, inter-domain dynamics and small-scale conformational changes could be determined and related to the global protein states. Such structural relations can assist in uncovering previously unknown regulation sites.

The methodological advancements resulted in new insights into structural dynamics of Hsp90. The open state of was found to be highly flexible featuring dynamics on a broad range of timescales. The interface between C- and M-domain seems to be locally flexible but globally stable caused in whole or in part by interchanging salt-bridges that are not present in the closed state. The timescales and amplitudes of the large-scale inter-monomer fluctuations were integrated into Monte Carlo simulations to derive a most likely mechanistic model. Preliminary temperature-dependent measurements of the inter-monomer fluctuations might hint a cooperation of successive microstates. Moreover, molecular interactions with co-chaperones and nucleotides were measured in order to improve the understanding of the coupling between co-chaperone binding, nucleotide binding, ATP hydrolysis and conformational changes.

Finally, the dynamic interaction between Hsp90 and a partially folded client protein was analyzed. In presence of the model client, the fluctuations within Hsp90's open state were significantly reduced on the (sub-) millisecond timescale. However, the populations of Hsp90's global states remained unchanged indicating an orthogonal regulation mechanism. The observed conformational flexibility and fast client interaction suggest that Hsp90 has its role in facilitating fast conformational adaption for several different client proteins.

In summary, novel technologies enabled a more detailed mechanistic understanding of the Hsp90 machinery. It is anticipated that the presented methods and insights prepare the ground for the detection and understanding of many more dynamic multi-domain proteins and dynamic multi-protein interactions.

CONTENTS

1 INTRODUCTION

The variety of the world is based on the interplay between processes that take place at different space- and timescales. Elementary particles form atoms and molecules, which in turn build up more complex molecules as, for example, proteins. The latter serve as building blocks for cells, whose dynamic interplay represents the basis of organs, life and finally society at the top of the hierarchy. Lower and higher levels mutually interact, similarly to the strict hierarchy of fractals (Figure 1). In contrast to the simple world of fractals, though, - where an identical law applies to each level of reproduction - in real life, novel and unpredictable effects can emerge as we move from the lowest to the highest level. As structural complexity at a given level can be caused by very specific driving forces, individual descriptions and experiments, as well as interdisciplinary approaches are necessary to understand them.

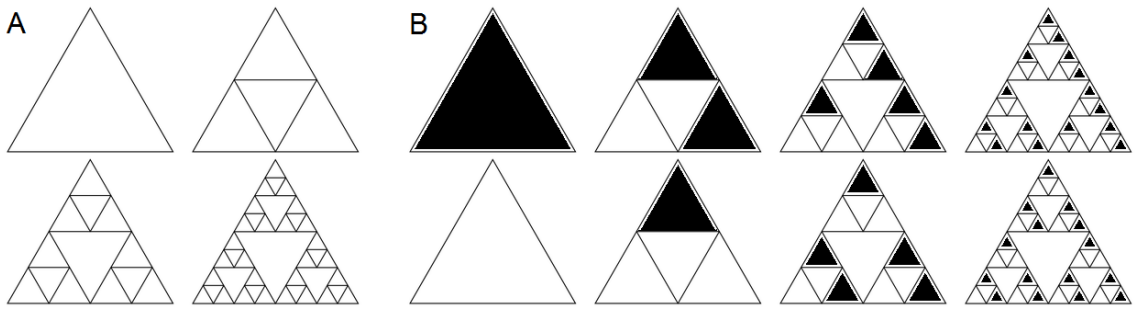


Figure 1: A: Four hierarchical ranks of the fractal Sierpinski triangle. Zooming-in at higher levels resembles the patterns at lower levels. B: Additional randomness at lower levels enormously increases variety at higher levels. Nevertheless, the entire ensemble of higher level patterns is predictable from lover levels.

In the fascinating world of proteins, complexity refers to the temporal and spatial development of structural conformations. A polypeptide with only 100 amino acids enables more possible protein combinations than the number of estimated atoms in the world, just to mention an example of

increasing complexity on higher levels. Such arbitrary combination provides one of several cell's basic toolboxes. Each of about 20.000 protein encoding-genes in the human cell can produce numerous forms of a protein thanks to nested mechanisms, such as alternative splicing or post-translational modifications (*1, 2*). However, complexity is not only given by sequence and structure. The motions within single proteins enable interconversion between conformational ensembles of structures with small- and large-scale differences at equilibrium. Conformational changes occur on a huge range of timescales, from nanoseconds to seconds, affecting single residues, small elements or complete domains. Depending on their structural and dynamical patterns, proteins can interact with each other, thus creating new protein complexes for a certain lifetime. For this reason, nearly arbitrary complex signaling pathways can emerge within a cell. Further complexity arises from novel proteins and novel higher level pathways that can emerge at any time due to lower level random errors during gene replication, transcription or translation.

In principle, each protein has a specific function defined by structural and dynamic patterns. Its tasks within the cell include, *inter alia*, cell signaling, transport, enzymatic function, regulation, mechanical adaption and molecular chaperoning. Fascinating molecular motors have been evolved, molecular multi-component machines that work far away from thermal equilibrium. Some of them have been uncovered in the last decade by using modern single-molecule techniques. Some molecular motors transport, for example, protons through the cell membrane, thereby connecting the atomistic level with higher levels.

Predicting protein structure and function from the peptide sequence is challenging, even for small and rigid structural motifs, because of the huge variety of structural parameters (*3, 4*). Even if, one day, all protein structures are identified, the mostly unknown higher level architectures will remain challenging. The next challenge is the identification and analysis of protein-protein interaction networks, whereat at this level diversity is given by conformational states, conformational disorder, complexed state before and after interaction etc.

The heat shock protein 90 is a perfect example for a high degree of complexity of possible multiple interaction partners. The highly abundant and ubiquitous chaperone interacts with several co-chaperones while managing the interaction with hundreds of client proteins (*5-8*). It is essential for the activation of a large set of signal transduction and key regulatory proteins (*9-11*) – also called client proteins – and has been assigned critical functions in chaperoning oncogenic proteins (*12-14*) and evolution (*15*). Clients are supposed to bind comprehensively to several Hsp90 domains, such as the tau protein binding to the N- and M-domain (*16*), staphylococcal nuclease binding predominantly to the C- and M-domain (*17*) or the tumor suppressor protein p53 binding dynamically to all three domains (*18, 19*). Thus, inter-domain dynamics of Hsp90 seem to play an important role for client binding. However, the understanding of the nature of molecular interactions between chaperone, co-chaperone and client is still at the beginning. The interdependency of the co-chaperone network and client binding remains questioning. Beside

direct folding assistance, Hsp90 is supposed to locally concentrate co-chaperones to a complex that can act on the client protein.

Protein-protein interaction networks help to model local and global interactions between chaperones, co-chaperones and clients. The network in Figure 2 (left) is based on experimental detection of binary high-affinity interactions using analytical methods such as two-hybrid screening or affinity purification (*20*). However, such assays are mostly limited to stable and binary interactions and lack information on important interaction characteristics such as timescales, conformational dependency, complexed state dependency and specificity. Detailed measurements are required to characterize individual conformation-dependent interactions and their interdependency with the global pathways and to finally understand the nested hierarchy between local and global protein interactions and the global pathways organized by the cell.

A better understanding of the transient interactions of the Hsp90 machinery necessitates the detection of intra-molecular inter-domain dynamics. In addition, the overall flexibility of Hsp90's open state has challenged several structure determination methods.

This work focuses on providing a dynamic structural picture of eukaryotic Hsp90. It develops a novel method for the determination of dynamic structural ensembles based on single-molecule Förster resonance energy transfer (FRET) (Figure 2 right). This is a key approach for gaining a comprehensive picture of small- and large-scale conformations and fluctuations at the micro- to milleseconds timescale. It enables the analysis of fast inter-domain dynamics in interdependence with the global state, and thus bridges the gap to determining transient protein-protein interactions.

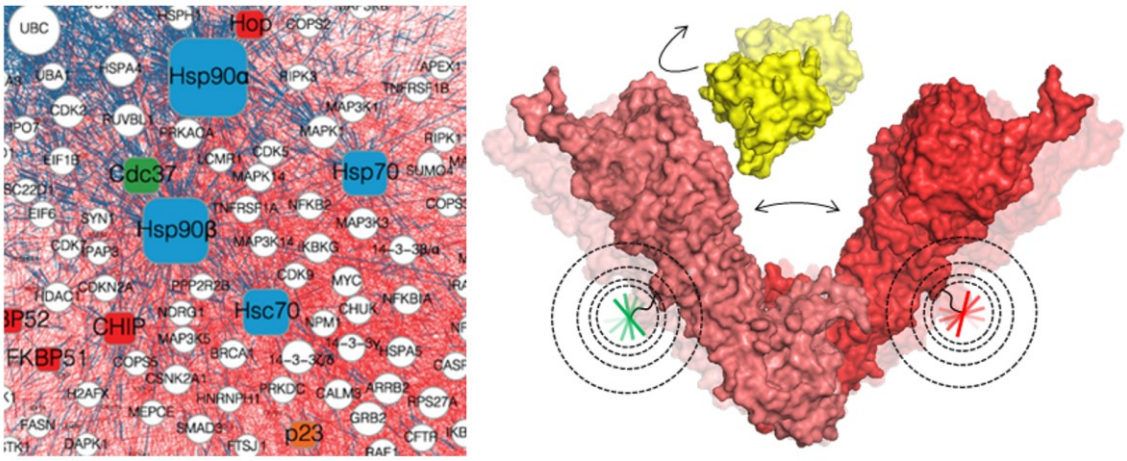


Figure 2: Left: Protein-protein interaction network between chaperones (blue), co-chaperones (red) and other interactors (white) such as client proteins. The blue and red lines indicate interologs and human interactions, respectively (from (*20*)). Right: Schematics of measuring the molecular interaction between the highly dynamic multi-domain protein Hsp90 (reddish surface) and the client protein staphylococcal nuclease (yellow surface) indicating a time window of 1 ms. The time-resolved distance distribution between donor

and acceptor dye labels – represented as rotating dipoles (green and red, respectively) at a timescale of 1 ns – is calculated via the energy transfer efficiency.

The hybrid method enables to answer further open questions. The coupling between Hsp90's conformational changes, nucleotide binding and ATP hydrolysis remains fragmentary and puzzling even in the absence of interaction partners. Similarly, the nucleotide dependent symmetry of the dimer is not fully understood. It is still unclear how clients interact with Hsp90 and how their interaction is synchronized with nucleotide and co-chaperone binding. The release mechanism is probably even more important. Several mechanisms have been reported covering a remarkable timespan from microseconds for ATP-lid movement (*21*), over milliseconds for the dynamics of the negatively charged linker between N- and M-domains (*22*), over seconds for large conformational rearrangements (*23*), up to minutes for the ATPase rate. These and other flexible elements might play a role in client, co-chaperone or self-regulation. Further stand-alone observations have been made in the last years, such as communications between C- and N-domain, dimer asymmetry, single turnover ATPase rates, as well as, interactions with clients, co-chaperones and nucleotides.

A comprehensive picture for the description of the Hsp90 machinery is still missing. The interplay of Hsp90's conformational states and ATP resulted already in complex rate networks (*23*). In order to bring the above observations into a global context and to be able to describe higher order interactions, a simpler basic conformational model becomes necessary. The presented dynamic approach sheds light on these issues. Above all, it enables the observation of interaction with a client protein at the sub-millisecond timescale. Furthermore, systematically varying physicochemical parameters allows interesting speculations about the origin of the fluctuations. All things considered, a mechanism for transient client interaction is proposed.

Determination methods of structure and dynamics

For a long time, researchers have been benefitting from powerful methods such as x-ray crystallography that tremendously increased the understanding in molecular biology by providing three-dimensional models of proteins at atomic resolution. These models represent rigid and sometimes non-native protein conformations. Theoretical models use them to predict binding free energies of protein–protein complexes (*24*). However, transient and diffusive interactions and structural intermediates are frequently overseen, because they are difficult to detect, difficult to classify and thus difficult to integrate in a comprehensive picture. Transient interactions may serve as a previously underestimated mode of protein interaction and regulation. It seems likely that the numerous interactions within the Hsp90 machinery occur on very different timescales and different hierarchical sublevels, which are linked to the inter-domain dynamics of Hsp90. Fast inter-domain dynamics and fast fluctuations of small loops located at the inner surface of Hsp90

may form the basis for dynamic interactions. An important question one may ask in this context is whether binding specificity is in contradiction to transient interaction. The first step for understanding these interactions is changing the focus from a static to a dynamic protein picture.

Before understanding transient interactions, one has to detect and understand related intra-molecular inter-domain dynamics. Nuclear magnetic resonance (NMR) spectroscopy of proteins in solution opened a door for the understanding of local dynamics, but is limited to small distances. Recent advances in NMR relaxation methods allow measuring the conformational entropy of the protein in its free and complexed state (*25, 26*). Using cryo-electron microscopy, inter-domain arrangements can be roughly classified (*27*). A rough idea about the global flexibility can be obtained by small-angle x-ray scattering (*17*). However, these current methods are either temporally or spatially limited and are therefore unable to resolve both fast and large inter-domain fluctuations usually occurring on the micro- to millisecond timescale.

With single-molecule methods, ensemble averaging is prevented and dynamics of local and global conformational changes can be determined simultaneously. The dynamics of the protein's accessible conformations can be transformed into free energy landscapes or other physical models. For analyzing the biological function, this abstraction is probably better suited than hypothetical real-time imaging. With single-molecule force spectroscopy beautiful mechanisms of protein machineries have been uncovered such as the rectified movement of the myosin-actin-complex (*28, 29*). In contrast to force-based methods, single-molecule Förster resonance energy transfer (smFRET) enables direct observation of near-native inter-molecular and intra-molecular dynamics on a broad range of timescales from nanoseconds to minutes (*30-34*). Hybrid methods that combine, for example, single-molecule FRET and cryo-electron microscopy enable further insights such as the analysis of short-lived hidden intermediates (*35*). The succession of a certain number of protein states and their transition probabilities can be analyzed with hidden Markov modelling of single-molecule fluorescence trajectories of immobilized proteins. In this manner, both the working cycle and the overall function can be consecutively evaluated. Multi-component interactions can be analyzed similarly, as recently accomplished for Hsp90, the co-chaperone p23 and ATP. A p23-induced coupling slightly increased the flux of the system (*36*). However, single-molecule approaches are limited by one-dimensional trajectories and - often - by a certain range of timescales.

Besides determination of Markovian dynamics, single-molecule FRET can be used to determine exact distances and distance distributions on fast timescales and thus has the potential to maintain the overall structure-dynamics relationship.

By combining known conformations and experimental distance distributions obtained for some pairs of residues, relevant slow modes of the protein can be analyzed (*37*), to obtain ideas about possible conformational changes. More recent developments add a biasing potential to

conventional MD simulation to facilitate the sampling of a structural ensemble, whose ensemble-averaged observables agree with the experimental distance distributions (*38, 39*). These methods are suited for small proteins and, regarding labeling, they are better suited for small spin labels than for large dye labels, because the accessible space of spin labels is negligible.

Single-molecule FRET based distances have been combined with structural information from x-ray crystallography to find novel static interaction sides (*40-42*). In contrast to molecular dynamics approaches, the domains are assumed to remain rigid on the observed timescale. This allows for the calculation of exact dye accessible volumes that have significant impact on the certainty for structure determination.

The approach presented herein uses self-consistent FRET networks to determine both structural ensembles and their dynamics. Time-resolved distance distributions prevent ensemble averaging and, at the same time, enable determination of inter-domain dynamics and dynamics of protein-protein interactions.

Furthermore, for such a detailed analysis of structural dynamics, accuracy and confidence of single-molecule FRET has to be brought to an unprecedented level. FRET describes a non-radiative and non-contact energy transfer between two dipoles; however, this electrodynamic phenomenon is not yet fully understood for several specific cases. Additionally, accurate intra- and inter-molecular distance determination is compromised by over- or underestimation of Förster's approximations for weak dipole-dipole coupling and by incomplete considerations of the numerous photophysical parameters. As smFRET is gaining recognition in biomolecular studies and is becoming a fundamental tool for hundreds of researchers, these issues are increasingly debated in literature – sometimes with contradicting results. Therefore, a comprehensive analysis of physicochemical properties and verification by a representative data set with self-consistent distances and time-resolved anisotropies became necessary, and it is presented in this work.

The path from static structures to dynamic structures via single-molecule FRET requires different methodological developments and analytical investigations while cutting across several scientific disciplines. As a consequence, the results and discussions section of this work is divided into four subchapters, each of them containing a dedicated small introduction and conclusion.

2 THEORY

2.1 Fluorescence

Photoluminescence means delayed radiation of light as a result of absorption of light. In contrast to scattered light, the excited electron of a molecule needs to be in resonance with the observed system. The direction and the spectrum of the emission are independent of the absorbed light. Photoluminescence as result of a singlet-singlet relaxation is called fluorescence and occurs on the nanoseconds timescale. The probability for the excited electron to reverse its spin (triplet state) and to relax to the ground state by emitting a photon, is relative small resulting in a relative long phosphorescence lifetimes. A term scheme illustrates possible electronic transitions (Figure 3A).

The singlet states are spitted into fast relaxing vibrational states, while some transitions are more probable than other transitions (Franck-Condon principle). The shape of the spectrum is determined by the transition probabilities between these vibrational states. Collision with water or other molecules lead to a broadening of the peaks and thus a more continuous spectrum. The absorption and emission spectra of two artificial fluorophores (i.e. dyes), Atto550 and Atto647N, in aqueous solution are shown in Figure 9. These dyes are derivatives of classic aromatic molecules being optimized especially with regards to the quantum efficiency, the probability of emitting a photon during singlet-singlet relaxation.

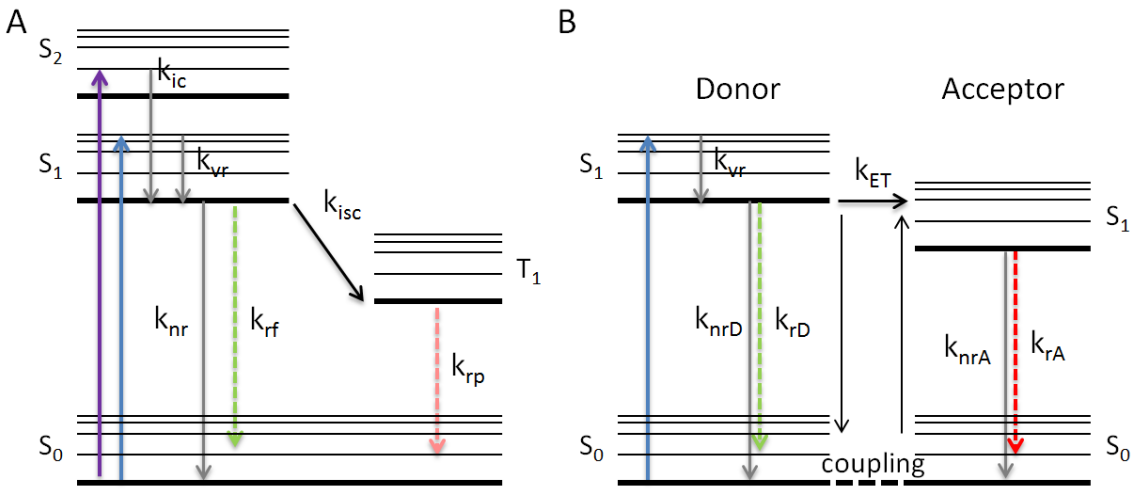


Figure 3: Jablonski diagram for (A) photoluminescence in general and (B) Förster resonance energy transfer with singlet states S_i, triplet states T_i and rates k for internal conversion *ic*, vibrionic relaxation *vr*, non-radiative decay *nr*, radiative decay *r*, triplet transfer *T* and energy transfer *ET*. Coulomb coupling leads to simultaneous deexcitation of the donor and excitation of the acceptor (B).

2.2 Förster resonance energy transfer

Resonance energy transfer (RET) is an electrodynamic phenomenon and is based on induced dipole-dipole coupling. The Coulomb interaction between two dye dipoles $V(43)$ that are far apart can be approximated with point dipoles for donor μ_D and acceptor μ_A:

$$V = \kappa \frac{|\mu_D||\mu_A|}{R_{DA}^3} \qquad (1)$$

where $\kappa = cos\beta_{DA} - 3cos\theta_D cos\theta_A$ represents the dipole orientation factor being extracted from the electric dipole moments, where β_{DA} is the angle between the dipole vectors and θ_D and θ_A are the angles between the respective dipoles and the dye-separation vector R_{DA}.

If the spectra of two chromophores differ from each other, the energy can be transferred nearly irreversible from the higher energy dye (donor) to the lower energy dye (acceptor). In the weak coupling regime the dimer spectrum equals the sum of donor and acceptor spectrum. That is why Fermi's Golden Rule from perturbation theory can be used to describe the transition rate from donor to acceptor:

$$k_{D^*A \to DA^*} = \frac{2\pi}{\hbar} \rho(E_{A^*D})|V|^2 \qquad (2)$$

with the perturbation operator (or electronic coupling) V and the density of final states ρ. The latter is replaced by the sum over all accessible transitions between the vibrational states of donor and acceptor. With some approximations one obtains:

$$k_{D^*A\to DA^*} = \frac{2\pi}{\hbar}\sum_{i,j,k,l} P_{D,i}P_{A,k}|V_{DA}|^2\left|\langle\chi_{D^*,i}|\chi_{D,j}\rangle\right|^2\left|\langle\chi_{A,k}|\chi_{A^*,l}\rangle\right|^2 \qquad (3)$$

with the Boltzmann occupation probabilities P of the vibrational states and the donor D and acceptor A. The observation that the vibrational overlap factors and the electronic coupling can be separated was first made by Förster (*44, 45*) and results in:

$$k_{D^*A\to DA^*} = \frac{2\pi}{\hbar}\frac{\kappa^2}{R^6}\sum_{i,j,k,l} P_{D,i}P_{A,k}\cdot|\mu_D|^2\left|\langle\chi_{D^*,i}|\chi_{D,j}\rangle\right|^2\cdot|\mu_A|^2\left|\langle\chi_{A,k}|\chi_{A^*,l}\rangle\right|^2 \qquad (4)$$

Förster arrives to an expression for the transfer rate, which depends on the independent spectral characteristics of donor and acceptor dye:

$$k_T(R) = \frac{1}{\tau_{D0}}\left(\frac{R_0}{R}\right)^6 \qquad (5)$$

τ_{D0} is the lifetime of the donor fluorescence in absence of the acceptor depending on radiative and non-radiative decays: $\tau_{D0}^{-1}=k_{D0r}+k_{D0nr}$ and R_0 is the so called Förster radius, at which the transfer rate equals the donor fluorescence rate. It includes constants for dipole-dipole coupling and dye spectra:

$${R_0}^6 = \frac{9000(ln10)}{128\pi^5}\cdot\frac{\kappa^2 Q_D}{N_A n^4}\cdot\int_0^\infty F_D(\lambda)\varepsilon_A(\lambda)\lambda^4 d\lambda \qquad (6)$$

The spectral overlap integral between donor emission and acceptor absorption is directly proportional to the energy transfer rate. It integrates all possible combinations of transitions over all frequencies at which simultaneous vibronic excitation of the donor and vibronic deexcitation of the acceptor occur. These vibronic levels depend on the environment of the dyes and the strength of this dependence is dye-characteristic, too. F_D is the donor fluorescence intensity in this case, N_A is the Avogadro's constant in this case (not to be confused with the acceptor fluorescence intensity N_A), Q_D is the quantum yield of the donor, n is the refractive index of the medium. The orientation factor κ^2 is assumed to be 2/3 if the dyes undergo fast isotropic motions. Dynamic orientation averaging is assured only if the dipoles sample all possible relative orientations on a timescale faster than the energy transfer rate. Otherwise, the transfer rates would have to be calculated individually for each relative dipole orientation. This has to be considered for the choice of the dyes and for the choice of the approach for structure generation (see Chapter 4.1 and 4.3). The Förster radius of the optimized FRET dye pair Atto550-Atto647N in aqueous solution is 65 Å (see Figure 9 for the respective overlap spectrum).

The transfer efficiency is defined as the transfer rate divided by all rates from the donor singlet state (see also Figure 3 right):

$$E = \frac{k_T(R)}{k_r + k_{nr} + k_T(R)} \qquad (7)$$

Equations (5) and (7) lead to a simple relation between the inter-dye distance and the FRET efficiency that can be observed in different ways (see Chapter 3.2.3):

$$E = \frac{1}{1 + \left(\frac{R}{R_0}\right)^6} \qquad (8)$$

After Förster published his theoretical paper in 1948, in 1954 Bowen and Livingston (*46*) observed energy transfer from 1-Chloroanthracene to perylene in spite of very low quantum yields. Even before Förster, Perrin already supposed distance dependency when observing increasing depolarization by increasing dye concentrations. Dexter and others extended Förster's theory considering, for example, overlapping electronic orbitals for very small distances. Aside from other coupling mechanisms that come into play at certain distance regimes, Förster has done several approximations including the ideal dipole approximation. Within the last 10 years several novel theories were suggested, providing more accurate and/or more general solutions for intermediate and weak coupling, for example, by using semiempirical quantum chemical methods (*47*) or by using modified Redfield equations (*48, 49*) or by extending Coulomb interaction with correlation and exchange contributions with the help of time-dependent density functional theory (*50*). The latter concludes that the dynamic averaging of the orientation factor cancels errors in the ideal dipole approximation, which breaks down in case of specific dipole orientations for distances up to 50 Å. Because there are also several practical limitations for distances smaller than 45 Å and practical reasons why dynamic averaging should be warranted (see Chapter 4.1), FRET is still an excellent approximation (and elegant solution) for distances larger than 45 Å, but only if the dyes are homogenous in sense of spectra and dipole orientations. Defining some rules and considering dye characteristics on the basis of a huge data set can revise FRET as a molecular ruler (see Chapter 4.1).

2.3 Protein structure & energetics

A protein is a peptide chain of numerous amino acids that are ligated during the mRNA transcription process. Protein function in and outside the cell is determined by its structure and dynamics, mainly by time-dependent conformational changes (*51-54*). The peptide chain undergoes an energetically driven self-assembly process called protein folding demonstrated by Hsp90 in Figure 4. The general picture is that secondary structure is formed first, followed by tertiary structure formation and finally assembling to quaternary structures, whereat higher level

structures can stabilize the underlying elements. Amino acid chains (primary structure) assemble to alpha helixes and beta sheets (secondary structure), which built up stable domains (tertiary structure) often with typical motives. Large proteins consist of several domains (quaternary structure). While secondary structure elements are mainly stabilized by hydrogen bonds and Van der Waals interactions, tertiary structures are governed by hydrophobic interactions, ionic interactions, hydrogen bonds and disulfide bonds. The assembly of protein domains is dominated by hydrophobic interactions, while other long ranging interactions such as ionic or hydrogen bonds increase specificity. These stabilizing energy contributions compete with destabilizing contributions such as conformational entropy. A rough average value for the free energy of protein folding can be estimated at 10 k_BT per 100 amino acids, whereat the intermediate enthalpy of two (charged) amino acid interactions can be 5 k_BT or more.

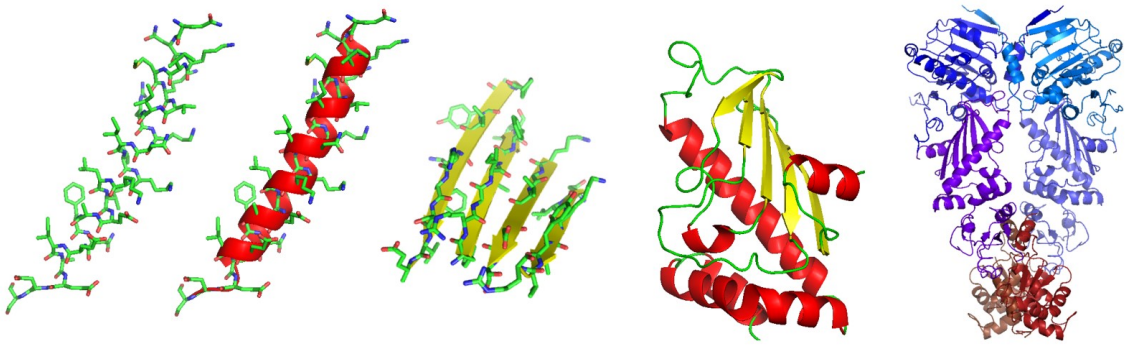


Figure 4: Protein structure levels shown for yeast Hsp90. From left to right: Peptide chain of amino acids with a specific sequence (primary structure), alpha helix and beta sheet (secondary structure), stable domain of secondary structure elements (tertiary structure) and arrangement of domains to the final protein structure (quaternary structure). Structures were created with PyMol and are not to scale.

The stability of a domain or global conformation (i.e. state) can be described by its lifetime (i.e. dwell time). A logarithmic scale of these dwell times is proportional to free energy (Figure 5). The domains of a protein can globally rearrange between different states with dwell times ranging from milliseconds to minutes and more, equivalent to about 5-20 k_BT. Larger domains can be usually considered as rigid building blocks on the timescale of these rearrangements. Smaller elements such as loops can undergo conformational changes within microseconds and translational motion of loops and single residues usually occurs within nanoseconds.

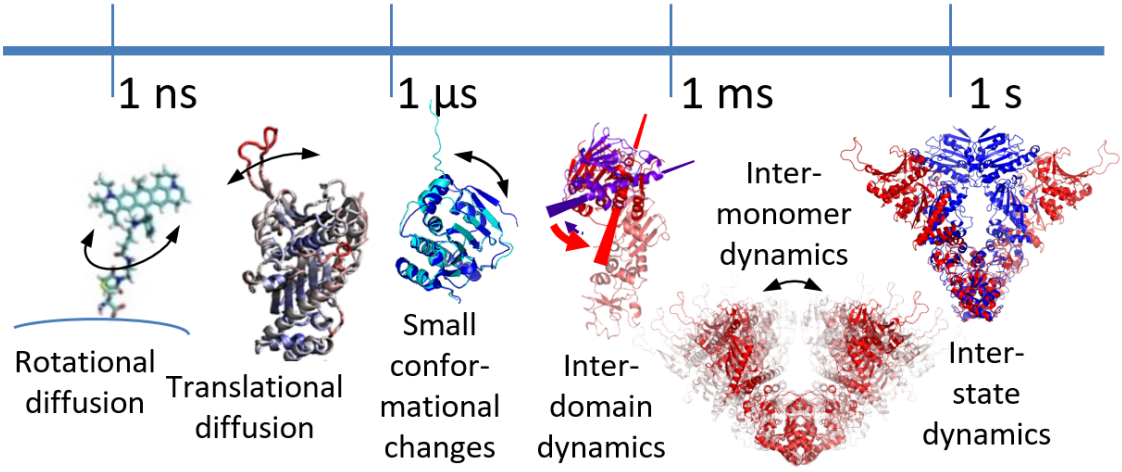


Figure 5: Structural and conformational dynamics of folded proteins on a broad range of timescales.

The conformational equilibrium of a protein can be dramatically changed through an interaction with another molecule. Figure 6 describes possible thermodynamic modes of action that can be distinguished by the free energy landscape or by the observed forward and backward rates. Protein-protein or protein-molecule interaction can lead to stabilization by decreasing an energy minimum and or to a catalytic effect by decreasing the activation free energy. While these are usually reversible processes, energy consumption (for example due to ATP hydrolysis) is often coupled to irreversible structural changes leading to global directionality.

The succession of a certain number of states and their transition probabilities can be analyzed with hidden Markov modelling of single-molecule trajectories of the protein under near-native conditions (*36*). However, information is in most cases limited by few state observables, time resolution and conformational flexibility. The conformational flexibility within a state, as well as, diffusive interaction can also influence the kinetics (*55, 56*) and could be seen as an additional interaction mechanism.

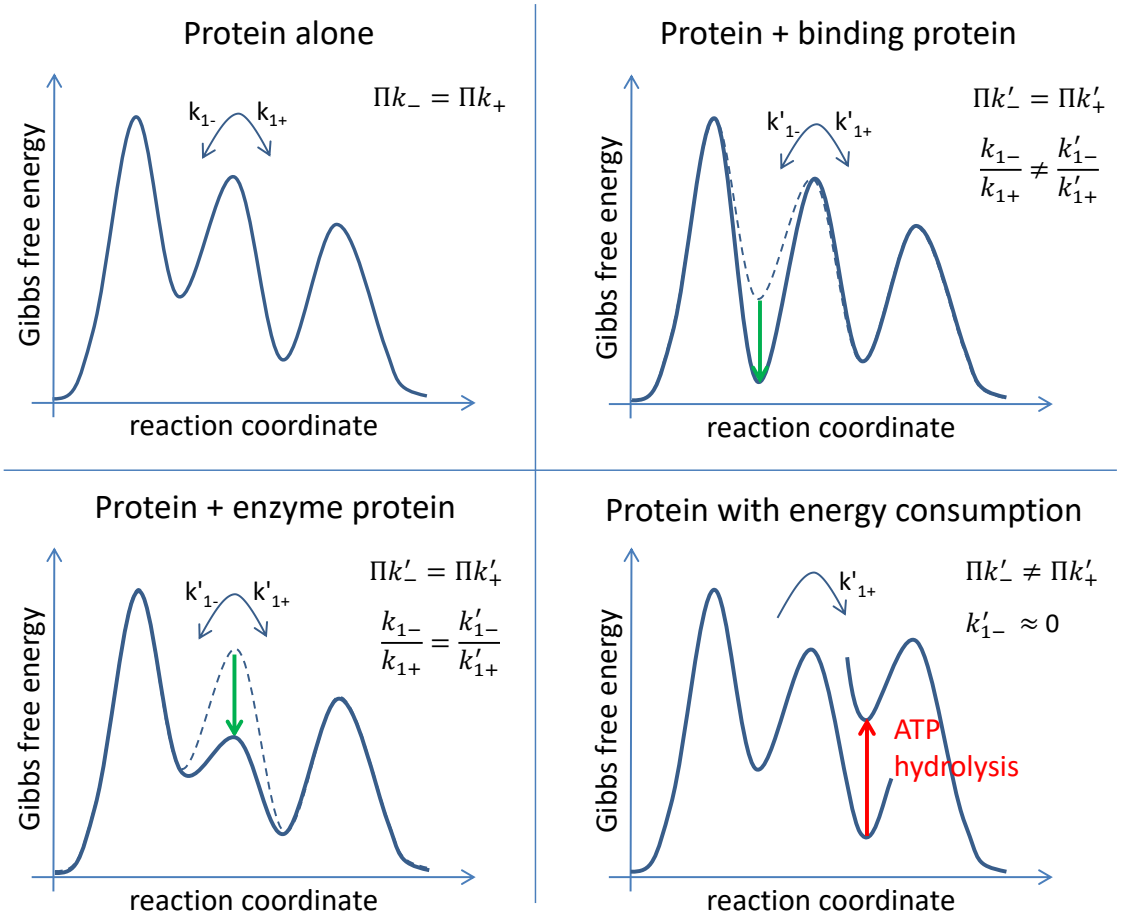


Figure 6: Thermodynamic modes of action for protein-protein interactions demonstrated with the free energy landscapes of a protein that is influenced by other proteins or nucleotides. Forward and backward rates between two states (k_+ and k_-) are observable parameters.

Thermodynamic quantities

A protein or multi-protein system is in thermodynamic equilibrium if no macroscopic changes such as flux of energy occur. The conformational changes of a protein can be enthalpy-driven (due to attractive or repulsive forces) or entropy-driven (for example due to water exclusion during dimerization). The change of Gibbs free energy of a protein depends on the change of enthalpy ΔH and change of entropy ΔS:

$$\Delta G = \Delta H - T * \Delta S \qquad (9)$$

In isolation, the protein spends most of the time in its lowest energy conformation. The rate k from one state to another of a protein, which is in equilibrium in canonical ensemble, can be described by the Arrhenius-like equation:

$$k = A \cdot e^{-\frac{\Delta G^{\ddagger}}{R \cdot T}} \qquad (10)$$

$\Delta G^{\ddagger}$ is the Gibbs free energy of activation, A the Arrhenius prefactor, which describes here the maximum frequency due to molecular collisions and can be estimated at 10^8 s^{-1} for proteins (*57, 58*) and R is the universal gas constant. Determining the dependency of the rate k on the temperature T, one obtains the activation free energy without knowledge of the prefactor. The dependency of the frequency factor A on the temperature can be neglected, if the reaction is not diffusion limited. Division of two Arrhenius equations for two Temperatures T_1 and T_2 yields:

$$\Delta G^{\ddagger} = R \cdot \frac{T_2 \cdot T_1}{T_2 - T_1} \cdot ln\left(\frac{k_1}{k_2}\right) \qquad (11)$$

The free energy difference ΔG between two states with equal prefactors is related to the equilibrium constant K between both states:

$$\Delta G = -RT * ln\left(\frac{k_{1\to2}}{k_{2\to1}}\right) \qquad (12)$$

Here K is expressed as the quotient of observed forward and backward rates between state 1 and state 2, but more generally it can be understood as the reaction quotient of activity values.

Transitions rate

The transition rate k from one protein state to another can be observed recording waiting time distributions. The probability that one observes an event after time t can be derived from the Poisson distribution (*59*) and is for simple systems:

$$P(t) = 1/k \cdot e^{-t/k} \qquad (13)$$

Ergodicity is given as long as heterogeneity among the observed molecules such as post modifications can be excluded.

From Stokes' law and Einstein's relation follows the diffusion coefficient:

$$D = \frac{\langle x^2 \rangle}{2t} = \frac{k_B T}{6\pi\eta R_H} \qquad (14)$$

with which diffusion limited rates can be estimated. Here η is the viscosity and R_H the hydrodynamic radius of the observed particle and $k_B T$ scales for the energy in molecular systems. Rotational diffusion can be calculated by the Stokes-Einstein-Debye equation, for example to describe the rotational motion of the dye dipole axis:

$$D_{rot} = \frac{k_B T}{8\pi\eta R_{rot}^3} \qquad (15)$$

The activation energy of a diffusion limited reaction can be estimated at 8-25 kJ/mol, mainly resulting from the temperature dependence of the viscosity of water.

3 METHODS & DEVELOPMENT

3.1 Optical setup

Within this work, a confocal setup for single-molecule fluorescence detection (schematically shown in Figure 7) was built. The main feature is the simultaneous polarization- and color-sensitive detection of single photons after excitation of freely diffusing single molecules. This allows detecting numerous fluorescence parameters such as FRET efficiency, fluorescence lifetime, stoichiometry and time-resolved anisotropy.

The setup design presented here is partially based on other published designs (*30, 60, 61*). Some of the implemented features were gained from great compendia for setup alignment (*62-64*), parameter optimization (*65*) and single-molecule FRET (*66-68*). Further features developed within this work and the main components of the setup are briefly described in the following.

A linearly polarized green diode laser (532 nm LDH-P-FA-530, PicoQuant) is focused wpith an Apo 60x water immersion objective (Nikon). The 50µm-pinhole reduces the detection volume, within which the dyes of a single molecule are excited. Fluorescence photons from FRET donor and acceptor are spectrally and polarization-sensitively separated and recorded with single photon avalanche photodiodes (PDM-50µm and τ-SPAD, PicoQuant) with picosecond resolution utilizing a commercial data acquisition system (HydraHarp 400, PicoQuant). A second red laser (640 nm LDH-D-C-640, PicoQuant) directly excites the acceptor dye to check its presence and to derive photophysical correction factors (Chapter 3.2.3). A third blue-green laser was introduced for additional selection criteria and brightness detection (see Chapter 6.2.1). All lasers are spatially aligned and temporally pulsed interleaved with repetition rates of 20 MHz enabling detection of fast conformational changes. The measurement chambers were passivated with PEG (see Chapter 6.1) to reduce surface interactions and preserve the analyte concentrations.

The detected single photons are time-correlated to a laser pulse with picosecond resolution (called micro-time) and accumulated in one histogram per detector (Figure 8 left). In a separate time channel, the photons are counted with microsecond resolution (called macro-time) to obtain photon bursts from single diffusing molecules (Figure 8 right). By combing micro-time data (fluorescence lifetimes) and macro-time data (efficiencies and stoichiometries, among others), sub-ensemble fluorescence lifetimes and sub-ensemble time-resolved anisotropies can be calculated (see Figure 24). The laser intensities were adjusted such that on average 50 photon counts per millisecond are obtained. With a threshold of 100 counts per millisecond a compromise was found between a sufficient signal to noise ratio of 10 (for distinguishing several sample specific efficiencies) and a low probability of simultaneous excitation of donor and acceptor with the green laser. The latter probability is ca. 1 % with a laser pulse repetition rate of 20 MHz, a direct excitation probability of the acceptor with the green laser of 10 % and when assuming a total quantum efficiency of 10 % (including quantum efficiencies of dye and setup).

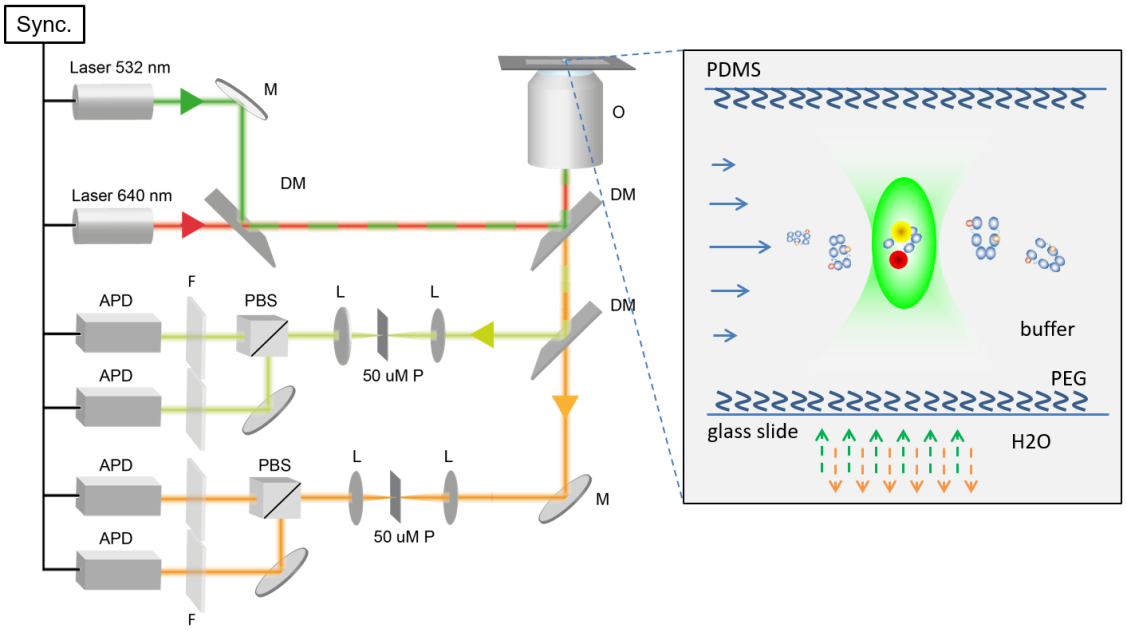


Figure 7: Schematics of the confocal setup with pulsed interleaved excitation and polarization- and color-sensitive detection. The most important elements are specified: Objective (O), dichroic mirror (DM), pinhole (P), polarizing beam splitter (BPS), spectral filter (F), avalanche photo diode (APD) and electronic picosecond synchronization of laser pulses and single photon counting (Sync). The measurement chamber (right) consists either of a PEG passivated glass slide and Teflon cavities or PDMS channels for combination with microfluidics.

The samples in this work were normally labeled with the high efficiency dyes Atto550, Atto532, Atto647N (AttoTec, Germany) or Alexa647 (Life Technologies, Germany), whereas the former two served as FRET donors and the latter two as FRET acceptors. The absorption and emission spectra of Atto550 and Atto647N (measured on a Lambda25 UV/VIS spectrometer, PerkinElmer

or a Jasco FP8500 spectra fluorimeter, Jasco, respectively) are plotted in Figure 9, together with their calculated spectral overlap and the transmission range of the applied spectral filters.

Such filters separate florescence from up to six orders of magnitude brighter excitation light exploiting the Stokes shift between respective photons. The dye brightness is an important guide value for optimized detection efficiency and is about 50.000 photons per molecule per second for Atto647N and confocal detection. Redox agents can improve brightness and stability (*69*).

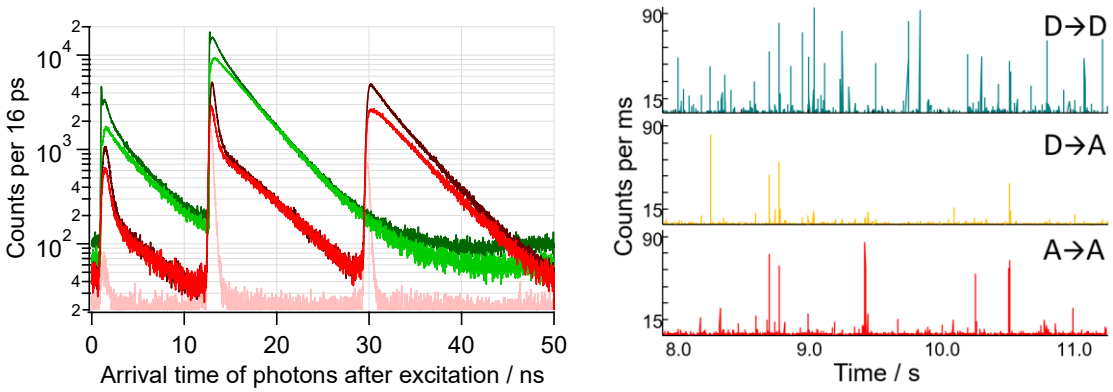


Figure 8: Left: Time correlated single photon counting (TCSPC) with three lasers and four detection channels measuring a test sample labeled with a donor and acceptor dye. The following three interleaved pulsed lasers each with a repetition rate of 20 MHz are used: 507 nm (left), 532 nm (middle) and 640 nm (right). The detector channels filter for "green" fluorescence photons arising from the donor dye (540-605 nm, green colored decays) and for "red" fluorescence photons mainly arising from the acceptor dye (655-710 nm, red colored decays) and additionally for parallel (light colors) and perpendicular (dark colors) polarization. The instrument response function exemplary shown for the "red+perpendicular" channel (pink), is used to deconvolute detection delays. Right: Photon bursts versus elapsed measurement time for the donor channel (green fluorescence) after donor excitation (with green laser, top), the acceptor channel (red fluorescence) after acceptor excitation (with red laser, middle) and the acceptor channel (red fluorescence) after acceptor excitation (with red laser, bottom).

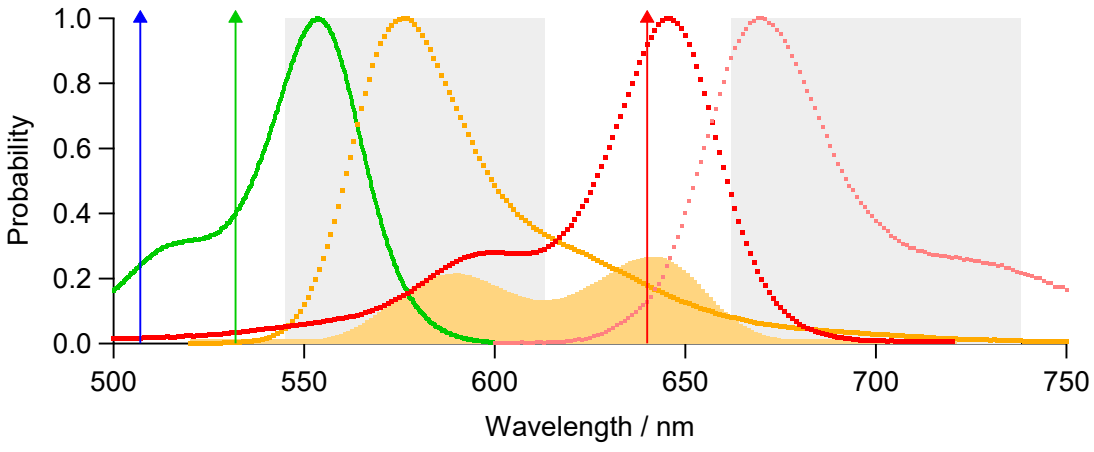


Figure 9: Absorption and emission spectra of Atto550 and Atto647N with carboxyl groups in aqueous solution (dots) are depicted, as well as, their overlap integral (orange area) calculated with equation (6), the wavelengths of the utilized lasers (arrows) and the transmission ranges of the utilized optical filters (gray).

A typical diffraction limited confocal volume for visible laser light, a numerical aperture of 1.2 and 50 µm pinholes is about one femtoliter. The smaller the confocal volume the higher is the tolerable concentration of the sample for single-molecule observations. However, in this work the confocal volumes were enlarged about 8 times above diffraction limitation to increase the burst intensities for fast scanning. Furthermore, the difference in the diffraction limited volumes of red and green light is thereby circumvented. Equal confocal volumes are important for FRET efficiency correction and for fluorescence cross correlation spectroscopy for analyzing sub-millisecond kinetics (see Chapter 3.2). The effective confocal volumes were determined by three-dimensional scanning of a bright, green and red fluorescent bead that is smaller than the diffraction limit of the smallest laser focus (see Figure 10). The beads were trapped in a polymer with a refractive index equivalent to water by dehydrating the same (MY-133-MC, mypolmers). The knowledge of the exact sizes is important to determine sample concentrations or diffusion coefficients. Independently, the effective volumes were verified by determining the diffusion time through each laser with a standard dye of known diffusion coefficient (see Chapter 3.2.1).

As the aperture of each laser (diameter of collimated beam) was adapted individually to obtain equal confocal volumes, the relative positions of the confocal foci cannot be stabilized by a glass fiber. Temperature stabilization (±1 K) reduced the drift of the foci to some few nanometers per hour. The relative positions of the foci were measured by scanning a fluorescent bead (Figure 10) and adjusted iteratively with piezo-controlled mirrors.

The discrepancy between small concentrations necessary for single-molecule measurements and higher concentrations necessary for low affinity protein complexes was overcome with a microfluidic device (see Chapter 6.3). The presented device is capable to dilute the sample 10.000 times in about 10 ms. Most measurements were performed in scanning mode. Scanning the focus relative to the sample with a spiral shape with about 1 mm/s is advantageous for several reasons – namely to obtain exponentially distributed burst intensities (Chapter 3.2.2), to obtain homogeneous passage times through the confocal foci and to realize novel brightness detection (see Chapter 6.2).

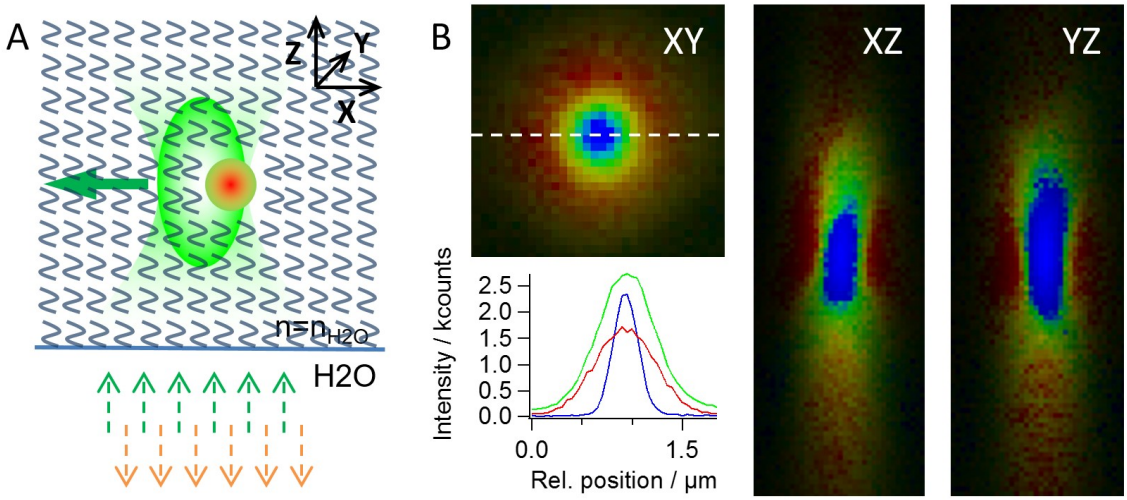


Figure 10: The confocal volumes and their relative positions are characterized simultaneously for all three lasers by three-dimensional scanning of a multi-spectral fluorescent bead. A: The sample (comprising beads fixed in a polymer) is scanned line-wise and uni-directional with a three-axis piezo actuator (PI, Germany) while being excited alternately with all lasers and while recording fluorescence photons in all detection channels. B: Different planes are shown for the overlapped laser foci that are colored according to their wavelength. The laser beam profiles indicate perfect overlap of the mean positions and equal Gaussian distributions for the green and red laser. The diffraction limited blue laser is used for novel brightness detection (see Chapter 6.2.1).

3.2 Fluorescence parameters

The four presented detection channels (green or red fluorescence with parallel or perpendicular polarization) with picosecond time resolution allow the determination of numerous fluorescence parameters. The most important parameters and further processing are briefly explained by means of concrete examples, either dye labeled DNA or Hsp90.

3.2.1 Fluorescence correlation spectroscopy

FCS is suited for determining concentration, hydrodynamic radius, molecular brightness, possible conformational rates and photophysical rates of any sample that is labeled with a FRET pair or fluorescent proteins, for example. The double stranded DNA labeled with the donor dye Atto550 and the acceptor dye Alexa647 from Figure 17 is used as an example to briefly demonstrate the potential of this method. The autocorrelation is calculated for the number of red photons N_A arriving at time t after the direct excitation of the acceptor:

$$G(\tau) = \frac{\langle N_A(t) N_A(t+\tau) \rangle}{\langle N_A(t) \rangle^2} \qquad (16)$$

The autocorrelation fit function (*70*) contains a diffusion term (right hand side) and in case of Alexa647 it should also contain a term for the triplet rate (left hand side):

$$G(\tau) = \left(1 + \sum_{j=1}^{2} \frac{T_j}{1 - T_j} e^{-\tau/\tau_{T,j}}\right) \frac{1}{\langle N \rangle} \left(1 + \left(\frac{\tau}{\tau_d}\right)^{\alpha}\right)^{-1} \left(1 + \frac{1}{K^2}\left(\frac{\tau}{\tau_d}\right)^{\alpha}\right)^{-1/2} \qquad (17)$$

with the diffusion time τ_d, the mean number of molecules in the confocal volume $\langle N \rangle$, the structure parameter $K = z_0/w_0$ (with effective focus diameters in z-direction and xy-direction estimated from a bead scan (see Chapter 3.1)), the relaxation time for the triplet path $\tau_T^{-1} = k_{TS} + k_{ex}k_{ST}/(k_{ex} + k_r + k_{nr})$ (with rates for laser excitation ex, singlet-triplet transition ST and triplet-singlet relaxation TS) and the fraction of molecules in the triplet state $T = k_{ex}k_{ST}/\big(k_{ex}(k_{ST} + k_{TS}) + k_{TS}(k_{ST} + k_r + k_{nr})\big)$. A value of less than 1 for the exponent α indicates anomalous diffusion and might indicate molecular crowding (*71*).

Fitting the autocorrelation of the fluorescence of Alexa647 labeled to DNA (Figure 11) reveals the diffusion time $\tau_d = 1.3\,ms$ und thus the diffusion coefficient:$D = w_0^2/(4\,\tau_d) \approx 62\,\mu m^2/s$. With equation (14) one obtains a hydrodynamic radius R_H = 3.6 nm, which fits to the size of the 38 base pair long B-DNA. The concentration of the sample is estimated at $C = \langle N \rangle/\big(\pi^{3/2} w_0^2 z_0 N_A\big) = 110\,pM$ (with the Avogadro constant N_A), which agrees well with the dilution factor and the concentration determined by absorbance.

The setup dependent molecular brightness of Alexa647 (bound to DNA) is $B = N/(\Delta t \langle N \rangle) \approx 13.000\ \mathrm{photons/s}$ per molecule. This is about four times darker than Atto647N under similar conditions. The relative high fraction of molecules in the triplet state (30% for Alexa647 vs. 3% for Atto647N) is certainly a reason for the reduced brightness (see Figure 23), besides the lower quantum yield (33% for Alexa647 vs. 64% for Atto647N).

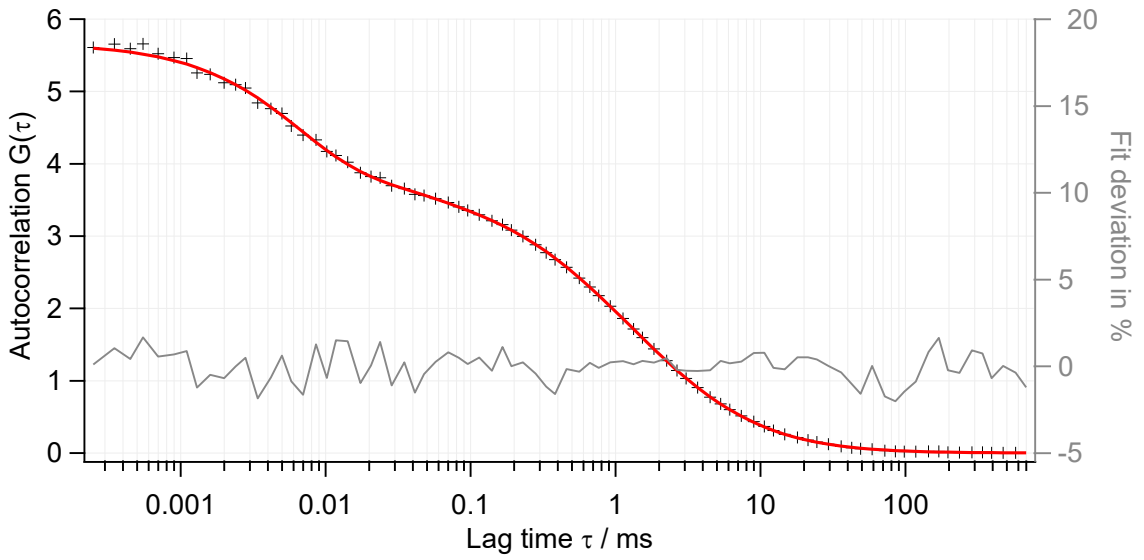


Figure 11: Fluorescence correlation spectroscopy: Autocorrelation of the red photons after excitation of Alexa647 labeled to double stranded DNA (black crosses) is fitted with equation (17) (red line) considering diffusion and intersystem crossing resulting in minimal residual deviations (gray line).

Although FCS is an ensemble measurement, protein heterogeneities can be revealed under certain conditions, for example, when the protein changes faster than 1 ms between two conformations that differ in FRET efficiency (see Chapter 3.2.6).

3.2.2 Photon clustering

Single-molecule measurements are suited to study heterogeneities in protein systems. Each time a freely diffusing molecule passes the confocal volumes of the excitation lasers, a fluorescence burst is recorded. The number of detected photons N is counted for each time interval $t+t_{bin}$ after the start of the measurement (macro-time). By scanning the sample relative to the confocal volume with constant velocity, the transition times of each molecule through the confocal volumes is constant and the intensities can be approximated with an exponential distribution (Figure 12). The photon intensity distribution is important for calculating correct shot noise limited FRET efficiency distributions (see Chapter 4.2).

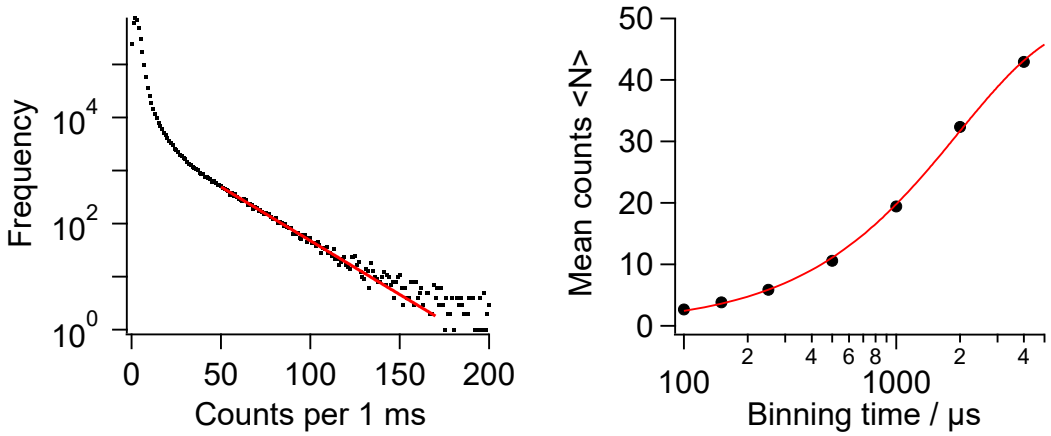


Figure 12: Left: Photon Counting Histogram (PCH) for the total number of photons n after donor excitation for a binning time of 1 ms in the scanning mode. Within the relevant range, the histogram can be approximated with an exponential distribution resulting in a mean intensity per bin of 21 counts. Right: Mean photon counts <*N*> as a function of time binning. The fitted saturation function gives a maximum mean number of photons per bin of 50 and an expected binning time of 2 ms, which corresponds to the mean passage time through the confocal laser spot.

Intensity thresholds are often set to reduce photons from background or from foreign particles that are usually less bright than the fluorescent dyes. When a lower threshold n_{th} is set, the expected number of photon counts per time bin increases. Interestingly, the threshold value – no

matter which size – is simply added to the expectation value for exponentially distributed photon intensities:

$$\boldsymbol{E}[n_{th} < X < \infty] = \int_{n_{th}}^{\infty} x \cdot e^{-x/\langle N \rangle} \Big/ \int_{n_{th}}^{\infty} e^{-x/\langle N \rangle} dx\, dx = n_{th} + \langle N \rangle \qquad (18)$$

Classically the photon traces are binned with constant time windows and an intensity threshold selects for a minimum brightness. A more sophisticated searching algorithm uses dynamic time windows and considers the inter-photon time, which includes also photons beneath the threshold. In contrast to total photon burst search, the dual color burst search DCBS verifies the consistency between green and red photons (*72*). This is necessary if the confocal volumes differ in size or if a dye suffers fast bleaching.

If one is interested in determining the timescale of distance fluctuations, the classical time binning is preferred to the inter-photon time binning. However, large time bins also include non-relevant sections with low intensities. That is why a combination of these algorisms is applied within this work. First the burst lengths are determined with DCBS, and then the bursts are chopped into very small time bins, which then are reintegrated to the wished binning time.

3.2.3 Transfer efficiency

The FRET efficiency can be determined from the percentage of the acceptor intensity N_A (acceptor photons arising only from energy transfer) in comparison to the total intensity $N=N_A+N_D$. For the measured intensities N_A^* and N_D^* the gamma factor γ, the fraction c of spectral crosstalk from donor emission into the detection channel of the acceptor, the fraction d of direct excitation of the acceptor from the donor excitation have to be considered:

$$E = \frac{N_A}{N_A + N_D} = \frac{N_A^* - dN_A^A - cN_D^*}{N_A^* - dN_A^A - cN_D^* + \gamma N_D^*} \qquad (19)$$

The gamma factor considers the quantum yields and detection efficiencies $\gamma=Q_A\eta_A/(Q_D\eta_D)$. The direct excitation factor $d=d_0/\beta$ depends on the beta factor $\beta=(\sigma_A I_A)/(\sigma_D I_D)$, which itself describes the ratio of laser intensities and absorption cross sections. N_A^A is the acceptor intensity upon acceptor excitation. The correction factors can be obtained either by separate measurements or more conveniently with a multi-parameter analysis by correlating efficiencies with stoichiometries (see Chapter 3.2.4).

With equation (8) and the Förster radius R_0 the inter-dye distance can be calculated. The Förster radius should be calculated separately for each dye environment. In aqueous solution the Förster radius for Atto550-COOH and Atto647N-COOH was calculated to be 65 Å (see Figure 9). The accuracy of R_0 depends very much on the orientation factor κ^2 (see Chapter 4.1).

Because shot noise, background and sample heterogeneities lead to a broadening of the expected efficiencies, at least 1000 events should be recorded in a histogram. If each photon burst contains more than about 100 counts, shot noise limited distributions can be approximated with Gaussians with the mean efficiencies equaling the maximum values. However, this procedure is incorrect, if the shot noise distributions are convoluted with dye or sample heterogeneities. Then a more sophisticated distribution analysis is required (Chapter 4.2). Processes that are much faster than the photon count rate, are averaged and usually do not broaden the efficiency distribution apart from shot noise.

In principle, the more straight-forward determination of the FRET efficiency is the donor lifetime, as it is quenched proportionally to the transfer rate:

$$E = \frac{k_T}{k_T + k_{D0}} = \frac{k_D - k_{D0}}{k_D} = 1 - \frac{\tau_D}{\tau_{D0}} \qquad (20)$$

The quenched donor lifetime τ_D declines with decreasing donor-acceptor distance:

$$R^6 = \frac{\tau_D}{\tau_{D0} - \tau_D} R_0^6 \qquad (21)$$

However, there are several practical limitations. In contrast to intensity based transfer efficiencies, fluorescence lifetimes can be more affected by fast processes such as linker motions and by background photons. This fact can be used to uncover very fast kinetics (see Chapter 3.2.6). But, different populations can be separated better by means of efficiencies than by means of lifetimes. Moreover, additional non-radiative rates can come into play for a certain state, but leave the globally determined reference lifetime τ_{D0} unaffected.

The accuracies of both approaches – lifetime based and intensity based distance determination – are verified in Chapter 4.1.

3.2.4 Multi-parameter analysis

The above introduced correction factors γ and β are obtained by correlating stoichiometries and FRET efficiencies of a two-state system or higher (a sample with at least two different expected FRET efficiencies) (*30, 73*). Therefore, the uncorrected stoichiometries $S^* = (N_A^* + N_D^*)/\left(N_A^* + N_D^* + N_A^{A*}\right)$ and uncorrected efficiencies $E^* = N_A^*/(N_A^* + N_D^*)$ are correlated. The corrected stoichiometry is defined as:

$$S = \frac{N_A^* - dN_A^{A*} - cN_D^* + \gamma N_D^*}{N_A^* - dN_A^{A*} - cN_D^* + \gamma N_D^* + N_A^{A*}/\beta} \qquad (22)$$

The correlation between stoichiometry and efficiency is given, because both parameters depend on the γ factor. When γ is not equal to 1, the different quantum efficiencies of donor and acceptor

dye lead to different stoichiometries at different transfer efficiencies. Before correcting γ and β, the cross talk parameter c and the direct excitation parameter d should be corrected. For this purpose, the donor-only and acceptor-only population of the samples are used. They represent molecules with only one donor or acceptor dye present, which cause non-zero efficiencies and non-zero stoichiometries, respectively. This has the advantage that the environment of the dyes are equal to the double-labeled sample and that no separate measurements are necessary. The crosstalk factor equals:

$$c = (1/E^*_{Donor} - 1)^{-1} \tag{23}$$

and the direct excitation factor equals:

$$d = \left(1/S^*_{Acceptor} - 1\right)^{-1} \tag{24}$$

By combining the uncorrected and corrected equations for efficiency and stoichiometry one obtains the following relationship between *S, E,* γ and β:

$$S = \frac{1}{1 + \gamma\beta + \beta(1-\gamma)E} \tag{25}$$

The parameters γ and β can be obtained by a least square fit of the scattered data when correlating *S* and *E*. Figure 13A shows all four sample populations and their correction functions and Figure 13B shows the same measurement after two correction iterations. At least two iterations of the upper procedure are necessary because *d* is related to β. In Figure 13B, the donor-only population is located at E=0, the acceptor-only population is located at S=0 and the double-labeled populations are parallel to *E* with S=0.5. The new parameters are γ=1, β=1, c=0 and d=0.

In this work, it is shown that also local gamma and beta factors of single states can be determined if the efficiency broadening is dominated by distance fluctuations (see Chapter 4.1).

Correlation of other photon burst parameters can be used to discover certain dye or protein characteristics. For example, correlating donor lifetime and efficiencies uncovers fast dynamics or hidden short lived states of the dyes or the protein (Figure 14). Correlating static anisotropy and donor lifetime uncovers hidden states due to dye-sticking or protein-protein interaction.

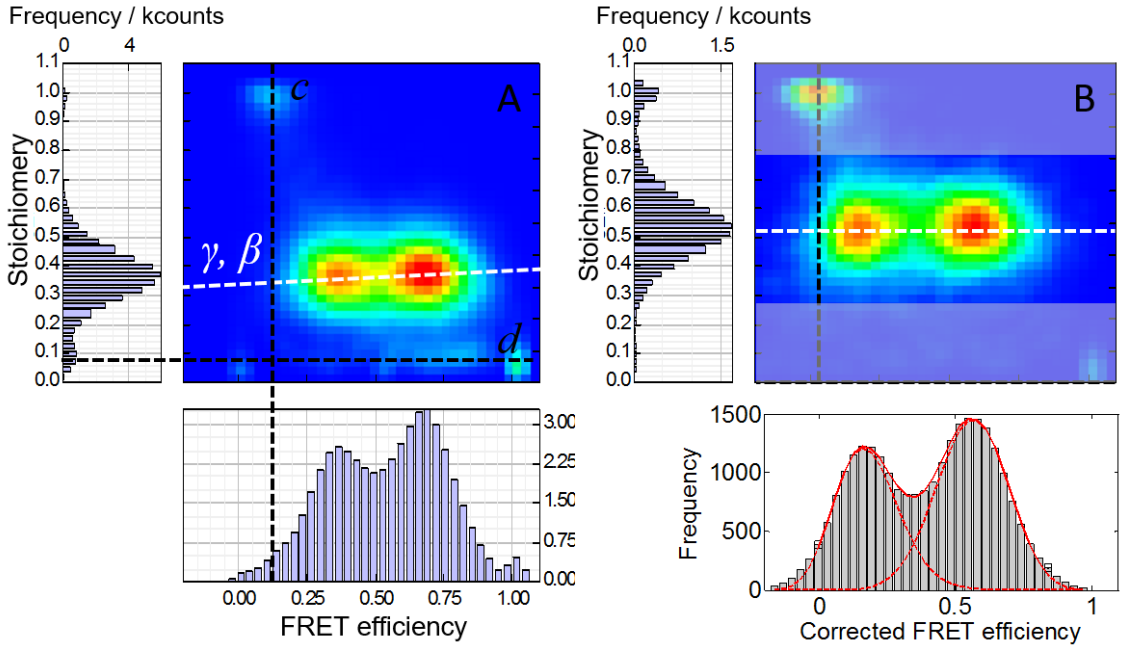


Figure 13: Uncorrected (A) and corrected (B) stoichiometry-efficiency plots. The parameters before correction are: γ=1.2 β=2, c=0.12 and d=0.12 and after correction: γ=1, β=1, c=0 and d=0.

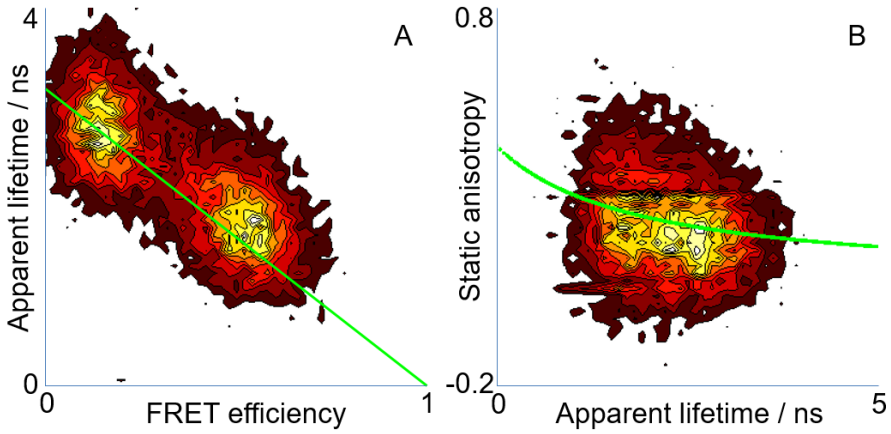


Figure 14: Two examples of multi-parameter fluorescence analysis: (A) The apparent donor lifetime (single exponential fit) is correlated with the corrected FRET efficiency for each single photon burst. Populations of static and non-quenched samples obey equation (20) (green line). Populations of fast fluctuating states deviate from the static line towards larger apparent lifetimes. (B) Static donor anisotropy versus apparent donor lifetime. Populations that deviate from the Perrin relation (green line) indicate long-lived high anisotropy states.

3.2.5 Time-resolved anisotropy

While the static fluorescence anisotropy is often used to detect protein-protein interactions, the time-resolved fluorescence anisotropy is especially important to determine the mobility of the

FRET dyes, which in turn is important for the orientation factor and accurate distances (see Chapter 4.1.4 for further discussions).

The fundamental anisotropy of freely diffusion dipoles is 0.4. However, a small depolarization term leads to a maximal possible anisotropy of about 0.38 for the dyes used in this work. The anisotropy can be calculated from the intensities N of parallel and perpendicularly polarized fluorescence after excitation with linearly polarized laser light:

$$A = \frac{N_{||} - GN_{\perp}}{N_{||} + 2GN_{\perp}} \qquad (26)$$

where G corrects for relative detection efficiencies. For objectives with a high numerical aperture depolarization factors have to be considered ($l1$=0.0308 $l2$=0.0368 for NA=1.2). See (*74*) for further details.

An equation describing time-resolved anisotropy is derived in Chapter 4.1.4. In principle, the anisotropy decays are fitted with a bi-exponential function, one term for fast rotational diffusion of the dye and another term for slower translational diffusion of the dye or the molecule to which the dye is bound. Figure 15 shows two examples, one for freely diffusing Atto647N-NHS and one for Atto647N bound to double stranded DNA. In presence of the DNA, the translational dye motions are restricted at larger timescales.

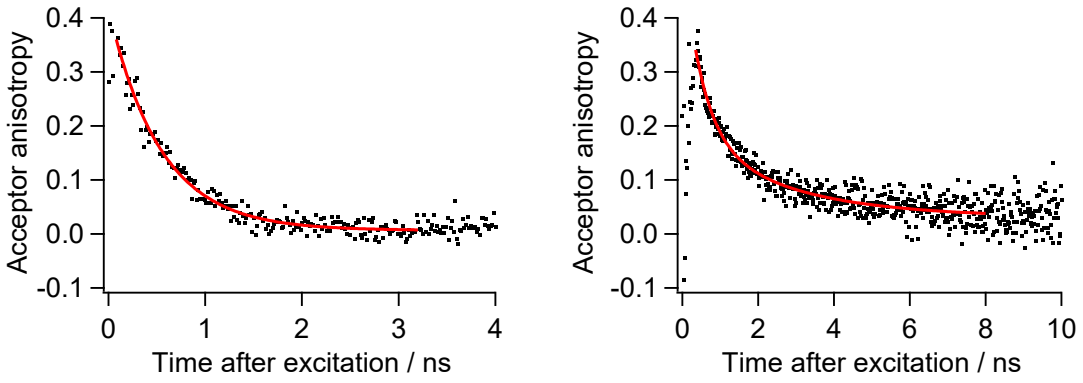


Figure 15: Time-resolved anisotropy of freely diffusing Atto647N-NHS (left) and Atto647N bound to double stranded 38-base-pair B-DNA (right). The fit coefficients are r_1=0.35±0.001, τ_1=0.53±0.01 ns, r_∞=0.006±0.001 (left) r_1=0.17±0.03, τ_1=0.51±0.1 ns, r_2=0.14±0.02, τ_2=2.6±0.6 ns, r_∞=0.03±0.01 (right).

3.2.6 Kinetics from fluorescence

A multi-domain protein in solution undergoes conformational changes that happen on a broad range of timescales. There are slow conformational changes of global domain arrangements and there are fast conformational changes of small elements and inter-domain fluctuations. Large conformational changes can be the result of the combination of smaller conformational changes.

As FRET occurs within nanoseconds it is especially suited to describe the dynamics of the conformational ensemble of a protein. Limitations such as signal-to-noise ratio and bleaching are the reason why different methods must be applied to capture dynamics on various timescales. The kinetic rates obtained by each method can be determined independently as long as they differ at least by an order of magnitude. This includes the assumption of ergodicity, i.e. the kinetic heterogeneity is assumed to be the same for individual molecules observed for long timescales and for different molecules observed for shorter timescales. However, one should take care of very slow conformational changes that are slower than the observation time and of very fast conformational changes that are averaged. Figure 68 summarizes the fluorescence based methods and corresponding timescales that were used in this work.

Slow conformational changes of freely diffusing FRET labeled molecules can be monitored by disequilibrating the ensemble, e.g. by the addition of nucleotides. In contrast to standard fluorescence microscopy, with a confocal microscope changes of small populations can be tracked, because snapshots of single molecules are recorded. But, due to limited statistics, only dwell times larger than 10 seconds can be observed.

Conformational transition rates at equilibrium can be determined by analyzing the transitions in the FRET efficiency time traces. Using a confocal microscope with an avalanche photodiode, the accessible time window lies between 1 ms and 10 s, depending on the laser intensity. Using a TIRF setup with an up-to-date EMCCD camera, the time resolution is limited to about 20 ms, but numerous molecules can be observed in parallel. Due to heterogeneities in extinction coefficients and quantum yields when using a TIRF setup, one should check the anti-correlation between donor intensity and FRET-sensitized acceptor intensity, as well as, the state of the acceptor dye with alternating direct excitation of the acceptor.

Faster equilibrium kinetics within a certain global state can be gained from photon bursts of freely diffusing molecules. The actual global state of each single molecule when passing through the confocal laser volumes can be identified via FRET-efficiency, stoichiometry and anisotropy.

Burst variance analysis (BVA) is a powerful tool to check for dynamics at the timescale between photon binning time and diffusion time (*75*). This method is insensitive to small subpopulations, with the advantage that linker dynamics and background photons can be ignored. On the other hand, it is sensitive to small differences in FRET efficiencies as required for the analysis of small fluctuations.

Fluorescence lifetime analysis reflects dynamics occurring between 10 ns and 1 ms (*76, 77*). This method is especially sensitive to small dynamic populations with large differences in FRET efficiencies. Thus, it can be seen as complementary to the burst variance analysis.

Based on the idea of the burst variance analysis, a more detailed analysis was developed within this work (Chapter 4.2). First, structural heterogeneities are separated from shot noise with a

probability distribution analysis and described with different models whereat Gaussian distributions resulted in optimal fits. Second, the timescale of the distance fluctuations is revealed by varying the binning time of the photon traces between 0.1 and 10 ms and plotting the standard deviation of the Gaussian distance distributions versus their observation time.

Fluorescence correlation spectroscopy of selected FRET efficiency populations accesses relaxation rates at timescales between 10 ns and diffusion time. With cross correlations of donor and acceptor channels protein kinetics that are even slower than the average diffusion time can be revealed. By dividing two different cross correlation functions (*78*):

$$G_{xy}(\tau) = T(\tau)X_{xy}(\tau, k_{12}, k_{21}, E_1, E_2), \qquad (27)$$

the diffusion term $T(\tau)$ is eliminated and the rates k_{12} and k_{21} or the FRET efficiencies E_1 or E_2 can be extracted, in the ideal case. Normally, the applicability is limited, as the diffusion times may differ depending on the effective sizes of the confocal volumes. That is one of the reasons why, within this work, the confocal volumes are enlarged above their diffraction limits and adapted to each other. Photophysical fluctuations and dark state blinking can still affect the remaining term X_{xy}, depending on the fluorescence channels x and y to be correlated. In contrast to (*78*), herein, the autocorrelation of the acceptor photons after donor excitation is divided by the autocorrelation of the acceptor photons after acceptor excitation. In this manner, the correlation term of the triplet relaxation rate of the acceptor can be canceled out. An additional advantage is that the acceptor photons after acceptor excitation are usually not influenced by the protein kinetics, so that the division of the autocorrelations can be fitted with simple (multi-) exponential functions. Figure 58 shows the divided autocorrelations G_{DA}/G_{AA} for an intra-monomer FRET pair of Hsp90 in presence of ADP (red) and AMPPNP (blue). The open ADP-bound state contains a dominant average rate of about 2 ms^{-1} and some smaller populated faster rates. The closed AMPPNP-bound state contains only a small populated rate of about 10 ms^{-1} and thus is as rigid as expected. As an additional control, the same procedure was done with a relative rigid double stranded DNA molecule. In this case, no kinetics could be observed between 1 µs and 1 ms.

3.3 Molecular dynamics simulation

The following molecular dynamics (MD) tools have been implemented by Martin Zacharias and Florian Kandzia.

Restrained MD simulations: Molecular dynamics simulations and energy minimization to generate Hsp90 structural models were performed using the *sander* module of the Amber14 molecular simulation package (*79*). During structural modeling the core structures of the N-, M- and C-domains were restrained to experimental reference structures by adding distance restraints between Cα atoms within 15 Å (force constant 3 kcal mol^{-1} $Å^{-2}$). This allows translational and

rotational mobility but only limited internal flexibility in each domain. Additionally, the core structure of the C-domains was kept close to the geometry of the x-ray structure, due to the corresponding experimental distances being equal in the presence of ADP and AMPPNP. For the structure generation steps, an implicit solvent model with a distance dependent dielectric constant (ε=4r) was employed.

As start structure a domain arrangement as found in the Grp94 homolog was used with the sequence of yeast Hsp90. The connection between N- and M-domains was modeled by a short Gly_{10} linker insert between residue 216 and 261. For inclusion of the FRET derived distance data, pseudo atoms with a van der Waals radius of 6.5 Å, representing approximately the size of the dyes, were connected to the Cβ atoms of the dye labeled residues using a pseudo bond (linker) of 11 Å. The distance between each donor-acceptor pseudo atom pair was allowed to vary within an interval given by the experimentally determined FRET derived distance range. Beyond the allowed distance interval further variation was penalized with a force constant of 3 kcal mol^{-1} $Å^{-2}$. In addition, the optimal spatial MN-arrangement in the dimer (see Chapter 4.3.4) served as additional positional restraints (on heavy atoms, with a force constant 1.0 kcal mol^{-1} $Å^{-2}$) during MD-based structure generation. In order to keep the structure symmetric with respect to the spatial arrangement of each monomer, the option to penalize the difference between two distances implemented in s*ander* module of the Amber14 package was employed. The option was applied to a subset of equivalent backbone distances within each monomer and between monomers. MD simulations were performed for 0.5 ns at 450 K followed by 0.5 ns at 300 K using a time step of 0.001 ps. Structures were finally energy minimized within 7500 steps of conjugate gradient minimization. Structural models of the open Hsp90 structure were re-evaluated by the FPS approach to check compatibility with the FRET data. The final mean-structure yielded an average deviation from the FRET distance data of $\chi^2 < 0.3$.

Unrestraint MD simulations: In addition to restrained MD simulations in implicit solvent for structure generation, unrestrained MD simulations starting from the mean open ADP-bound Hsp90 structure were performed in the presence of explicit solvent and surrounding ions. The structure was solvated in a truncated octahedral box with explicit TIP3P water molecules (*80*) and neutralized with sodium and chloride ions by means of the leap module and employing the parm14SB force field (*81*). Long range electrostatic interactions were calculated with the Particle Mesh Ewald (PME) method (*82*) and a real space cutoff radius of 9Å. During 0.5 ns equilibration the system was heated up to 300 K while heavy atoms of the protein were harmonically restrained (25 kcal mol^{-1} $Å^{-2}$) to positions in the starting structure. After removing positional restraints gradually during another 0.15 ns the structures were being equilibrated for 0.2 ns in an unrestrained simulation. The simulation was extended to 100 ns at 300 K and a pressure of 1 bar. For comparison explicit solvent MD simulations were started from the closed crystal structure of Hsp90 (pdb: 2cg9) after removing the p23 cofactor and supplemented with the same $(Gly)_{10}$ linker

between N- and M-domain as used for modeling the open Hsp90 structures. Equilibration followed the same protocol and simulations were extended to 100 ns.

The generated trajectories were evaluated using the cpptraj module of the Amber14 package. Calculations of the root mean square deviation (RMSD) of the backbone and root mean square fluctuations (RMSF) were performed using the cpptraj module. The RMSF was calculated as the mean over all heavy atoms of each residue.

Buried surface areas: For a subset of 200 snapshots obtained either from the simulation of the closed structure or from the simulation of the open structure, the buried surface area (BSA) of the CM interface was calculated. Therefore, the solvent accessible surface area of the C, M and CM domains has been calculated atom wise using the SHRAKE algorithm (*83*). This was done for both monomers of the Hsp90 dimer and the arithmetic mean was used in the following. For each residue, the normalized total contribution as well as the number of contributions to the BSA was determined. The same procedure was repeated for the NM interface.

3.4 Biochemistry

Mutagenesis and purification: Yeast Hsp90 (Hsp82) was used as a construct with a C-terminally inserted coiled-coil motif followed by a strep-tag (*23*). Cysteine point mutations were created with the QuickChange Lightening Kit (Stratagene, USA).

Hsp90 was expressed 4 h from pET28a in e.coli BL21DE3 cod+ (Stratagene, USA) after induction at OD_{600} 0.8 with 1 mM IPTG. Cells were harvested by centrifugation, re-suspended in PBS with 1 mM DTT and lysed in a French Press (Constant Systems). Lysate was centrifuged and protein purified on a Strep-Tactin Superflow column (IBA, Germany) according to the manufacturer's protocol. The elution was further purified by anion exchange chromatography on a HiTrap Q column (GE Healthcare, USA) using a linear gradient with increasing salt concentration (40 mM Hepes, 50 mM / 1 M NaCl, 1 mM DTT, pH 7.5) followed by gel filtration in 40 mM Hepes, 150 mM KCL, pH 7.5 on a Superdex 200 HiLoad 16/60 (GE Healthcare, USA). Protein purity was checked by SDS PAGE. P23 and Δ131Δ was expressed and purified as described earlier (*36, 84*).

Fluorescent labeling: Protein was site-specifically labeled with dyes via cysteine-maleimide chemistry. Dyes used in this study were Atto532, Atto550 as donor fluorophores, Atto647N (all AttoTec, Germany) and Alexa647 (Life Technologies, Germany) as acceptor fluorophores. Single cysteine mutants were incubated with 10 mM TCEP for 30 minutes at room temperature. TCEP was removed and protein buffered into PBS pH 6.7 using PD 10 Trap columns (GE Healthcare). The reduced protein was incubated with a threefold excess of maleimide dye for 1 h at room temperature. Free dye was removed with a PD-10 Trap column or in case of hydrophilic dyes by >10 dilution steps in a Vivaspin500 centrifugal concentrator (Sartorius Stedim Biotech,

Germany). Labeling efficiency depended on the labeling position and the dye, but generally exceeded 70 %.

Hetero dimers: Homo dimers (i.e. Hsp90 dimers with two equal dyes at the same position per monomer) were mixed with different homo dimers and incubated for 30 min at 42°C for monomer exchange. This resulted in 50 % hetero dimers that have one acceptor at one monomer and one donor dye at the other monomer. By correlating FRET efficiency with stoichiometry in a scatter plot, hetero dimers were selected.

In contrast, intra-monomer mutants (with two labels within one monomer) were exchanged with a 200-fold excess of WT-Hsp90 to exclude inter-monomer FRET.

Protein activity was checked by a regenerating ATPase assay (*85*). ATPase rates were measured at 30°C in 40 mM Hepes, 150 mM KCl, 10 mM MgCl2, pH 7.5. ATPase background could be detected by specific inhibition of Hsp90 with radicicol (Sigma) and was subtracted.

4 RESULTS & DISCUSSIONS

4.1 Molecular ruler under revision

Single-molecule FRET is a popular dynamic ruler for proteins under native conditions. It is extensively used to characterize the real-time dynamics between defined positions in a protein under near native conditions (*30-34*). For these experiments, two FRET dyes (donor and acceptor dye) are attached to the protein under investigation at defined positions. However, extracting accurate distances from such measurements depends on photophysics and unknown dye dipole orientations and accessible volumes.

Often, different approaches for FRET based distance determination are presented from their best side. This provided motivation to compare and analyze different measurement parameters, different detection methods, and different test systems, each in closed analogy with the underlying physics.

The first question to be addressed is which measurement parameter – transfer efficiency or donor lifetime – is better suited to obtain precise and confident distances for a broad range of distances (40-90 Å) and multiple molecule states, considering photophysics and experimental conditions.

Experimental uncertainties can dominate the theoretical uncertainties and are therefore particularly discussed in the subsequent subchapters. After comparing model distances with measured FRET efficiencies for a DNA test system, some rules were defined to exclude errors due to photophysical artifacts. In contrast to a mainly one-dimensional DNA molecule, a protein surface can dramatically restrict dye mobility. That is why more than 100 distances of a multi-domain protein were compared to the respective x-ray structure, to derive representative anisotropy criteria for sufficient dye mobility. On the basis of these test systems, the accuracy and confidence in distance determination via FRET could be improved.

Immobilization of single molecules, for example with a TIRF setup, has the disadvantage of large heterogeneities in extinction coefficients and quantum yields and therefore of large distributions of correction factors. Therefore, the cutting-edge confocal setup built herein (see Chapter 3.1) was used for the experimental verifications.

4.1.1 FRET parameters

The following theoretical estimation demonstrates the potential accuracies for both measurement parameters, the intensity based transfer efficiency and the donor lifetime (see Chapters 2.2 and 3.2.3 for the physical definitions). While the fluorescence lifetime is exponentially distributed, the transfer efficiencies can be perfectly approximated with a binomial distribution if the number of photons per molecule event is larger than the bin size of the efficiency histogram (see Chapter 4.2 for the explanation). The standard error of the mean of a sample of exponential or binomial random variables – not the standard deviation – is used to calculate the distance error for a FRET state with an efficiency of E=0.25 or a donor lifetime of τ_D=3 ns (if the donor lifetime in absence of the acceptor is τ_{D0}=4 ns), each with n=100 photons and a single-molecule event (photon burst) k=1. With the standard error $\sigma_E = \sqrt{nE(1-E)/k}$ of the mean of a sample of k binomial random variables, the propagated relative distance error for a single efficiency event (k=1) is about:

$$\frac{\Delta R_E}{R} = \left(1/\left(E - \frac{\sqrt{E(1-E)}}{\sqrt{kn}}\right) - 1\right)^{1/6} \Big/ (1/E - 1)^{1/6} - 1 = 4.3\ \% \qquad (28)$$

With the standard error $\sigma_\tau = \tau_D/\sqrt{kn}$ of the mean of a sample of k exponential random variables, the propagated relative distance error for a single lifetime burst (k=1) is about:

$$\frac{\Delta R_\tau}{R} = \left(\frac{\tau_{D0}}{\tau_D + \tau_D/\sqrt{kn}} - 1\right)^{-1/6} \Big/ \left(\frac{\tau_{D0}}{\tau_D} - 1\right)^{-1/6} - 1 = 7.9\ \% \qquad (29)$$

For k=100 molecule events both distance errors fall below 1 % and thus become negligible. The considerations for k=1 are, however, important for multiple molecule states (see below). Besides this statistic uncertainty, another error occurs during fitting. Here, the background has much more impact on the asymmetric lifetime distribution than on the efficiency distribution that is symmetric in most cases.

More relevant is the case of multiple states. In most cases, at least two states are present, one molecule state and one donor-only state. When assuming a donor-only state with E=0 and τ_D=4 ns and a molecule state with E=0.25 and τ_D =3 ns, each with n=50 photons, the efficiency distributions are still separable. But, the lifetime distributions are not separable at the single-molecule level, because the standard error of a sample of exponential random variables (n=50, k=1) is larger than the difference of the average lifetimes. When summing up several events (k>1),

separability and accuracy increase for both parameters; however, there are some practical limitations for the lifetime parameter. Fitting simulated data with two different expected lifetimes (3 ns and 4 ns) with k=5000 events and n=100 photons and 1 % background photons, yields average lifetimes with 10 % deviations from the expected values (Figure 73). These considerations show that the efficiency parameter is better suited than the lifetime parameter for single-molecule detection.

There is one rare case, in which the lifetime parameter might advantage over the efficiency parameter. That is, when the molecule has two states that are interchanging on a timescale that is shorter than the binning time of the photon traces, which is usually about 1 ms. Such states cannot be separated by the efficiency parameter due to averaging. But, they can, in principle, be distinguished by the lifetime parameter, because the average fluorescence lifetime is usually smaller than 10 ns. Distinction is possible if and only if both states largely differ in their average lifetimes and if numerous molecule events are summed up. However, a huge drawback is that all dynamics that occur between nanoseconds and minutes are convoluted with each other. By deconvolution of the donor-only lifetime, dye linker dynamics, anisotropy fluctuations, background and so on, one might be able to distinguish the underlying molecule state(s), but probably with large ambiguity. In contrast, in case of efficiency observation, troublesome linker dynamics and anisotropy fluctuations are averaged out.

4.1.2 DNA test system

Double stranded DNA is a widely used molecular test system because of its stability, relatively large persistence length, possible synthetization with variable modifications and existence of accurate structural models. A model DNA with various FRET dye combinations was designed and then measured with different fluorescence based methods. Additionally, sets of samples were sent to numerous independent laboratories for performing independent measurements based on different protocols (the results of this worldwide comparison project will be published elsewhere). Atto550 is chosen as the donor dye and is attached either at position A or B that are either 15 or 23 base pairs away from the position to which the acceptor Atto647N (1) or Alexa647 (2) is bound, resulting in the samples 1A, 1B, 2A and 2B. The dyes are labeled via C2-amino-linker to a thymidine that is known to cause the least fluorescence quenching among all nucleotides. In contrast to Alexa647, the other dyes feature comparably large lifetimes of 4 ns, relatively small molecular sizes and dipole orientations that are perpendicular to the linker, assuring fast rotational relaxation compared to the transfer rate (Figure 16).

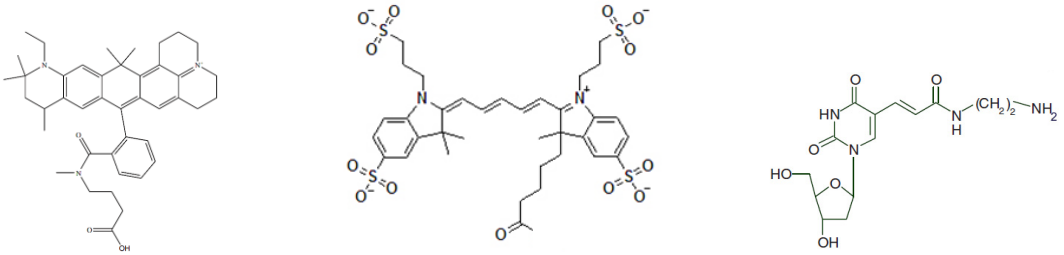


Figure 16: Structural formulas of Atto647N (left) and Alexa647 (middle) and the dye-linker synthetized to thymidine (right).

Model distances

To compare the measured FRET efficiencies and distances with real distances, model distances are generated reflecting the real distances as closely as possible. The model for double-stranded B-DNA is generated with 3D-DART. Each dye, donor and acceptor, is modeled as an ellipsoid which is bound via a cylindrical linker to certain positions of the model DNA (*41*). Dye accessible volumes assuming spatially equally distributed positions are generated for both dyes (Figure 17). Next, the average distance from all possible combinations of inter-dye distances is calculated. Notably, this average distance must not be the same as the distance between the centers of mass of the accessible volumes. In order to estimate the error of the approximated center of mass distance, Monte Carlo simulations are performed for two normally distributed position uncertainties in 3D (Figure 18) at different distances. They show that the Euclidean distance distribution converges to a normal distribution, if the mean distance is about three times larger than the standard deviations. I.e. with a dye-linker length of 1.5 nm, the mean inter-dye distance has to be at least 4.5 nm for negligible deviations between the ensemble average distance and the center of mass distances.

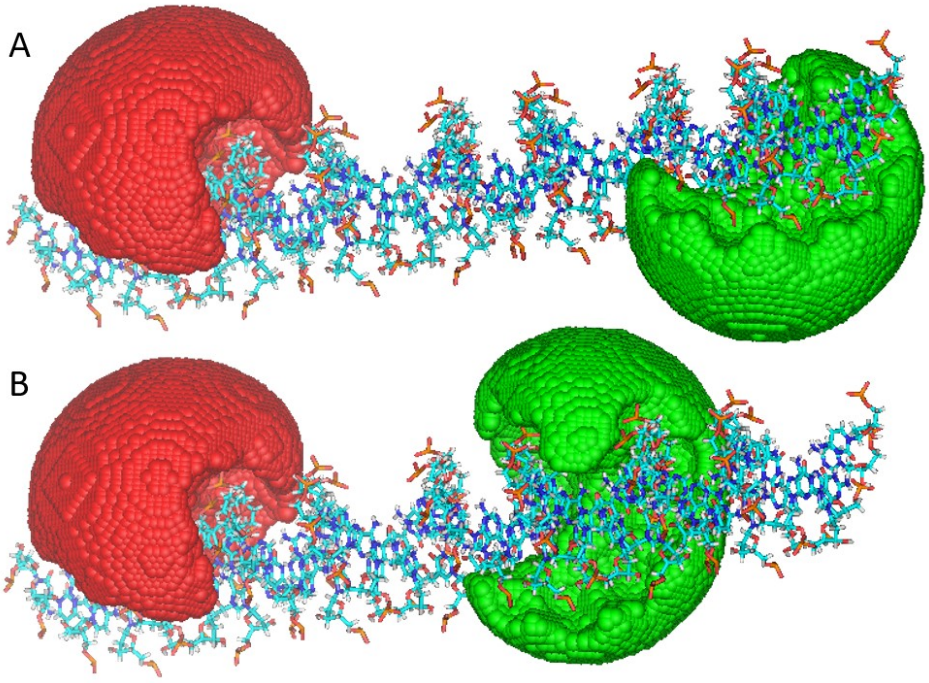


Figure 17: Accessible volumes of donor and acceptor dye – Atto550 (green) and Atto6347N (red) (Alexa647 is very similar) – that are bound to thymines of the depicted double stranded B-DNA and separated with either 23 (A) or 15 (B) base pairs.

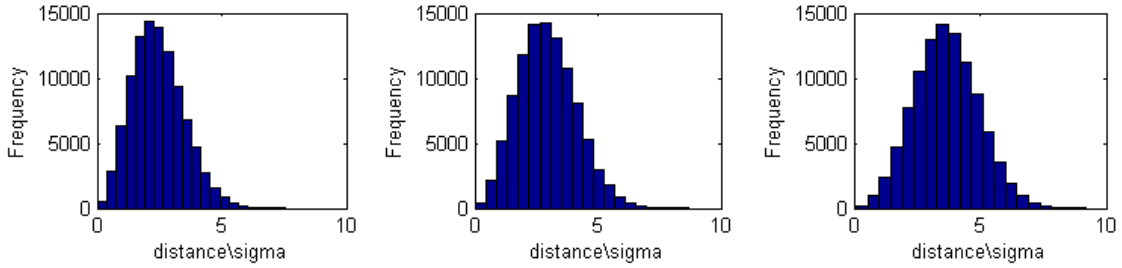


Figure 18: Monte Carlo simulations of two normal distributed position uncertainties in 3D with different relations between mean distance and standard deviation μ/σ = 1 (left), 2 (middle) and 3(right).

Timescales of averaging

The timescale of the dye motions has to be considered, too. If the dye motions are faster than the energy transfer rate, the transfer rates are probably averaged. If the dye motions are slower than the energy transfer rate, the transfer efficiencies from all possible combinations of inter-dye distances are averaged. The rotational relaxation time of most dyes is usually faster than the average transfer rate. But, the translational motion depends on the dye accessible volume and can be slower. If the dye linker is larger than the dye radius, the sampled space is usually larger than the dye itself.

The time-resolved anisotropies of the acceptor dyes reveal two exponential decays, respectively, a fast one (< 1 ns) representing the rotational diffusion of the dye and a slow one (> 3 ns) probably representing the translational motion of the dye within its accessible volume (Figure 19). Therefore, FRET efficiency averaging is used to determine the model distances. Notably, the error due to distance averaging instead of efficiency averaging is smaller than 5 % for distances larger than 60 Å.

Given that the dyes are internally labeled to double-stranded DNA, the probability for sticking of the dyes to the DNA is reduced to a minimum and the accessible volumes are still large. The time-resolved anisotropies indicate fast isotropic movement of all dyes. This guarantees sufficient dynamic averaging of relative dipole orientation, so that the orientation factor can be estimated at 2/3. See next Chapter for a detailed anisotropy discussion.

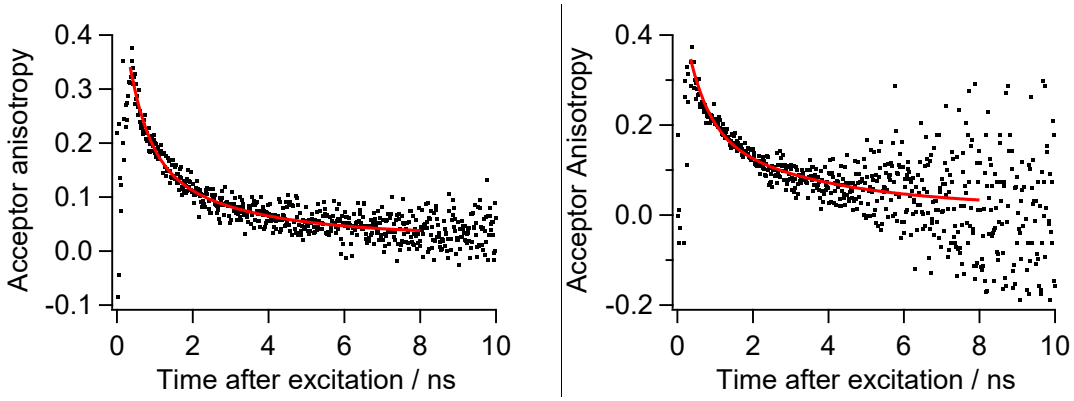


Figure 19: Time-resolved acceptor anisotropies for Atto647N (left) and Alexa647 (right). The fit coefficients are r_1=0.17±0.03, τ_1=0.51±0.1 ns, r_2=0.14±0.02, τ_2=2.6±0.6 ns, r_∞=0.03±0.01 (left) and r_1=0.16±0.03, τ_1=0.54±0.2 ns, r_2=0.17±0.04, τ_2=3.2±0.6 ns, r_∞=0.02±0.02 (right).

Measured distances

The measured FRET efficiency distributions were fitted with the distance distribution analysis derived in Chapter 4.2 (Figure 20). For each sample, an individual Förster radius R_0 is calculated with equation (6) considering donor quantum efficiency, spectral overlap and extinction coefficient. The resulting average FRET efficiencies and efficiency-averaged distances are compared with other parameters in Table 1. The gamma factor is determined in two ways, either locally by considering the efficiency-stoichiometry relation separately for each population or globally by considering the efficiency-stoichiometry relation simultaneously for all FRET efficiency populations. The local gamma factor depends only on the dye characteristics of one molecule state. It can be determined only if the confocal volumes sizes are adapted to each other and only if the efficiency broadening is dominated by distance fluctuations. The global gamma factor depends on the dye characteristics of all molecule states and can be determined only if the quantum efficiencies and extinction coefficients of donor and acceptor are state-independent. The comparison between local and global gamma factors can be used as a quality criterion as discussed below. Table 1 compares the measured FRET efficiencies and distances for locally (l) and globally (g) fitted gamma factors with the respective model values.

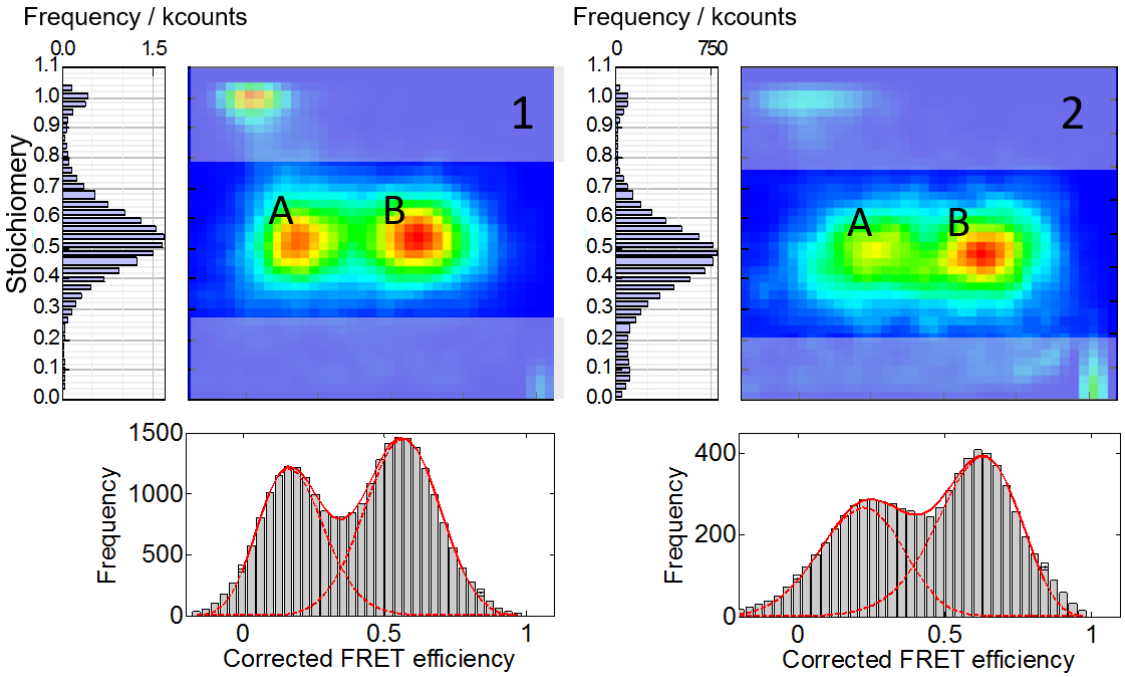


Figure 20: Stoichiometries versus corrected FRET efficiencies for mixed DNA-species with either 15 (A) or 23 (B) base pair separation, composed of Atto550 and Atto647N (1) or Atto550 and Alexa647 (2), respectively. The relation between stoichiometry and efficiency is used to determine the correction parameters *c*, *d*, *γ* and *β* (see Chapter 3.2.4). The corrected FRET efficiencies were then fitted with the distance distribution analysis derived in Chapter 4.2 (red line).

Sample	FRET-efficiency <E>	R_0/Å	Measured R_{DA}/Å	Model <E_{DA}>	Model R_{DA}/Å	Distance deviation in %
1A) 23b Atto647N	0.17(g) 0.15(l)	64	84(g) 86(l)	0.153	85.1	**2±3(g) 1±3(l)**
1B) 15b Atto647N	0.56(g) 0.57(l)	64	62(g) 61(l)	0.596	60.0	**3±2(g) 2±2(l)**
2A) 23b Alexa647	0.22(g) 0.12(l)	71	88(g) 99(l)	0.245	85.6	**2±3(g) 16±6(l)**
2B) 15b Alexa647	0.63(g) 0.53(l)	71	65(g) 70(l)	0.717	60.8	**7±3(g) 14±3(l)**

Table 1: Comparison of FRET efficiencies and distances with model distances for all four DNA samples. The donor dye is Atto550 for all samples. The FRET efficiencies are corrected with a local (l) or global (g) gamma factor. The Förster radius R_0 is calculated with sample-specific parameters. The model distances are calculated from FRET averaged inter-dye distance distributions considering the dye accessible volumes. The standard error (right hand side of ±) of the systematic deviation (left hand side of ±) between model and measured distance is propagated from at least three measurements.

Discussion

The measured distances for the Atto647N samples, 1A and 1B, agree perfectly with the model distances if averaging of the FRET efficiencies is considered. For both samples, the local gamma factors slightly deviate from the global ones and lead to slightly better results. Differences in the confocal volumes or differences in the extinction coefficients would lead to local gamma factors that both deviate in the same direction from the value obtained with a global gamma factor. As this is not the case, there must be other non-detectable local heterogeneities for the donor dyes, because the acceptor dyes are located at identical positions. The donor lifetimes might reveal differences in local quantum yields. But, Figure 21 shows that they are nearly identical for both samples in absence of the acceptor. Also, the correlation between donor lifetimes and FRET efficiencies does not show local quenching of the donor (Figure 22). The fraction of triplet quenching, however, is slightly larger for the donor in sample 1A (4 %) than for the donor in sample 1B (2 %) and could explain the deviating local gamma factor for sample 1A, because intersystem crossing competes with FRET. Nevertheless, the final deviations from the model distances for sample 1A and 1B with local or global gamma factor correction are negligible within an error of 3 %.

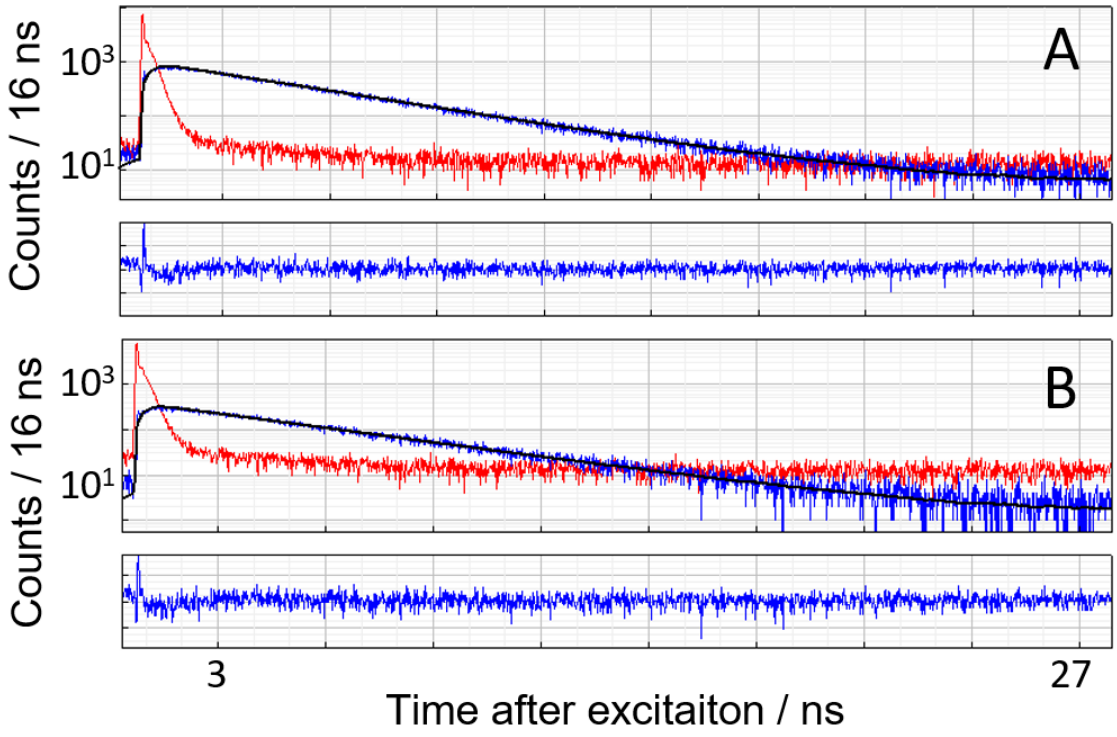


Figure 21: Time correlated photon counting histogram for donor fluorescence after donor excitation of double stranded DNA without acceptor for sample 1A (A) and 1B (B). The instrument response function (red curve) is convoluted with the fit function. The fitted donor lifetimes are 3.82 ns (A) and 3.79 ns (B).

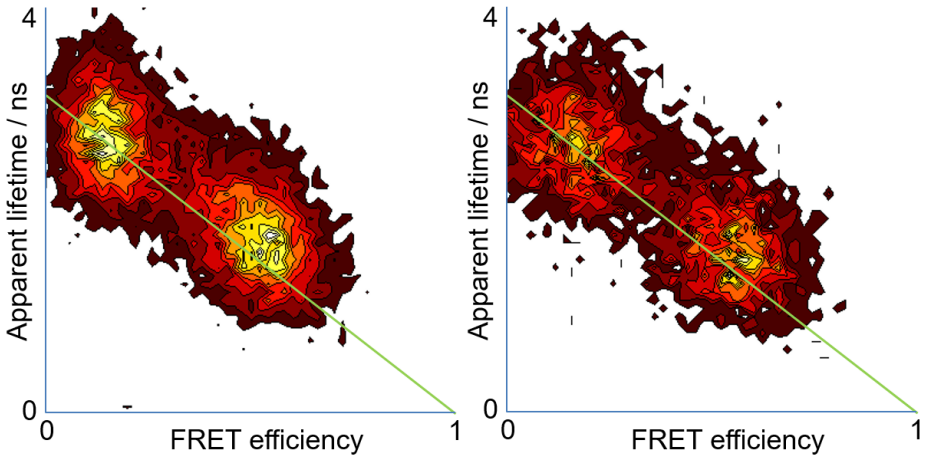


Figure 22: Correlating apparent donor lifetimes and FRET efficiencies for each single-molecule event for samples 1A and 1B (left) and samples 2A and 2B (right).

The critical problem cases are the samples 2A and 2B, where only the Atto647N dye is replaced by Alexa647 when comparing to samples 1A and 1B. For sample 2A the local gamma factor correction leads to strong deviations from the model distance. For 2B the measured distances strongly deviate in the same direction for both local and global gamma factor correction. The Förster radius has been calculated individually for each sample and fits with the literature value. Different extinction coefficients for the donor dyes are unlikely, because sample 1A and 1B have equal donor positions and their results agree with each other within 3 %. Sample dynamics such as bending of the DNA should be averaged, while wrong averaging would cause equal deviations for all samples. Other labs measured the same deviation for this sample with different setups and methods.

All these considerations lead to the conclusion that the wrongdoer must be the acceptor dye. It turned out that for Alexa647 there is a 10 times larger fraction of molecules being in a dark state than for Atto647N (Figure 23). These dark states can be caused by cis-trans isomerization or intersystem crossing (electron being in the triplet state). In the following the latter case is assumed. As the rates into the triplet state and from the triplet state back (also called blinking rates) are in the range of the photon count rate (1-10 µs), consecutive photons are correlated and the apparent acceptor intensity decreases. If the acceptor intensity was near saturation, it would not increase linearly with transfer efficiency. This could explain why the globally corrected sample 2A agrees with the model distance and sample 2B does not. Another reason for the non-linear acceptor intensity increase could be an increasing blinking probability with increasing FRET rates, indicated by a larger triplet state fraction for sample 2B than for sample 2A (black and green curves in Figure 23 right, respectively). In both cases a global gamma factor cannot correct for blinking, because the apparent saturation effect depends on the FRET rate. The introduction of a novel correction factor becomes necessary (see Chapter 4.1.3).

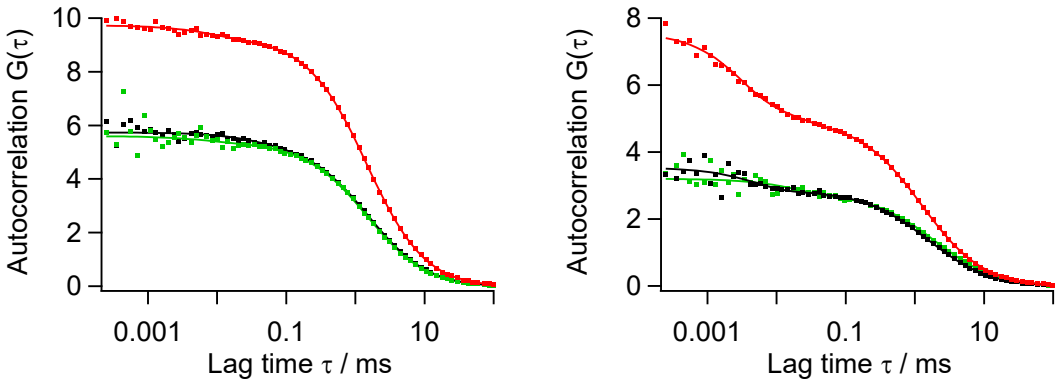


Figure 23: Left: Autocorrelations of acceptor (Atto647N) photons after donor (Atto550) excitation for sample 1A (green) and sample 1B (black) and autocorrelation of acceptor photons after acceptor excitation (red). Right: Autocorrelations of acceptor (Alexa647) photons after donor (Atto550) excitation for sample 2A (green) and sample 2B (black) and autocorrelation of acceptor photons after acceptor excitation (red). The autocorrelations were fitted with diffusion and triplet terms (see Chapter 3.2.1). The fits for acceptor excitation revealed a triplet fraction of T=3 % and a correlation time of τ=3 ms for the Atto647N sample and T=33 % and τ= 7 ms for the Alexa647 sample.

In contrast to intensity based transfer efficiencies, the arrival times of the donor photons are not averaged over several photons. Fast switching states like the acceptor being in the triplet state, cannot be distinguished by the transfer efficiencies, but by the lifetime distribution. Indeed, Figure 24 clearly reveals two donor lifetimes for the selected FRET population of sample 2B. The smaller lifetime probably represents the normal FRET state and the larger unquenched lifetime probably represents the dark state of the acceptor which influences also the donor. However, the smaller lifetime does not fit to the expected lifetime of 1.1 ns calculated from the respective efficiency. One reason for this might be that lifetimes are not only superimposed with blinking rates, but also with other comparably fast processes such as slow dye motions. On the other side, it is statistically nearly impossible to distinguish different expected lifetimes, especially lifetime distributions. Notably, for all samples the distances calculated via lifetime are underestimated.

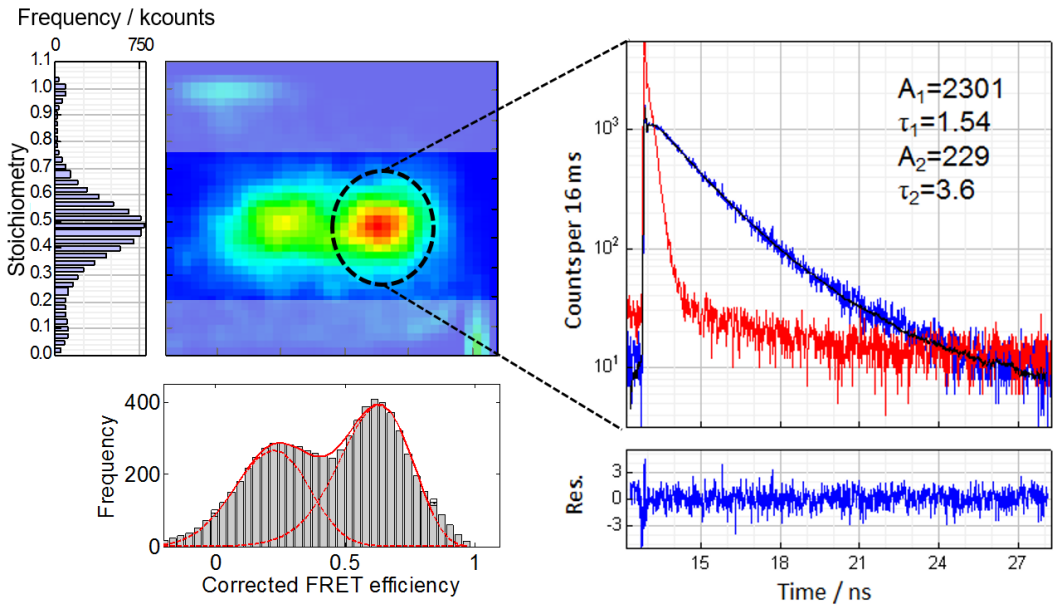


Figure 24: Donor lifetimes were selected from all photon bursts that belong to the high efficiency population of sample 2B (left) and were fitted with a two-exponential function (right) considering the instrument response function (red).

It turned out that disagreement of local and global gamma factors can be caused by many artifacts such as mismatch of confocal volumes, position dependent dye heterogeneities (such as locally different quantum yields or extinction coefficients) or large triplet fractions. This fact can be utilized as a quality criterion: Local and global gamma-factors differing by more than 3 % are questionable and require further state-dependent analysis of lifetimes, anisotropies and fluorescence correlations.

4.1.3 Introduction of a triplet disturbance factor

In the last section, a simple criterion (comparison of local and global gamma factor) was derived that indicates intensity fluctuations of the dyes due to so-called dark states. It is assumed that triplet blinking (intersystem crossing into the triplet state) is the cause of these dark states. As triplet blinking of the acceptor dye impacts the FRET efficiency, a correction factor is introduced in the following.

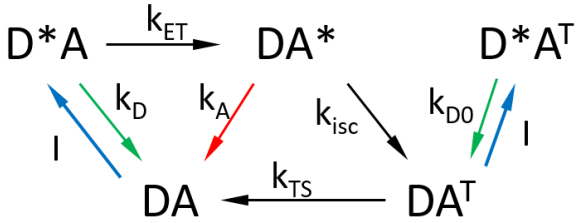


Figure 25: Schematics of the transfer processes illustrating the state D^*A^T in which the donor is excited in the singlet state, but energy transfer to the acceptor that is excited in the triplet state is very unlikely.

When the acceptor electron is in the triplet state, the molecule is "invisible" for a certain amount of time. When the donor gets excited to the singlet state during that time, energy transfer to the acceptor is very unlikely, leading to reduction of the average transfer efficiency (see Figure 25). The question has to be raised on which timescale the blinking rate occurs. Triplet rates that are slower than the diffusion time of about 1 ms end up in the donor-only population of the stoichiometry-efficiency scatter plot and can be ignored. Triplet rates between the photon count rate (about 100 ms^{-1}) and diffusion time will completely affect the average transfer efficiency. And triplet rates faster than the photon count rate, but slower than the laser excitation rate will partly affect the average transfer efficiency. Finally, triplet rates that are faster than the laser excitation rate, but slower than the transfer rate, should be negligible, because the triplet excitation relaxes before the next potential energy transfer can occur.

The beta factor $\beta=\sigma_A I_A/(\sigma_D I_D)$ corrects for the relative absorption cross sections of the dyes and the relative laser intensities and is gained by correlating FRET efficiencies and stoichiometries (see Figure 20 and Chapter 3.2.4). Here, a third factor p_r representing the probability to be in the relaxed state (i.e. not to be in the triplet or singlet state) is added to the beta factor resulting in $\beta=\sigma_A I_A p_{rA}/(\sigma_D I_D p_{rD})$. p_{rA} for the acceptor can be obtained by dividing the global beta factor β of samples 2A and 2B by the global beta factor β^* of the reference sample 1A and 2B. As the laser intensities are equal for all measurements, all intensity factors eliminate each other. Since the donor dyes and their positions are exactly identical, their cross sections and probabilities p_{rD} also vanish. Only the ratio of the absorption cross-sections of the acceptors (equaling the ratio of their extinction coefficients) and the probabilities p_{rA} remain:

$$p_{rA} = \frac{\beta \varepsilon_A^*}{\beta^* p_{rA}^* \varepsilon_A} \qquad (30)$$

The probability for the acceptor to be not excited in the triplet state is p_{rA}=65% and agrees with the fraction of molecules in the triplet state $T=1-p_{rA}$=33% that is complementary obtained by fitting the fluorescence autocorrelation of the acceptor (red curve in Figure 23 right) with equation (17). The triplet de-excitation rate k_{TS}=100 ms^{-1} is obtained by the same fit and is faster than the average photon count rate k_{cr}=40 ms^{-1}, but slower than the laser excitation rate and thus partly affects the average transfer efficiency.

The ratio between the average photon count rate k_{cr} and the sum of the latter and the triplet-singlet de-excitation rate k_{TS} accounts for the probability of the acceptor electron still being trapped in the triplet state, when the donor is excited and a photon is detected. This probability multiplied with the fraction of acceptor molecules being in the triplet state, is now introduced as the triplet disturbance factor c_T:

$$c_T = \frac{1 - p_{rA}}{1 + \frac{k_{TS}}{k_{cr}}} \qquad (31)$$

The corrected FRET efficiency without triplet disturbance E^T can be obtained from the apparent FRET efficiency E in the following manner:

$$E^T = \frac{E}{(1 - c_T)} \qquad (32)$$

This correction is applied to the globally gamma-corrected values in Table 1 and finally results in an excellent agreement with the model distances (Table 2). Remarkably, the triplet disturbance factor c_T perfectly agrees with the ratio of the donor lifetime amplitudes A_D/A_{D0}=0.1 for sample 2B (see Figure 24).

Sample	**<E>**	**p_{rA}**	**$1+k_t/k_c$**	**c_T**	**R_0**	**<E^T>**	**R^T/Å**	**<E_{DA}>**	**R_{DA}/Å**	**Deviation in %**
1A) 23b	0.17	0.97	3.5	0.01	64	0.17	84	0.153	85.1	2±3
1B) 15b	0.56	0.97	3.5	0.01	64	0.57	61	0.596	60.0	2±2
2A) 23b	0.22	0.65	3.5	0.1	71	0.24	86	0.245	85.6	1±4
2B) 15b	0.63	0.65	3.5	0.1	71	0.71	61	0.717	60.8	1±2

Table 2: Correction for triplet disturbance and comparison of corrected expected FRET efficiencies and corrected distances with model distances.

Triplet rates faster than the transfer rate are very unlikely. But, one could think of other effects that lead to a modified apparent extinction coefficient of the acceptor such as cis-trans isomerization, which in turn proportionally modifies the transfer rate by the factor p_ε. In this case, one could either define the corrected Förster radius R_0^ε:

$$R_0{}^{\varepsilon} = R_0 \cdot p_\varepsilon{}^{1/6} \qquad (33)$$

or the corrected efficiency E^ε:

$$E^{\varepsilon} = \frac{1}{1 + p_\varepsilon(1/E - 1)} \qquad (34)$$

Finally, the local gamma factors of the samples 2A and 2B are likely modified by the triplet disturbance. Also, the broadening of the FRET efficiencies seems to be dominated by triplet disturbance. As a result, the correlation between stoichiometry and efficiency is not dominated by quantum yields any more. So, the usage of a global gamma factor is recommended, and the local quantum yields should be checked separately, e.g. by correlating the acceptor and donor

lifetimes with efficiencies. Notably, the comparison of the local and global gamma factors can be used to distinguish, whether the efficiency broadening exceeding shot noise arises from distance fluctuations or photophysical fluctuations. A local gamma factor that is larger than the global one, possibly originates from additional phosphorescence photons after direct excitation of the acceptor's triplet state. Conclusively, the matching results imply that the modeling of the dye accessible volumes and their homogeneous position density sufficiently reflects real dye motions.

4.1.4 Protein test system

The rules derived for the DNA test system apply also to proteins. In addition, the dye mobility needs to be checked, because of a more inhomogeneous protein surface and unpredictable steric restriction upon conformational change.

An average orientation factor of $<\kappa^2>=2/3$ is correct if the relative dye dipole orientations sample all possible combinations on a timescale that is faster than the energy transfer. However, depending on dye characteristics and environment, the dye motions are often restricted on different timescales. The fluorescence anisotropy of donor and acceptor can be used as a measure to estimate possible relative dye dipole orientations. Although the largest possible inter-dye distance error due to restricted dye motion was theoretically calculated to be about ±23 % for an anisotropy of 0.25 (*86*), realistic worst case errors can be much smaller depending on the dyes used. The dyes used in this work, Atto550 and Atto647N, have a polarization axis perpendicular to the linker, allowing faster isotropic motion than a parallel oriented axis. For reliable distance determination, an experimental value is needed that represents sufficient dynamic averaging of the dye dipole orientations. To this end, within this work, a large data set of distances and time-resolved anisotropies was measured that covers different dyes, different protein environments, and different dye-separations and was compared with the x-ray structure of Hsp90.

Time-resolved anisotropy measurements are used to analyze the timescales of dye motions. In (*41*) local motions of the dye ρ_{local} and global motions of the protein ρ_{global} are superimposed:

$$r(t) = \left((r_0 - r_\infty)e^{-t/\rho_{local}} + r_\infty\right) e^{-t/\rho_{global}} \qquad (35)$$

However, the rotational relaxation time of Hsp90 is estimated at 1 µs with equation (15), and is therefore negligible for the detectable anisotropy decay. Indeed, some dye positions at Hsp90 don't reveal an anisotropy decrease beyond 2 ns, but others do (Figure 26 A and B, respectively), confirming that the slower detectable anisotropy decay mainly arises from dye motions. As the protein motion is equal in both cases, the slow decay must originate from slow translational dye diffusion, meaning that the dye slowly samples the protein surface at the timescale of several nanoseconds. The following equation accounts for rotational and translational dye motion and is used for all measurements:

$$r(t) = (r_0 - r_1 - r_\infty)e^{-t/\tau_{c1}} + r_1 e^{-t/\tau_{c2}} + r_\infty \qquad (36)$$

Here, $r_2=r_0-r_1-r_\infty$ is the fraction for rotational diffusion and r_1 the fraction for translational diffusion. The fundamental anisotropy r_0 is estimated at 0.38. Notably, the rotational correlation times were slower for the DNA samples than for all Hsp90 positions, hinting towards charge-charge interactions between dye and DNA.

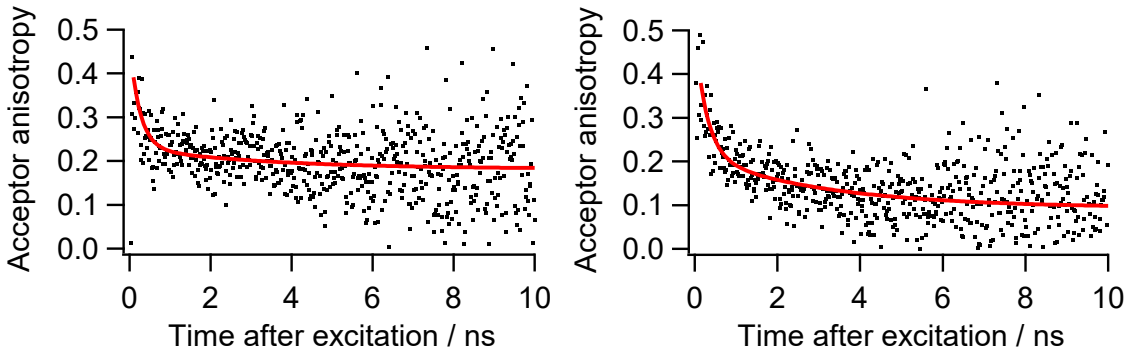


Figure 26: Two examples for time-resolved acceptor anisotropies after direct excitation of the acceptor located at different positions in Hsp90. The rotational correlation times are τ_{c1} =0.25 ns (left) and τ_{c1} =0.29 ns (right). The left measurement revealed a second significant correlation time of τ_c=3 ns. The residual anisotropies at the mean transfer time $r(t=\tau_T \approx 2\ ns)$ are 0.2 (left) and 0.15 (right) and the remaining residual anisotropies $r(t\rightarrow\infty)$ are 0.18 (left) and 0.1 (right).

In the following, the question will be addressed, whether the anisotropy value at transfer time $r(t=\tau_T)$ or the remaining anisotropy $r(t\rightarrow\infty)$ serves as a better criterion for evaluating the measured distances. The anisotropy at the time of energy transfer considers dynamic averaging of the dipole orientation factor. In contrast, the residual anisotropy beyond transfer time considers further slow dye reorientations, which lead to a distribution of transfer rates. The question arises how large the error is when partially averaging over the remaining distribution of transfer rates, although averaging over orientation factors is assumed with $\langle\kappa^2\rangle=2/3$.

As the dipole axes of the chosen dyes – Atto550 and Atto647N – are oriented perpendicular to the linker axis, fast rotational diffusion is facilitated despite restrictions from the protein surface. To estimate the errors for these dyes, numeric calculations were performed assuming depolarizations of 10 % and restricted dye-linkers with perpendicular dipole axes (resulting in fast depolarizations within a plane). The calculated transfer rates deviated from the assumed distances with an error of about 20 %. Remarkably, in this case, slow translational dye motion reduces the error, although the transfer rates and not the orientation factors are partially averaged. In case of sufficient dynamic averaging of dipole orientations (due to fast rotational relaxation), slower dipole reorientations yield very small deviations between assumed averaging over orientation factors and real averaging over transfer rates (for $\kappa^2=2/3$):

$$\sqrt[6]{\langle\kappa^2(1\pm0.5)\rangle} \approx \langle\sqrt[6]{\kappa^2(1\pm0.5)}\rangle \qquad (37)$$

These findings fit well with the total of measured distances. The geometric mean anisotropies at different times $r(t=\tau_T)$ and $r(t\rightarrow\infty)$ were calculated for each measured FRET pair and then compared with the deviation between the measured distance and the model distance obtained from the x-ray structure of Hsp90's closed conformation.

The geometric mean of the acceptor anisotropy r_A and donor anisotropy r_D turned out to be a representative value for the transfer anisotropy, because an isotropic dye can to a large extent compensate for an increased anisotropy of the other dye:

$$r = \sqrt{r_A}\sqrt{r_D} \qquad (38)$$

Notably, this observation fits to the Soleillet theorem (*87*) that describes the transfer anisotropy as products of depolarization factors.

Then, an upper threshold for the geometric mean anisotropy was varied, and for each threshold, the root mean square deviation (RMSD) for the remaining distances was calculated. Interestingly, the resulting RMSD curves are similar for both threshold criteria, $r(t=\tau_T)$ (Figure 27 A) and $r(t\rightarrow\infty)$ (Figure 27 B) indicating predominant anisotropy decay before energy transfer and thus sufficient dynamic averaging. Notably, when using $r(t\rightarrow\infty)$ as a threshold criterion, a smaller amount of good distances (i.e. distances with small deviations from the x-ray structure) is rejected than when using $r(t=\tau_T)$. This means that slow dye motions which occur in addition to sufficient fast isotropic motions normally lead to additional convergence to $\langle\kappa^2\rangle=2/3$, and thereby, to a more accurate distance determination. That is why the geometric mean of the residual dye anisotropies $r(t\rightarrow\infty)$ has been chosen as selection criterion.

Sufficient fast isotropic motion was given in all cases in which the residual anisotropy was below the threshold value of 0.21. Labeling positions with a large anisotropy at the time of energy transfer had also a large residual anisotropy. This is probably because a hindered rotational motion of the dye is usually caused by a hindered translational motion of the dye due to a small surface accessible volume.

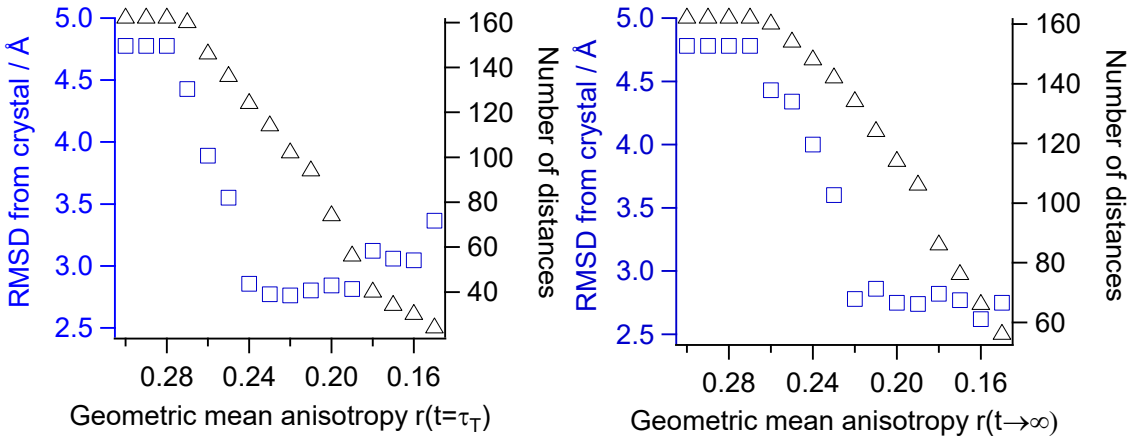


Figure 27: RMSD from x-ray structure and remaining number of distances versus upper threshold for *r(t=τT)* (left) and *r(t→∞)* (right).

Figure 28 shows that nearly all distances that strongly deviate from the crystal structure (>8 Å) belong to dyes that have a high mean anisotropy. An upper threshold of 0.21 for the geometric mean anisotropy perfectly rejects these troublemakers (red squares). An acceptor anisotropy that is much lower than the anisotropy of other dyes measured at the same positon, is an indicator for back transfer from acceptor to donor or for strong coupling due to too small dye separations (< 40 Å) and usually leads to a wrong distance (triangles). Furthermore, distances smaller than 45 Å have such a small transfer dwell time that the relative dipole orientations are not sufficiently dynamically averaged. The remaining distances (green circles) deviate with an RMSD of 2.8 Å from the respective x-ray structure. Badly resolved or dynamic structural elements at the protein surface lead to inaccurate prediction of the dye accessible volumes, especially if the calculated accessible volumes are very small. Besides, the volume sizes could be uses as an additional selection criterion. The outliers with 5-8 Å deviation are located mainly at the C-domains, where the x-ray structure is less well resolved. They might be caused by fast dissociation events at the CM-interface. Distinct spectroscopic features such as a larger fraction of molecules in the triplet state were not found for these outliers.

Importantly, for the presented approach, the lifetimes should be large compared to the rotational relaxation time and the acceptor lifetime should be at least as large as the donor lifetime.

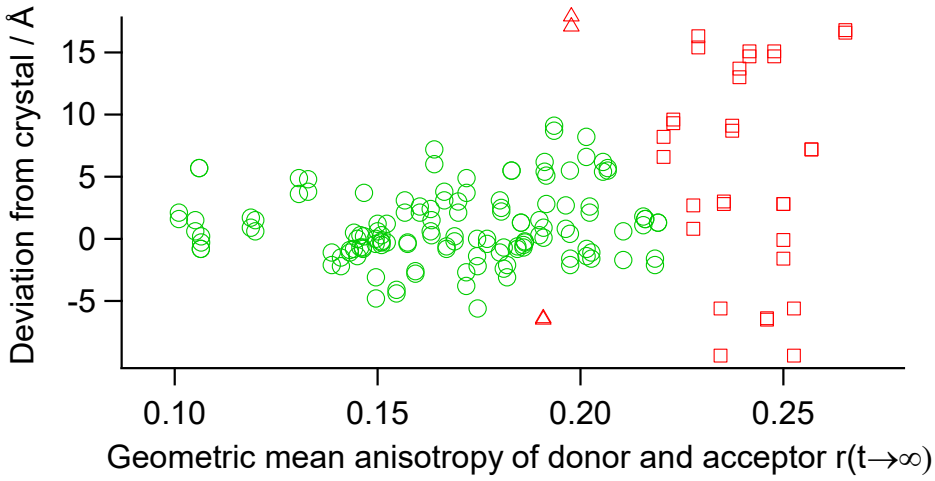


Figure 28: Deviation of the measured distances from the x-ray crystal structure 2cg9 versus the geometric mean of the donor and acceptor anisotropy respecting $r(t\rightarrow\infty)$. 81% of all measurements are under the chosen threshold $r_{TH}(t\rightarrow\infty)$=0.22 (green circles); 17% are above the threshold (red squares); and 2% are under the threshold, but were rejected by other criteria than anisotropy (red triangles).

As an alternative, anisotropy independent procedure, one can simply swap donor and acceptor positions to change the environment of each dye and then check both distances for consistency. In such a case, other dyes with greater differences in the geometries of the polarization axes would be helpful.

In contrast to the presented approach, some authors use the FRET sensitized acceptor anisotropy to further restrict the transfer rate errors or to even predict transfer rate distributions. However, the data presented here does not show any correlation between FRET sensitized anisotropy and distance deviations. More than 95 % of all dye positions possess a very low FRET sensitized acceptor anisotropy.

In conclusion, the geometric mean of the residual dye anisotropies can be used as a selection criterion for proteins, provided that small dyes (as Atto550 or Atto647N) are used with a relatively large linker (>15 Å), a dipole axis that is perpendicular to the linker and a large lifetime (>3 ns). The threshold value of 0.21 for the geometric mean anisotropy of both dyes is proposed to remain valid for other proteins under the above conditions. Deviating low acceptor anisotropy is an indicator for too small dye separations and should be additionally used as a criterion.

4.2 Distance distribution and fluctuation analysis

In the first part, a probability distribution analysis has been derived in order to separate FRET based distance distributions from shot noise, while accounting for the manipulated statistics of pre-corrected photon traces. In the second part, an approach for analyzing the timescale of the distance fluctuations is presented.

In other publications, uncorrected FRET efficiency histograms were fitted with a probability distribution analysis considering photophysical parameters from separate measurements (*72, 88, 89*). In contrast, here, the binned photon traces were corrected for quantum efficiencies, relative detection efficiencies, relative extinction coefficients, crosstalk, direct excitation and donor- and acceptor-only populations. This is realized iteratively by correlating efficiencies and stoichiometries from corrected photon bursts similar to (*30*). Additionally, these parameters have been verified in a consistency network. Another advantage of pre-correcting photon traces is that the efficiency populations are equally weighted and independent from the intensity threshold necessary for burst searching. The excessive photons, for example from crosstalk, are then considered as a source of additional shot noise broadening.

Other sources of broadening have been considered, too. Background photons have been neglected having a signal to noise ratio of at least 50. Fluctuations of the dipole orientation factor κ^2 or fast distance fluctuations due to translational diffusion of the dyes have been ignored, because they are many orders of magnitudes faster than the photon count rate. Blinking and bleaching of the dyes were reduced by fast laser scanning and by using stable dyes. These artifacts have been additionally checked by comparing local and global gamma factors.

Furthermore, it is demonstrated that the efficiency distribution of an average value of the measured intensity distribution perfectly agrees with the efficiency distribution of a distribution of intensities. This approximation allows fast maximum likelihood optimization and thus enables a global analysis of numerous FRET efficiency histograms.

4.2.1 Analysis for corrected transfer efficiencies

The photon counts per channel and time bin are Poisson distributed. If consecutive photons are not correlated, the shot noise limited FRET efficiency distribution is a combination of independent Poisson processes $Z=X/(X+Y)$ and can be described by the joint probability (*90*):

$$P\left(E=\frac{N_A}{N_A+N_D}\middle|\langle E\rangle,n\right)=\frac{(\langle E\rangle n)^{N_A}}{N_A!}\frac{\left((1-\langle E\rangle)n\right)^{N_D}}{N_D!}e^{-n} \qquad (39)$$

Here, $\langle E\rangle$ is the expected FRET efficiency, $n=\langle N_A+N_D\rangle=\langle N_{AD}\rangle$ the expected number of total photons and N_D and N_A are the number of detected donor or acceptor photons, respectively. Considering every combination of N_A and N_D for different n is computationally extensive, especially when the distributions are convoluted with numerous distance distributions. Gopich and Szabo suggest an approximation either by a Gaussian or a beta distribution. However, Monte Carlo simulations (implemented in Matlab) show a considerable deviation, especially for low and high efficiencies (Figure 29 top left).

Another description of the shot noise limited FRET efficiency distribution is a binomial distribution with the probability ϵ of observing N_A acceptor photons from n total photons (*89*):

$$P_\epsilon(N_A|n) = \epsilon^{N_A}(1-\epsilon)^{n-N_A}\frac{n!}{N_A!\,(n-N_A)!} \qquad (40)$$

Interestingly, extending the joint distribution (39) with $N_{AD}!$ leads to the following equation:

$$P\left(E = \frac{N_A}{N_A+N_D}\middle|\langle E\rangle, n\right) = \langle E\rangle^{N_A}(1-\langle E\rangle)^{N_D}\frac{N_{AD}!}{N_A!\,N_D!}\cdot\frac{n^{N_{AD}}}{N_{AD}!}e^{-n} \qquad (41)$$

with a binomial term on the left side and a Poisson term on the right side. Now, with the restriction $n=N_{AD}$, the Poisson term is constant, so that equation (41) results in a binomial distribution. This restriction is not valid for low intensities, because in reality donor and acceptor intensities are independently Poisson distributed. Interestingly, Monte Carlo simulations show a perfect agreement between two independent Poisson distributed random variables and a binomial distribution when the mean intensity n is the same as or larger than the number of bins chosen for the efficiency histogram (see Figure 29 top left for $n=n_{bins}=30$). The binomial distribution is generated for the mean total intensity n and then interpolated to the number of FRET efficiency bins n_{bins} to overcome the discontinuity of the binomial distribution.

On the experimental side, a random number is added to each intensity value before calculating the efficiencies:

$$r = \left[-\frac{n_{bins}}{2n};\frac{n_{bins}}{2n}\right] \qquad (42)$$

That way, discretization artifacts (i.e. binning artifacts) due to rational numbers are overcome and an optimal bin number as supposed by (*72*) is not necessary any more. Also in the Monte Carlo simulations, random numbers r are added to the Poison distributed intensities, in order to be most closely to the experiment. They resulted in perfect agreement of the simulated efficiency histograms with the binomial distributions (Figure 29).

Now, additional acceptor photons due to spectral crosstalk c and direct excitation d of the acceptor from the green laser are considered. The expectation value of the sum of independent Poisson random variables equals the Poisson distribution of the sum of the expectation values: $n_A^*=n_A+n_c+n_d$. Thus, the corrected FRET sensitized acceptor intensities can be calculated with $N_A=N_A^*-n_c-n_d$. However, after subtraction of absolute values from Poisson distributed intensities, the mean and the variance have not the expected relation anymore. This discrepancy was overcome by deriving corrected probability density distributions. In short, expected corrected efficiencies $\langle E\rangle$ are transferred to uncorrected ones $\langle E^*\rangle$ followed by a coordinate transformation from E^* to E.

The uncorrected FRET sensitized acceptor intensity:

$$n_A^* = n(\langle E\rangle + d + (1 - \langle E\rangle)c/\gamma) \tag{43}$$

depends on the corrected acceptor intensity $n_A = n\langle E\rangle$, the correction factor $\gamma = Q_A\eta_A/(Q_D\eta_D)$ for different quantum efficiencies Q and detection efficiencies η, the correction factor c for cross talk and the correction factor d for direct excitation of the acceptor by the donor laser. With the donor intensity $n_D^* = n(1 - \langle E\rangle)/\gamma$ and $\langle E^*\rangle = n_A^*/(n_A^* + n_D^*)$ one obtains the expected uncorrected efficiency:

$$\langle E^*\rangle = \frac{n(\langle E\rangle(1 - c/\gamma) + d + c/\gamma)}{n(\langle E\rangle(1 - c/\gamma - 1/\gamma) + d + c/\gamma + 1/\gamma)} \tag{44}$$

and the expected uncorrected total intensity:

$$n^* = n(\langle E\rangle(1 - c/\gamma - 1/\gamma) + d + c/\gamma + 1/\gamma). \tag{45}$$

This leads to the probability density distribution of uncorrected shot noise limited FRET efficiencies, which considers the variances of the original uncorrected photon traces:

$$P_{E^*}(N_A^*|n^*) = \langle E^*\rangle^{N_A^*}(1 - \langle E^*\rangle)^{n^* - N_A^*}\frac{n^*!}{N_A^*!\,(n^* - N_A^*)!} \tag{46}$$

With the control variable N_A^* the distribution is directly transformed to the corrected efficiencies:

$$E(N_A^*) = \frac{N_A^* - n(d + (1 - \langle E\rangle)c/\gamma)}{n} \tag{47}$$

to finally obtain the corrected shot noise limited FRET efficiency probability density distribution:

$$P_{E(N_A^*)} = P_{E^*}(N_A^*) \tag{48}$$

Figure 29 top left, shows simulated shot noise limited efficiency distributions for $n=n_{bin}=30$, $\gamma=1$, $c=d=0$ and 5 different expected FRET efficiencies $\langle E\rangle$ (coloured bars). The binomial distribution (dashed line) perfectly agrees with the simulated data, while the beta distribution (points) shows clear deviations, especially for low and high efficiencies. To further approach the experiment, exponential distributed intensities n and a lower intensity threshold n_{th} are added to the simulations (Figure 29 top right). The burst intensities in a scanning confocal setup can be approximated with an exponential distribution (see Chapter 3.2.2). For an expected value of $<N>=21$ photon counts with a threshold $n_{th}=30$ counts, the total mean intensity becomes: $n=n_{th}+<N>=51$. Interestingly, the sum of several shot noise limited efficiency distributions from exponential distributed intensities equals a single shot noise distribution with the expected intensity n. Also, a binomial distribution with an expected number of $n=51$ photons interpolated

to n_{bin}=30 perfectly agrees to the simulated efficiency distributions (Figure 29 top right). Finally, the photons from cross talk and direct excitation were added to the simulation as additional Poisson random variables and the acceptor intensity n_A is increased to γn_A. Typical values for Atto550 and Atto647N and the setup used in this work were taken: c=d=0.1 and γ=1.2. The resulting shot noise limited efficiencies from uncorrected intensities are plotted in Figure 29 bottom left. The efficiencies are shifted to higher values and the fits with binomial distributions show less events for the low efficiencies than for the high efficiencies, because high efficiency events have more acceptor photons and thus exceed the threshold more often. Additionally, the uncorrected mean intensity n* and thus also the variance of the efficiency distribution depends on the expected efficiency itself. In contrast, the corrected efficiencies from corrected intensities are equally weighted on the complete efficiency range (Figure 29 bottom right). The broadening of the lower expected efficiencies due to previous subtraction of additional acceptor photons is clearly seen by comparing with Figure 29 top right.

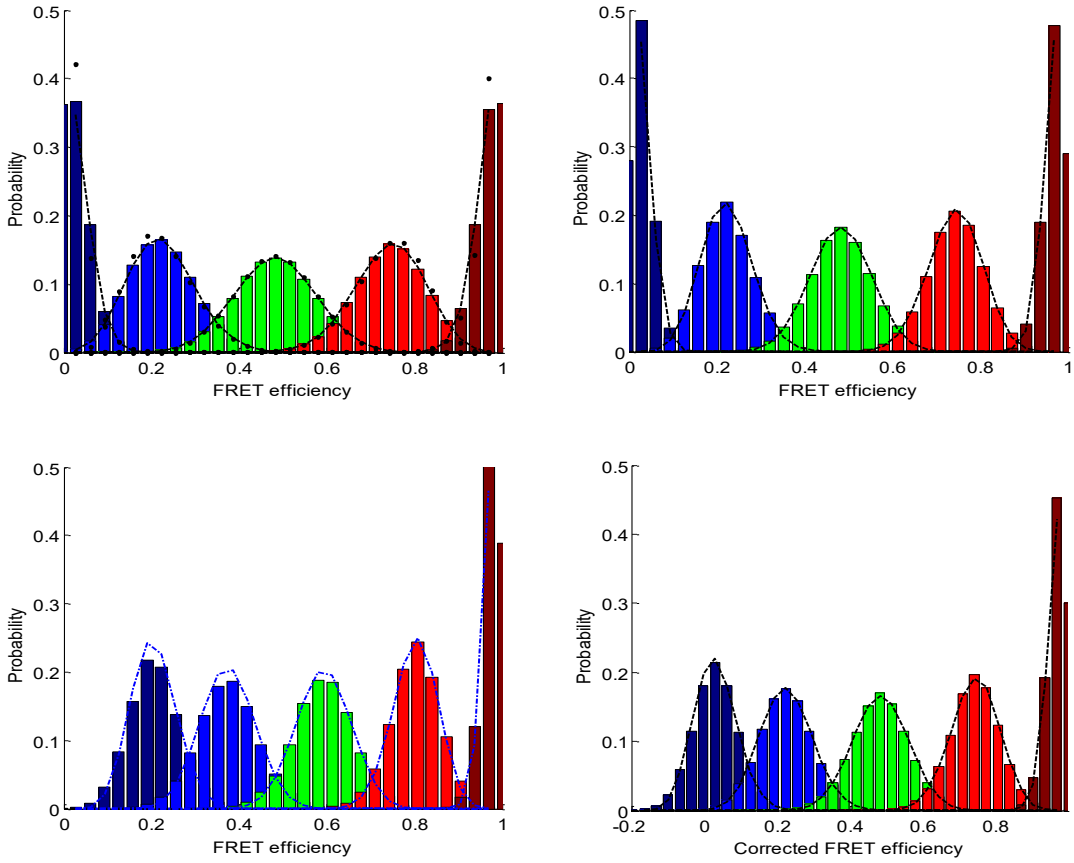


Figure 29: Monte Carlo simulations for five different expected FRET efficiencies (colored bars) and a total number of 100.000 photon bursts considering Poisson distributed fluorescence intensities. To minimize the formation of spikes from rational numbers, a very small random number was added to each intensity value. Top left: Distributions for an expected total intensity *n*=30 are perfectly described with binomial distributions (dashed line), but deviate from beta distributions (dots), especially for low and high efficiencies. Top

right: Distribution for exponential distributed total intensities with expectation values of 21 photons and a threshold of 30 photons yielding a total average number of 51 photons. Binomial distributions with *n*=51 photons (dashed lines) perfectly describe the simulated data. Bottom: Additional Poisson distributions were simulated for a gamma factor of 1.2, 10 % leakage and 10 % direct excitation. Bottom left: The uncorrected intensities lead to different weighting of the uncorrected efficiencies. Bottom right: Pre-corrected intensities (by subtraction with a constant number) lead to equal weighting of the corrected efficiencies. The corrected efficiency histograms are fitted with corrected binomial distributions that take into account the original intensity fluctuations (dashed lines).

To obtain the convolution between distance fluctuations and shot noise $P_E(R,n)$, the normalized corrected shot noise limited FRET efficiency distributions P_E are numerically superimposed with the distance probability distribution $P\big(\langle \mathrm{E} \rangle(\mathrm{R}_{DA})\big)$:

$$P_E(R,n) = \sum_{k=1}^{n} P\left(\int_{\langle E\rangle=(k-1)/n}^{k/n} \langle E\rangle(R_{DA}) d\langle E\rangle\right) P_E(E|n, \langle E\rangle = \frac{2k-1}{2n}) \qquad (49)$$

The integration over all distance probabilities per FRET efficiency bin $(k-1)/n < \langle E\rangle < k/n$ is especially important for low and high efficiencies:

$$\int_{\langle E\rangle=(k-1)/n}^{k/n} \langle E\rangle(R_{DA}) d\langle E\rangle = \frac{1}{2}\left[1 + erf\left(\frac{(1/\langle E\rangle - 1)^{1/6} R_0 - R_{DA}}{\sigma\sqrt{2}}\right)\right]_{(k-1)/n}^{k/n} \qquad (50)$$

Here, R_0 is the Förster radius, R_{DA} the mean distance between both dyes and σ_{DA} the standard deviation of the Gaussian distribution. The probability density distribution for n bins is interpolated to the number of bins used in the experiment n_{bin}.

A gradient-descent algorithm (*91*) with stochastically varied starting parameters is used to find the global minimum of the χ^2 error between experimental and theoretical efficiencies. For a faster minimization procedure, the shot noise limited distributions P_E are calculated for each bin $\langle E\rangle(k)$ before minimization. The only free parameters are the ones searched for: mean distance R_{DAi}, distance distribution σ_{DAi} and the relative fraction for i states. That way, it is possible to find multiple distance populations (i.e. states) for one FRET pair. This approach is especially useful for low efficiency distributions as required for the large open structures of Hsp90.

The experimental efficiency distribution obtained from an inter-monomer dye pair of the dynamic open state of Hsp90 (Figure 30, bars) is fitted with (49) (red lines) and then verified with Monte Carlo simulations (circles). The assumption of Gaussian distributed distances describes the experiment surprisingly well. When assuming rigid states instead, the distributions cannot be described with one, two or three rigid states. Five or more states have to be assumed, but then

they are weighted by a Gaussian distribution in the distance space, again. The time-dependence of the distance fluctuations are analyzed in Chapter 4.2.2.

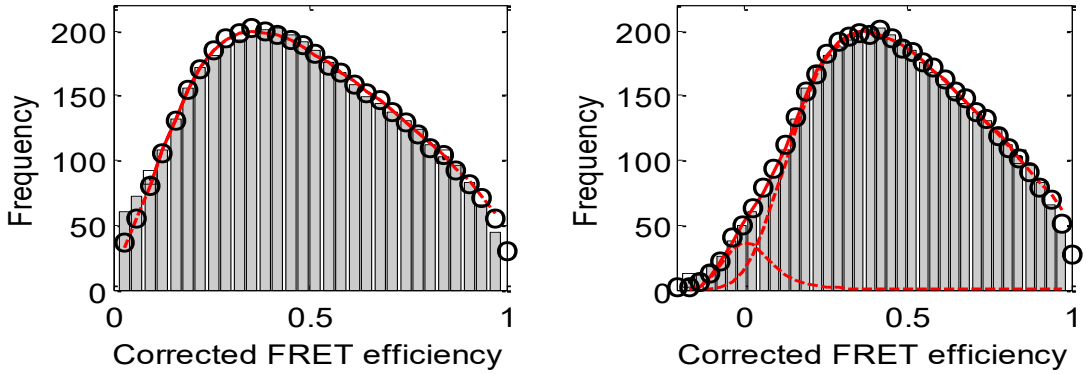


Figure 30: Corrected FRET efficiency histogram of the dynamic open state of Hsp90 was fitted with the probability distribution analysis that was derived in this work (red line) and was verified with a Monte Carlo Simulation (black circles) assuming either one Gaussian distance distribution (left) or two distributions (right). The photon binning time was 1 ms.

For relative rigid molecules such as double stranded DNA and the AMPPNP-bound closed conformation of Hsp90 (see Chapter 4.1.2 and 4.4.1), the broadening of the efficiencies should be limited to shot noise (i.e. Poisson statistics). However, a minimum distance distribution with a standard deviation of 2-3 Å was observed for all measured rigid molecules. A similar value has been reported by other groups (*89*), but could not be explained satisfactorily. It was suggested that intersystem crossing rates that are faster than the photon count rate cannot affect the broadening of efficiencies (*90*). In contrast, Monte Carlo simulations show notable convergence to shot noise limited efficiencies only if the intersystem crossing rate is about 10 times faster than the photon count rate (Figure 31). That would explain why the efficiency broadening for the rigid DNA samples exceeds shot noise.

Figure 32 demonstrates the applicability of the distance distribution analysis for several expected intensities n and distance standard deviations σ. If the distance-limited efficiency distribution is equal to or larger than the shot-noise limited efficiency distribution for a certain expected efficiency $\langle E \rangle$, the distance distribution can still be separated from shot noise. When detecting more than 100 photons per molecule event, distance fluctuations larger than 5 Å with expected efficiencies between 0.15 and 0.95 can be separated accurately from shot noise.

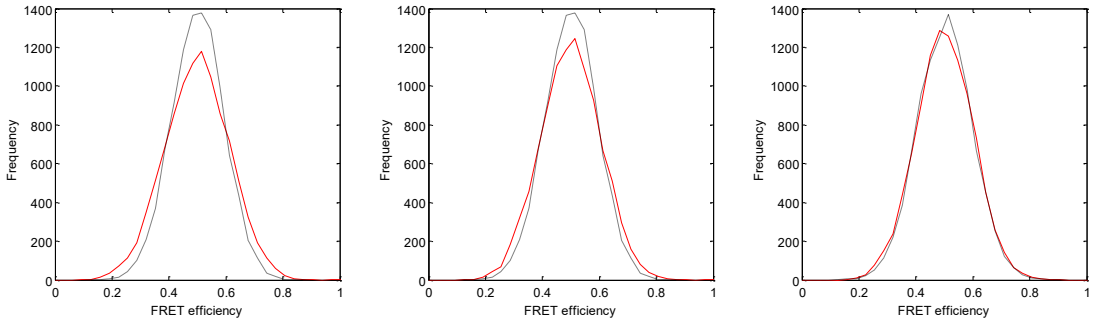


Figure 31: Monte Carlo simulations of efficiency broadening due to shot noise only (black dashed lines) and due to shot noise and intersystem crossing (efficiencies fluctuating between 0 and 1) (red line) for 30 photons and for blinking rates twice (left) or four times (middle) or 10 times (right) faster than the average photon count rate of 100 ms^{-1}.

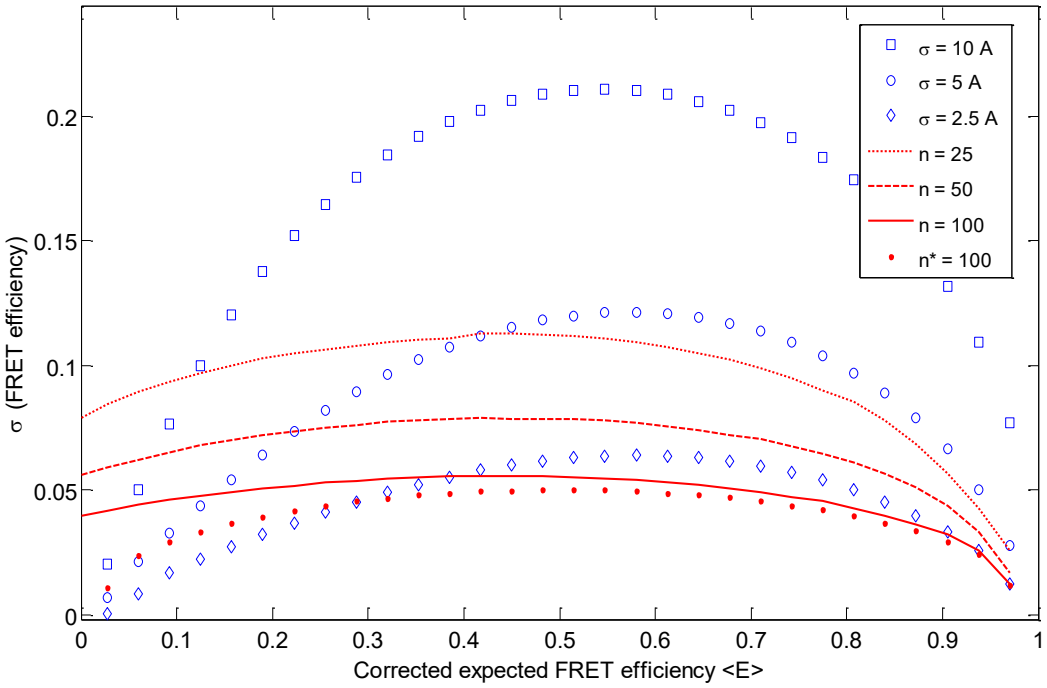


Figure 32: Standard deviations of efficiency distributions that arise only from shot noise (red) or only from Gaussian distributed distances (blue symbols) in dependence on the expected mean efficiency. For the shot-noise limited distributions, the number of expected photons *n* was varied while using either the correction factors *c*=*d*=0.1 and *γ*=1.2 (red lines) or no correction factors representing ideal FRET dyes (red points). For the distance distributions, the Gaussian standard deviations *σ* were varied and the mean value was *μ*=*R*$_0$ in each case. Each symbol represents the standard deviation of one distribution.

In principle, describing the raw data with the help of simulations is probably more accurate than data processing, but is often much more time-consuming and often has many degrees of freedom. Pre-correction of intensities and pre-calculation of shot noise limited efficiency histograms

(before minimization) while considering the original photon statistics allows a fast deconvolution of distance distributions. This enables fast global analysis of distances and thus nearly arbitrary distance models, as well as, simultaneous fitting of efficiencies and stoichiometries. Moreover, the corrected efficiencies and the underlying fluctuations can be directly correlated with stoichiometries to distinguish whether the fluctuations arise from structural heterogeneities or from photophysical heterogeneities such as intersystem crossing. Figure 33 exemplary demonstrates how the distance distribution (middle) can be assigned to the respective stoichiometries (left) to check for quenching, and to the respective lifetimes (right) to check for dynamics. Apart from these considerations, pre-correction of intensity traces leads to equally weighted expected efficiencies thus enables state assignment via populations.

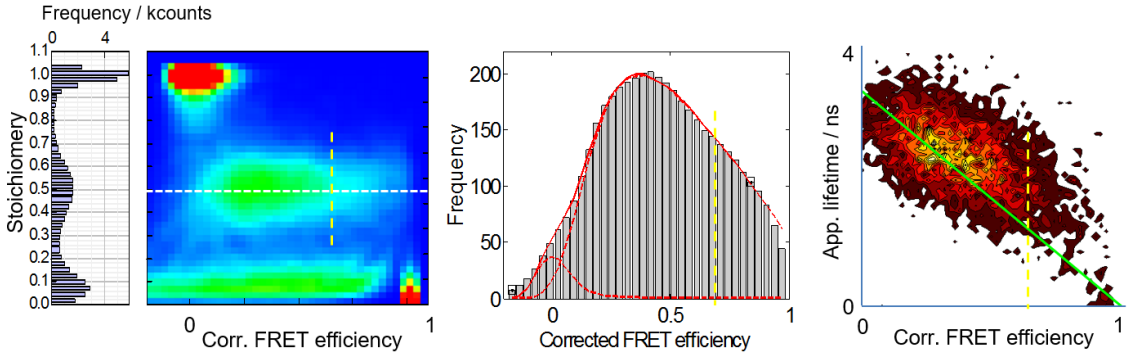


Figure 33: Distances from the distance distribution analysis of corrected FRET efficiencies (middle) can be directly compared with respective stoichiometries (left) to check for quenching and respective lifetimes (right) to check for respective dynamics. The stoichiometry-efficiency relation indicates that the efficiency broadening is dominated by distance fluctuations and not by photophysics.

4.2.2 Time-correlated distance distribution analysis

To further understand the underlying mechanistic model of the dynamic open state of Hps90, the timescale of the measured distance fluctuations was analyzed.

Conformational changes that are slower than the binning time of the fluorescence intensities, result in a convolution of their expected efficiencies as Dirac delta functions with shot noise. Conformational changes that are faster than the binning time, result in a convolution of the equilibrium average of their efficiencies with shot noise. Conformational changes that are 10 times faster than the photon count rate, do not notably affect the efficiency broadening beside shot noise. Small distance fluctuations are most likely averaged at a timescale close to the photon count rate. Now, the amplitude by which the fluctuations change with binning time can reveal the averaging effect and thus the timescale of the dynamics.

The burst variance analysis (*75*) is a good indicator for the timescale(s) of distance fluctuations. Fluctuations that occur at timescales slower than diffusion time (burst duration), result in bursts

with different efficiencies. In this case, when dividing the bursts into smaller bins, the variance of each single burst equals the variance of shot noise. Fluctuations that occur at timescales faster than diffusion time and slower than binning time, result in a variance of each divided burst that is larger than the corresponding shot-noise limit. For the open state of Hsp90, the variances were much larger than shot noise indicating fluctuations that are faster than the average burst duration of 2 ms (see Figure 58). The stoichiometry-efficiency relation additionally suggests that these fluctuations are caused by distance fluctuations (Figure 33).

To further encircle the timescale of these distance fluctuations, Gaussian distance distributions were separated from shot noise for different binning times (Figure 35); and then, their standard deviations were plotted versus the binning time, i.e. observation time (see Figure 47). It should be noted that by choosing smaller time bins with a constant intensity threshold, the average count rate increases, and thus, faster kinetics may be considered.

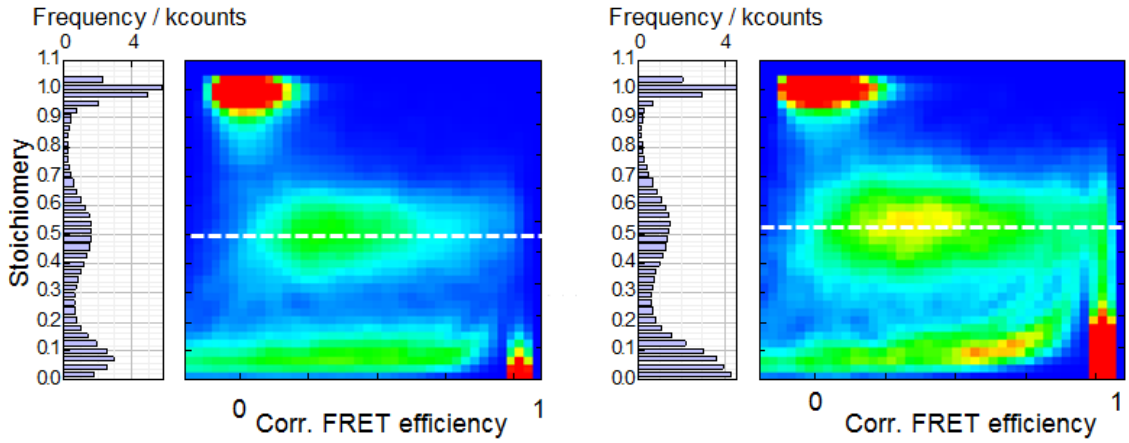


Figure 34: Stoichiometry-efficiency correlations for binning times of 1000 µs (left) and 200 µs (right) indicating that distance fluctuations are the main source of the observed FRET efficiency broadening beyond shot noise.

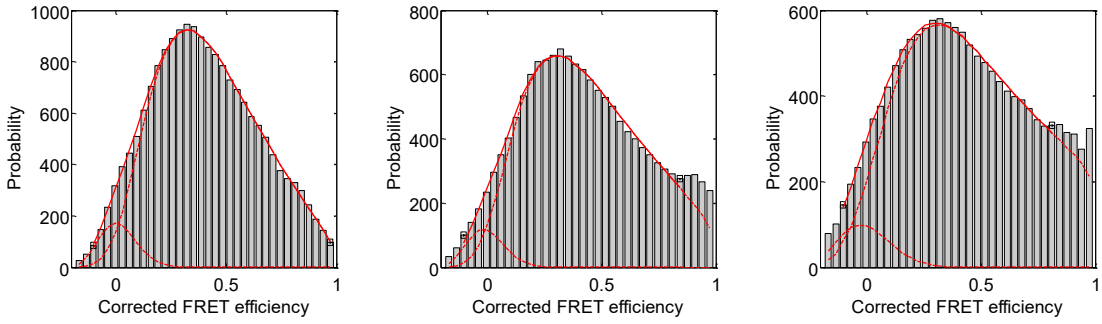


Figure 35: Distance distribution analysis of a FRET pair in the open ADP-bound state of Hps90 for different binning times: 2000 µs (left), 600 µs (middle) and 200 µs (right) resulting in Gaussian distance distributions with standard deviations of 10.5 Å (left), 12 Å (middle) and 14 Å (right).

For Hsp90, the standard deviations of the distance distributions increased slowly when decreasing the binning time from 10 ms down to 100 µs. A single two-state system or different states with similar rates are rather unlikely, because averaging would happen much faster. To further narrow the underlying mechanistic model of Hps90 in its open state, different kinetic models (A-D) were simulated (Figure 36); and then these models were compared with the experimental data. For this purpose, kinetic information from complementary fluorescence based methods were taken into account (see Chapter 3.2.6 and Figure 58).

Model A describes a simple two-state system that fluctuates between two expected efficiencies 0.25 and 0.75 with the rates $k_{12}=10\ ms^{-1}$ and $k_{21}=20\ ms^{-1}$ (Figure 36 A). For a binning time of 0.1 ms, three peaks are clearly visible, two static populations and one dynamic population due to averaging. For a binning time of 10 ms, an average and nearly shot-noise limited population remains. Such a two-state system cannot explain the strong efficiency broadening and the measured dependency on the observation time.

Model B describes an ensemble of states that fluctuate with a Gaussian distance distribution at a certain timescale. A Gaussian distribution would, for example, reflect the probability density of a diffusive process. One might think of a reduced diffusion constant due to reduced domain mobility within the multi-domain protein. The Gaussian fluctuations with $\mu_0=R_0$ and $\sigma=R_0/5$ were simulated at a timescale of 0.1 ms. These are nearly completely averaged for a binning time of 1 ms and thus do not reflect the measurements.

Model C describes two independent sources of Gaussian distributed fluctuations at two different timescales 0.1 ms and 1 ms each with $\mu_0=R_0$ and $\sigma=R_0/7$. The amplitude of the fluctuations decreases slowly with increasing binning time. This behavior well reflects the temporal dependency of the experimental efficiency histograms. Similar simulation results are obtained by combining a three-state system at slower timescales (> 1 ms) and Gaussian fluctuations at faster timescales (< 0.1 ms).

Model D describes a continuous random walk. Brownian motion between two molecules (i.e. two Hsp90 domains) with an average diffusion coefficient of $(R_0/12)^2/(0.1\ ms)$ according to an increased viscosity of $\eta=1$ kg/(s·m) due to a decreased domain mobility was simulated together with a Gaussian potential with $\mu_0=R_0$ and $\sigma=R_0/3$. In this case, the efficiency broadening decreases even more slowly with increasing binning time compared to Model C. Model D might also hold for the assumption that Hsp90'open state consists of multiple conformations with similar energy barriers between them. In future, model D might be extended by a random walk through a fractal-like structure.

Model C and D both reflect the experimental data fairly well. It became pretty clear that Hsp90's open state is better described by a distribution of states and a distribution of rates rather than single ones. As thermal fluctuations at the size of amino acids occur at a timescale of about 10 ns,

the distance fluctuations at a timescale of about 1 ms should roughly correspond to a free energy of about 10 k_BT. See Chapter 4.4.2 for further discussions about possible underlying structural mechanisms. To further encircle a comprehensive structural and kinetic model, in future, immobilized proteins will be measured in a confocal setup to access the complete time range between 0.1 and 100 ms.

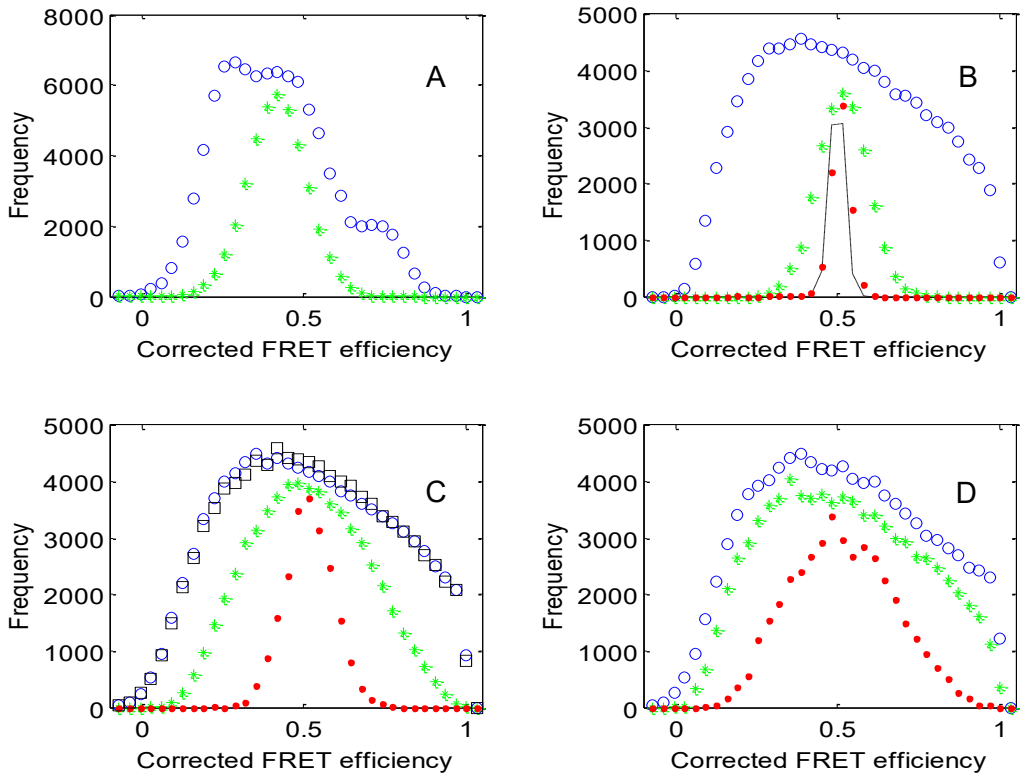


Figure 36: Monte Carlo simulations of efficiencies with different kinetic models for comparison with the observed inter-monomer distances fluctuations. The intensity threshold is set to 50 photons per time bin. (A) Two state model with k_{12}=10 ms^{-1} and k_{21}=20 ms^{-1} for a binning time of 0.1 ms (blue) and 1 ms (green). Additional faster rates does not further increase the efficiency broadening (B) Gaussian distributed spatial fluctuations around $\mu_0=R_0$ with $\sigma=R_0/5$ at 0.1 ms and for a binning time of 0.1 ms (blue), 1 ms (green) and 10 ms (red) and for comparison the shot noise limited efficiency distribution (black dashed line). (C) Two independent sources of Gaussian distributed spatial fluctuations around $\mu_0=R_0$ with $\sigma=R_0/7$ at two timescales 0.1 ms and 1 ms and for a binning time of 0.01 ms (black), 0.1 ms (blue), 1 ms (green) and 10 ms (red). (D) Continuous random walk within a Gaussian potential with $\mu_0=R_0$ and $\sigma=R_0/3$ and an average diffusion coefficient of $(R_0/12)^2/(0.1\text{ ms})$ for a binning time of 0.1 ms (blue), 1 ms (green) and 10 ms (red).

From Monte Carlo simulations of different kinetic models, standard deviations of distance distributions were calculated in dependence on the observation time (Figure 37). Surprisingly, even for a simple two-state system, the distance distribution varies over a wide range of timescales, namely about three orders of magnitudes. The two-state curves (blue, red) can be

separated in two functional regions, an exponential region at smaller timescales and a power-law dominating at larger timescales. The exponential part is probably linked to the relaxation rate while the square-root dependency on larger observation times, is most likely originated in the Poisson statistics of the fluctuations. The steepest position τ of the curves seems to be related to the relaxation rates k_{12} and k_{21} in the following manner: $\tau=a/(k_{12}+k_{21})$ (see inlet). The constant a has a value of 4 for these curves. A random walk in a harmonic potential with a single diffusion coefficient resulted in a similar shaped curve (green). Only if rates on different timescales are involved, the overall shape might change. Different relaxation rates for a three-state system can be distinguished only if the relaxation rates differ by at least two orders of magnitudes (see violet curve). The violet curve comes close to the data of Hsp90's open state.

Finally, this approach seems to fill a gap in the fluorescence field. The big advantage in comparison to standard fluorescence correlation spectroscopy is the insensitivity to intensity fluctuations and thus the applicability in the range of the protein's diffusion time.

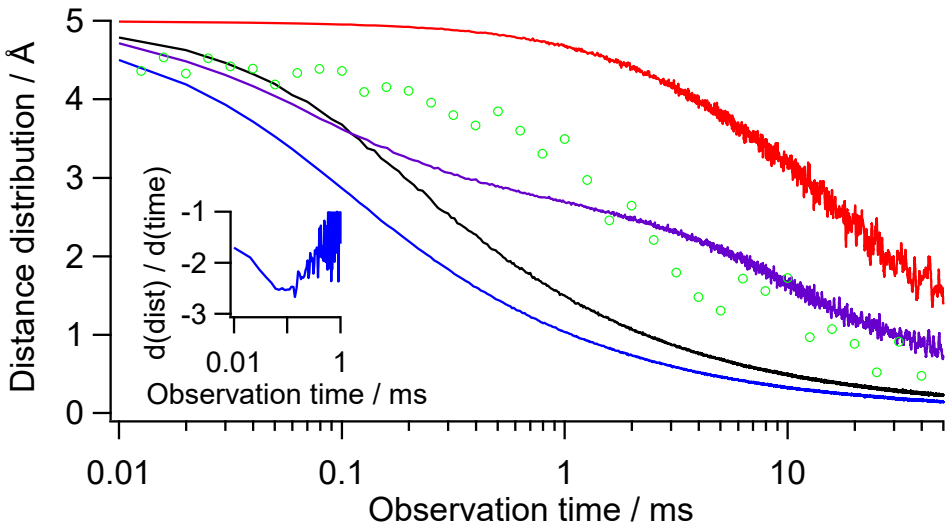


Figure 37: Simulated distance distributions versus observation time for two and three state systems with maximum fluctuations of 10 Å. The blue and red curves represent two-state systems with equal forward and backward rates: $k_{12}=k_{21}$=20 ms^{-1} (blue) and $k_{12}=k_{21}$=0.2 ms^{-1} (red). The black curve represents a two-state system with different forward and backward rates: k_{12}=20 ms^{-1} and k_{21}=0.2 ms^{-1}. The violet curve represents a three-state system with slow rates between state 1 and 2 and fast rates between state 2 and 3: $k_{12}=k_{21}$=0.2 ms^{-1}, $k_{23}=k_{32}$=20 ms^{-1} (comparable to model C). For lucidity, the black curve was multiplied with a factor of 5 and the violet curve with a factor of 1.4. The steepest positions reflect the correlation relaxation times of the rates. The derivative of the blue curve (inset) shows a minimum at 100 µs. The random walk simulation in a harmonic potential resulted in slow changes of the distance distribution (green) (comparable to model D).

4.3 FRET based dynamic protein structure

Determination of protein structure requires the protein or protein complex to be static on the observed timescale. Crystallization of the protein in one state enables the structure to be measured with x-ray diffraction, whereat averaging of several minutes is currently necessary to obtain atomic resolution for large proteins. Therefore, only stable protein states can be observed. With cryo-EM techniques the protein is frozen in near-native states that can be classified with software algorithms. However, certain states can be averaged and the information about dynamics and conformational flexibility is lost. These methods figuratively record millions of distances simultaneously on relative long timescales, while the protein is frozen in a predominant state. In contrast, using single-molecule FRET each distance is recorded in a separate measurement, but in a near native environment and with high temporal resolution down to nanoseconds. Principally, it is possible to measure up to four FRET based distances simultaneously, but at the expense of decreasing precision. The question arises how to bring the single FRET based distances together to a network.

First, FRET efficiency populations should be assigned to a certain state. The simplest way to identify a state is to shift the populations with nucleotides or interacting proteins. Also, occupation probabilities, distance fluctuations or dye anisotropies can be compared for state assignment. For Hsp90 the global states were assigned by comparing the dynamic ATP-bound state of Hsp90 with respective measurements of AMPPNP-bound and ADP-bound Hsp90. Further confidence in the assignment was archived during the final minimization between model structures and measured distances network.

Second, fluorescence based correction parameters can be verified for their self-consistency in order to exclude artifacts caused by photophysics. By comparing distances and position-specific properties of the AMPPNP-bound structure with the ADP-bound structure, further confidence was reached. The influence of the dyes on the protein itself was minimized by choosing suitable positions and has been controlled by versatile tests.

Different approaches for the docking of rigid protein structures have been scrutinized with the large dataset for Hsp90 and have been compared based on several physical parameters. The closed AMPPNP-bound state of the large multi-domain protein Hsp90 is perfectly suited for verification, because it is relative rigid on the observed timescale (see Chapter 4.4.1), and because model distances can be calculated with the existing high-resolution x-ray structure (pdb: 2cg9). Calculation of model distances, dye characteristics and anisotropy based selection criteria are developed in Chapter 4.1.

A sequence of rigid body docking, anisotropy-selection, and self-consistency check of redundant distances, allowed determining the accuracy of the individual distances. Network-independent and accurate single distances are useful for a free implementation of structure generation. Finally,

a method for dynamic structure generation has been developed that is based on a global analysis of both mean distances and time-resolved distance distributions. In contrast to spin labeling methods, these distributions are obtained from single-molecule trajectories at the micro- to millisecond timescale, so that averaging of the conformational ensemble is prevented.

4.3.1 Dye positions

FRET dye positions were chosen in such a way that they do not hinder protein function and conformational changes. Most positions are located in small loops between secondary structure elements to increase the accessible volume for the dyes and their mobility. The dimeric nature of the protein helped to considerably reduce the amount of cysteine mutations. With 15 mutations, 225 distances can be measured per state, in principle. As, within Hsp90's open state, the N-terminal domains (N-domains) are too far away from each other to be detectable with FRET, several intra-monomer FRET dye pairs were added to the network to bridge large distances. Negligible influence of the dyes on the protein itself was assumed if there was significant ATPase activity (i), if the populations changed depending on nucleotides (ii) and if the stoichiometries were small distributed indicating no aggregation (iii). Additionally, several positions were double checked by different dyes and by swapping donor and acceptor positions (iv).

4.3.2 Self-consistency

For each single measurement, photophysical correction parameters were calculated. Their consistency was verified by generating a parameter network with redundancies. About 10 distances (FRET dye pairs) are linked to each dye position and each distance was measured at least four times.

For example, for three FRET pairs – (1) between the positions A and B with the gamma factor γ_{AB}, (2) between the positions A and C with γ_{AC}, and (3) between the positions B and C with γ_{BC} – the relation $\gamma_{BC}=\gamma_{AB}/\gamma_{AC}$ must hold. In the case of disagreement for the minimal number of three pairs, it is difficult to locate a possible erroneous measurement. However, within the complete network of FRET pairs, a disagreement can be directly assigned to one single measurement. The respective measurement is excluded as the causes for such an outlier are normally setup misalignment, wrong labeling or contaminated buffer, which cannot be corrected afterwards. Another cause, which is discussed in Chapter 4.1, could be distance-dependent intensity fluctuations due to dark states. A strong indicator for such interfering dynamics is a deviation between the local and global gamma factor of more than about 20 %, which is used as a criterion for excluding such a FRET pair.

On the other hand, redundancy is given by calculating the gamma factor from stoichiometry and FRET efficiency for a specific FRET pair and then from donor and acceptor lifetimes assuming constant detection efficiencies. A disagreement of both values could originate from setup

misalignment, other wrong correction factors, blinking or bleaching. As it is difficult to correct for these kinds of artifacts after the measurement, such measurements were simply discarded.

Correction factors such as direct acceptor excitation or spectral crosstalk should be the same for each measurement. A possible disagreement within the network may hint towards contaminated buffer, position-specific spectral shift or changed excitation coefficient of the dye. Moreover, the position-specific fluorescence anisotropies of the dyes were used as fingerprints and compared within the network. One reason for a residual acceptor anisotropy that is lower than the others at that position within the network, could be energy transfer from the acceptor to the donor or strong coupling, which becomes relevant for distances smaller than 40 Å. In general, a deviating position-specific parameter can be used as a lab-check to prevent incorrect etiquettes.

A small conformational change within a domain may result in a change of the dye anisotropy at a certain positon. This anisotropy can be used to indicate small conformational changes that correlate with the changes of the global states. For Hsp90, the anisotropies were compared between the closed AMPPNP-bound structure and the open ADP-bound structure. Figure 38 shows the anisotropy differences of a dye located at the interface between M- and N-domain for three states.

Furthermore, for each single FRET pair, the distance distributions of the rigid closed and flexible open state of Hsp90 were directly compared with each other.

Last but not least, the FRET network can also be used to estimate reliable occupation probabilities and their errors.

The distances of the few excluded FRET pairs applying the above-mentioned quality criteria only deviated by few angstrom from the distance network.

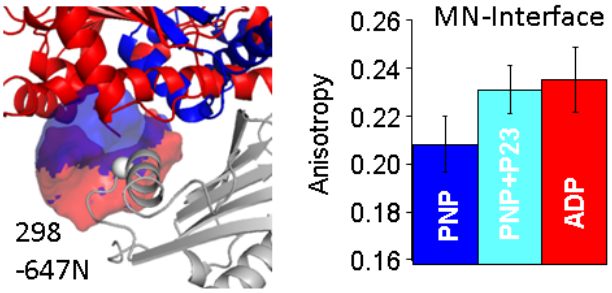


Figure 38: The anisotropy of Atto647 bound to position 298 of Hsp90 increases significantly (estimated from standard deviations) from the closed AMPPNP-bound to the open ADP-bound state, because the interface between N- and M-domain changes from the closed to the open state.

The consistency of the actual distances and their distributions is verified during structure arrangement (Chapter 4.3.4). Figure 39 demonstrates possible cases for systematic errors that are recognized within the network in case of sufficient redundant distances.

Finally, the anisotropy criterion derived in Chapter 4.1.4 is applied to the closed state Hsp90. The x-ray crystal structure of Hsp90's closed state was reproduced with an RMSD of 4.8 Å for all distances and 2.8 Å for anisotropy-selected distances. Distance deviations larger than 8 Å were stated for 17 of out 184 FRET pairs or 3 out of 148 anisotropy-selected FRET pairs. The fact that primarily distances of FRET pairs with high anisotropies deviated from the model structure, served as additional confidence.

Notably, when arranging the middle domains with the distance network as described in the following chapters, the arrangement for all distances was similar to the arrangement for the anisotropy-selected distances within an RMSD of 3 Å. This indicates that redundant distances are capable to outrival high-anisotropy distances and other single systematic deviations if and only if assuming constant dye accessible volumes.

It turned out that if less than about 10 % of all distances are wrong, they do not noticeably influence the result of the domain arrangement. As a rough guideline for domain docking, at least 20 FRET based distances should be used per domain pair in case of unselected distances and at least 10 distances in case of anisotropy selection.

The measured distance networks are listed in the appendix in Table 4.

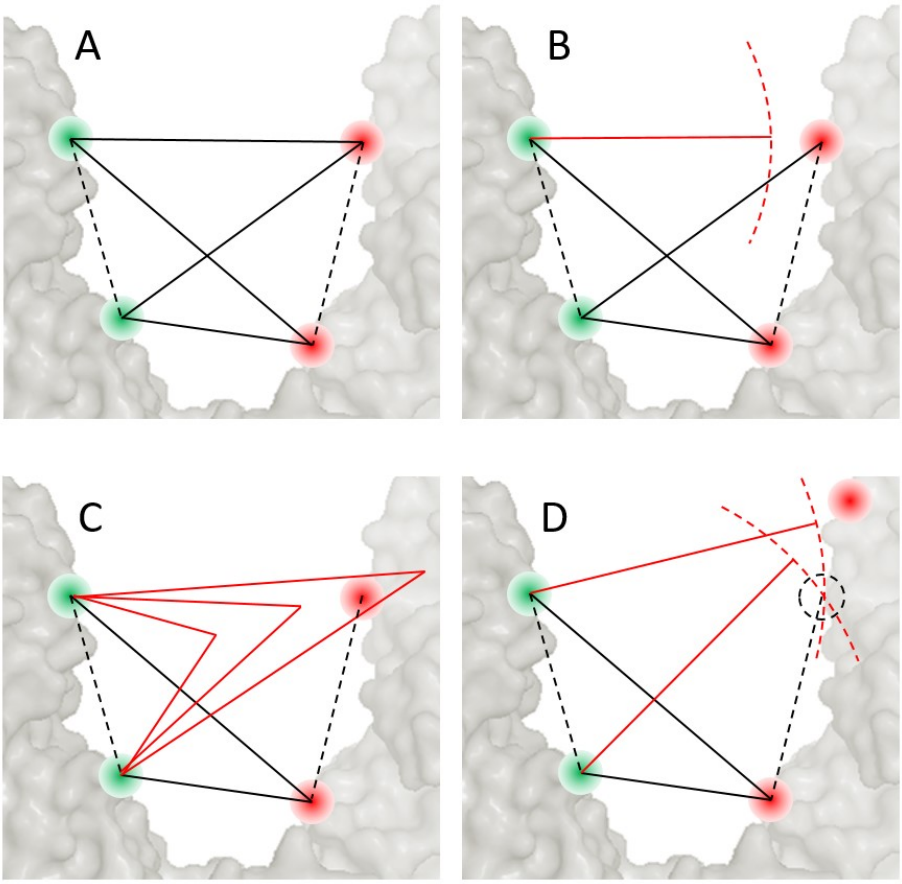


Figure 39: Possible systematic errors for FRET based distances demonstrated for two donors and two acceptors, resulting in four measurable distances (solid black lines) and 2

distances known form the x-ray structures of the single domains (dashed black lines). Verifying the self-consistency of the distance network enables monitoring deviating distances (red lines). Figure (A) represents an ideal self-consistent network with correct mean positions (dye accessible volumes are not illustrated) and correct mean distances. Figure (B) illustrates the case, in which the geometric mean anisotropy of the upper distance is relative large leading to a wrong consideration of the dipole orientation factor and therefore to a deviating distance. Such a single deviating distance will be directly recognized within a larger network. (C) Several distances to one position can deviate, for example, if the local quantum yield of the dye at this position is quenched. The gamma factor network indicates such possible quenched dye positions. Additionally, a larger distance network is capable to indicate such collectively deviating distances. (D) Several distances to one position can deviate, for example, if the dye accessible volume and thus the mean position are wrongly calculated. In this case, at least one of the distances will deviate already in this small four-pair network. The position error can be stated within a larger distance network.

4.3.3 Comparison of rigid body docking approaches

Two different approaches for rigid body docking have been recently published (*41, 42*). Both approaches share the assumption of rigid domains and model each dye as an ellipsoid and a stick capturing a certain accessible volume. In short, the NPS approach is a more theoretically elegant and the FPS approach is a more pragmatic solution, both coming with benefits and drawbacks.

The nano-poisitioning system (NPS) software considers the dyes to be restricted in their rotational and translational motion while the global relationship between numerous FRET dyes is used to limit possible dye positions and dipole orientations. The degrees of freedom are further reduced by means of measured dye anisotropies and FRET sensitized anisotropies. The main feature is the reduction of possible dipole orientations by calculating most likely posterior accessible volumes and position density distributions. Such a global statistical analysis is only beneficial and correct if most of the dyes are really restricted in their motion.

The FRET-restrained positioning and screening (FPS) approach assumes the opposite of the NPS approach, namely that the dyes are mobile and the transition dipole moment orientations are dynamically averaged. It accounts for averaging in different parameter spaces and for all possible combinations of distances between two homogeneously distributed dye position densities. It is assumed that the dyes sample their complete accessible space. Therefore the prior accessible volumes are not modified during the docking (minimization) procedure.

Dye accessible volumes

The first fundamental difference between both approaches is based on the handling of the dye accessible volumes before and during optimization. The NPS optimizes the probability density

functions of the prior accessible volumes during the docking process, following the idea that by diminishing the accessible volumes, the quantity of possible dye orientations and thus the transfer rate distributions are reduced. However, wrong distances or wrongly interpreted anisotropies artificially influence the posterior dye volumes in size and position. Due to the global statistical optimization process, wrong distances can be interpreted as correct ones while correct distances can be interpreted as wrong ones. If in reality the posterior volumes are comparable to the priors, the global analysis will likely increase uncertainty instead of decreasing it. In contrast, the FPS simply assumes constant dye accessible volumes with homogeneously distributed position probability densities during the minimization process. This has the advantage that the consistency of the distance network can be evaluated, if and only if the assumption that the dyes completely sample their accessible volumes is true in most cases. Wrong distances are not adapted by changing the position probability densities of the respective dyes and therefore can be identified as outliers. Outliers can arise from large anisotropies or from local conformational changes within a domain that was supposed to be rigid. The above considerations show that information about the real physics is essential, to decide about accuracy and confidence of both approaches.

The large dataset for Hsp90 revealed a significant anti-correlation between the modeled dye accessible volumes and the measured residual anisotropies (Figure 40, top). This suggests that the dyes sample to a large extend their overall accessible volume. Remaining deviations are probably due to different charges and hydrophobicity at the protein surface. Interestingly, there is nearly no high anisotropy where the accessible volume is large, indicating that there are probably no long-lasting states with the dye sticking to the protein. There is only one striking outlier (marked with a cross) indicating a partially restricted dye in spite of a very large accessible volume. Notably, this position has already been selected out with the criteria described in Chapter 4.3.1. A broad distribution of stoichiometries and a very small ATPase rate indicated that labeling at this position causes protein aggregation.

Figure 40 bottom demonstrates the impact of wrongly estimated dye accessible volumes on the error of the modeled distances. The deviations between model distances and measured distances become very large when, for example, calculating model distances between the cysteine's Cα atoms, which correspond to infinite small accessible volumes. Besides, nearly all Cα model distances underestimate the measured distances, because most of the dye accessible spheres are oriented outwards.

The the next question that is addressed is for which approach provide the available measurement parameters such as anisotropy more reliable quality criteria.

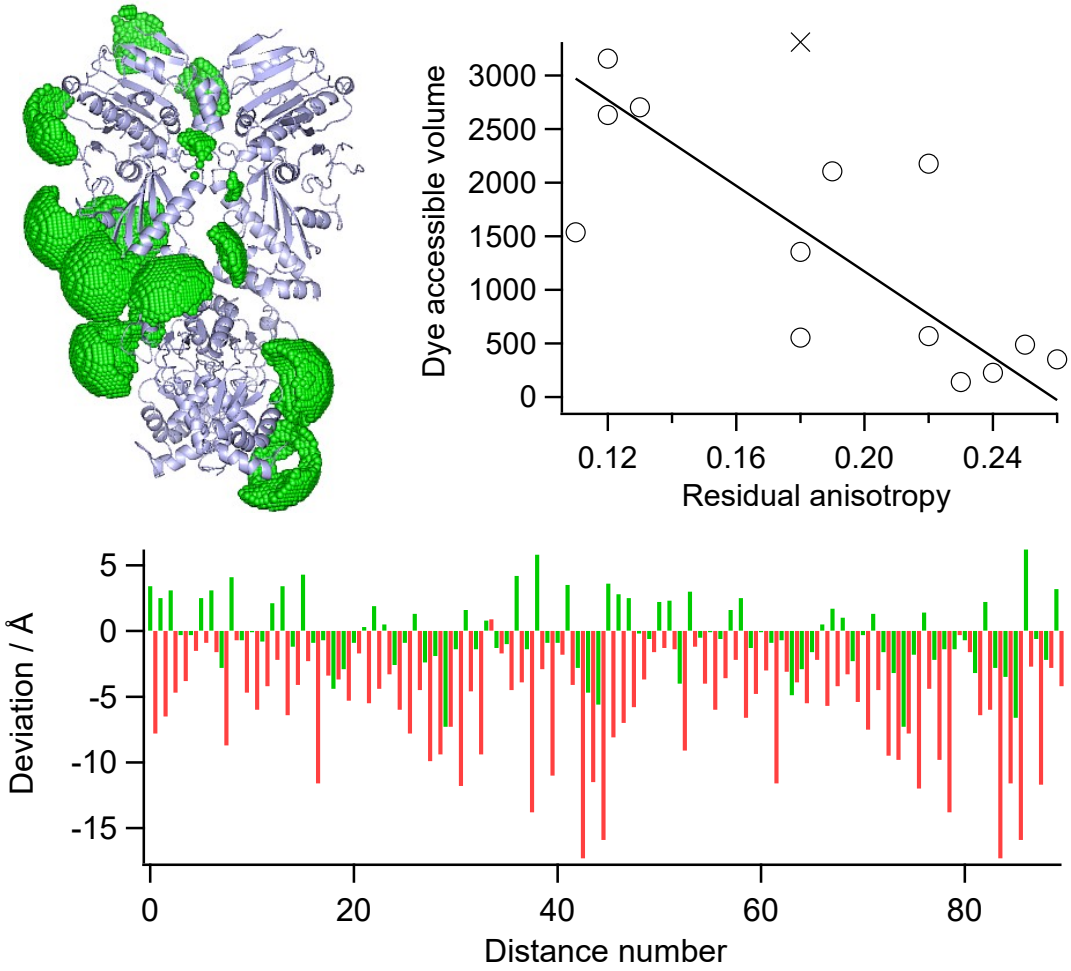


Figure 40: Top left: Dye accessible volumes for several amino acid positions to which the dyes are bound (shown for one monomer). The volumes consist of spatially equally distributed accessible positions above the surface of Hsp90's closed state while modeling the dye as an ellipsoid. Top right: Comparison of dye accessible volumes quantified as numbers of equidistant accessible positions with measured residual anisotropies, revealing a significant anti-correlation. The deviating amino acid 560 (marked as a cross) was already selected out by other criteria (see text). Bottom: Deviations between measured distances and model distances for the closed AMPPNP-bound conformation of Hsp90 (pdb: 2cg9). Model distances were either efficiency-averaged distances between dye accessible volumes (green) or distances between the cysteine's Cα atoms to which the dyes are bound (red).

Isotropic and anisotropic coupling

Another fundamental difference between both approaches is the handling of the transfer dipole moment orientations. In short, the NPS method can be seen an isotropic coupling approach (ICA), and the FPS method as an anisotropic coupling approach (ACA). At ideal isotropic coupling the dye dipole orientations are dynamically averaged and the orientation factor becomes $\kappa^2=2/3$ (Figure 41 left). At ideal anisotropic coupling the dipole orientation remains static for timescales longer than the transfer rate and κ becomes a value between 0 and 4 if the ideal dipole

approximation holds (Figure 41 right). It is important to note that the ideal dipole approximation applied by Förster breaks down for a completely anisotropic energy transfer and distances smaller than 50 Å (*50*). The real coupling can be identified with specific measurement parameters as discussed in the Chapter 4.1.4 and in the following.

Concerning the dye mobility, time-resolved anisotropy measurements revealed fast rotational relaxation times for the dyes at most labeling positons (< 1 ns) and suggest that the orientation factors are to a large extend dynamically averaged before energy transfer if the quenched donor lifetime is large enough (> 1 ns). Only about 10% of all measurements showed a larger anisotropy than 0.21, indicating a predominant number of measurements that fits the regime of isotropic coupling.

Focusing on the few cases with restricted translational dye diffusion, the highest measured anisotropy values were still considerably smaller than the fundamental anisotropy, which is about 0.38 for Atto550 and Atto647N. This indicates – additionally to the anti-correlation between accessible volumes and anisotropy – that the dyes were never completely fixed. If a dye moves more slowly than the transfer rate or sticks to the protein, single transfer rates would be required for each relative orientation leading to a distribution of transfer rates. This information is not deterministically accessible as the measured anisotropy is usually an averaged value over many orientations. Anisotropy populations caused by conformational changes that are slower than the photon count rate can be identified by plotting anisotropy versus donor lifetime in a scatter plot, but fast sticking states (< 10 µs) are averaged.

It must be noted that the *static* anisotropy can only be a good indicator for sufficient dynamic averaging if and only if the donor and acceptor lifetimes are larger than the rotational relaxation times of the dyes. In other words, if the lifetimes are too small (< 1 ns), the static anisotropy can be high, although the time-resolved anisotropy further decreases until transfer time.

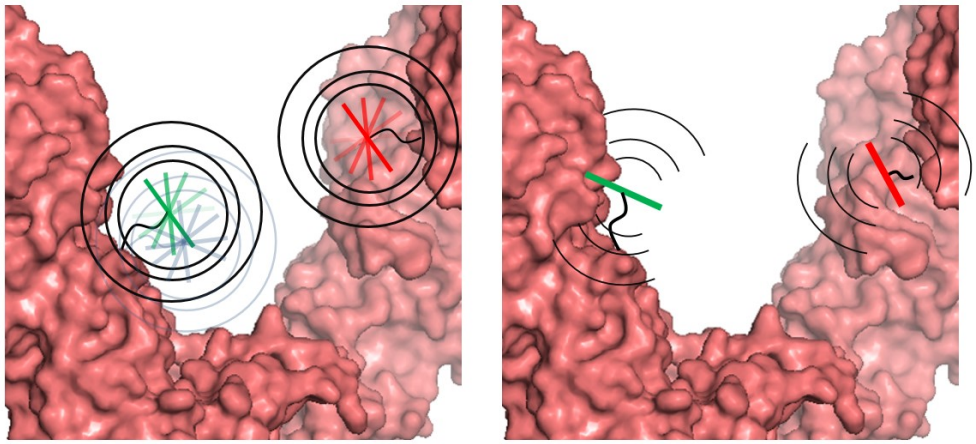


Figure 41: Comparison of an isotropic coupling (left) and anisotropic coupling (right) illustrating a time window of 1 ns. Left: If the relative dipole orientations are dynamically averaged before energy transfer, the average orientation factor is 2/3. Right: If the dyes

stick to the protein or are restricted by the surface for time periods longer than the energy transfer, the average orientation factor can deviate from 2/3 and is difficult to determine.

FRET sensitized anisotropy

Both approaches can use the FRET sensitized anisotropy – calculated from polarization sensitive acceptor florescence after donor excitation – additionally to the dye anisotropies as an indicator for the transfer anisotropy. However, within this work, no evidence was found for a correlation between the sensitized anisotropy and deviations from the crystal structure (see Chapter 4.1.4). First, the transfer anisotropy can be low although the dyes anisotropies are very high. Second, the FRET sensitized anisotropy does not represent the transfer anisotropy alone, because the excited acceptor further decreases the anisotropy during its fluorescence lifetime. An acceptor with a lifetime that is smaller than the rotational relaxation time, if existing, could largely decouple the transfer anisotropy from the acceptor anisotropy. However, this would be in contradiction to the large lifetime necessary for detectable acceptor anisotropy.

Anisotropy parameter and single-distance accuracy

The ICA and ACA approach are scrutinized by comparing 184 measured distances with their respective model distances considering the dye accessibility, followed by correlating the resulting deviations with the geometric mean of the measured residual anisotropies (see also Chapter 4.1.4). In form of an indirect proof, the orientation factor κ^2=2/3 is assumed for all calculated distances, in the following. In this case, small deviations at small anisotropies approve the ICA and large deviations at high anisotropies approve the ACA as it relates high anisotropies with a preferential orientation factor between 0 and 4. It has to be mentioned that there is a certain but small chance that the anisotropy is high, but the orientation factor is coincidentally 2/3.

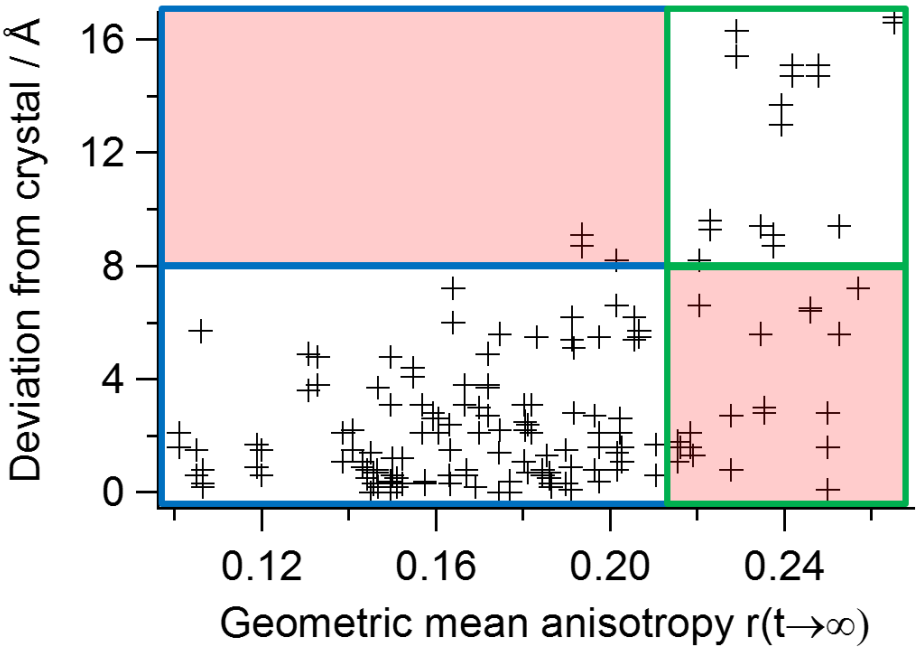


Figure 42: Deviation between calculated and model distances versus geometric mean anisotropy. The distances are calculated assuming the orientation factor to be 2/3. The blue and green frames illustrate the regions considered by the isotropic coupling approach (ICA) or the anisotropic coupling approach (ACA), respectively. The white areas indicate potentially correctly predicted distances (true positives and true negatives) and the red areas potentially wrongly predicted distances (false positives and false negatives).

Now, the amount of potential true positives, true negatives, false positives and false negatives are summed from the respective windows in Figure 42 for each approach (Table 3). The blue frames mark the measurements under the anisotropy threshold of 0.21 thus considered by the ICA and the green frames mark the measurements that are classified as positive for the ACA. The lower frames with a deviation of less than 8 Å represent the real positives for the ICA and the upper frames represent the potential real positives for the ACA (indirect proof). Thus, the white areas represent the truly classified measurements (true positives and true negatives or vice versa).

The ratios between the true positives and all positives indicate the approximate accuracy of the respective approach. The negatives have no influence on the accuracy, but on the fraction of data that cannot be used by the respective approach. Notably, the percentage of real positives (i.e. the data without anisotropy criterion) has a large impact on the approximate accuracy. Decreasing the tolerance limit of 8 Å for the maximal deviation decreases the approximate accuracy for the ICA and might increase the approximate accuracy for the ACA.

App-roach	Potential true positives	Potential true negatives	Potential false positives	Potential false negatives	Approx. accuracy* (Δ < 8 Å)	Real positives*
ICA	145	17	3	19	<98%	89%
ACA	17	145	19	3	≈50%	11%

Table 3: Approximate accuracy (number of true positives divided by the number of classified positives via anisotropy criterion) and percentage of real positives (number of true positives and false negatives divided by the total number) derived from Figure 42 for the ACA and ICA, respectively. The results imply that the truth of correct identification is strongly correlated to the geometric mean anisotropy. (*) The values can differ when using other dyes than Atto550 and Atto647N.

As a first conclusion, there is a much larger percentage of data that is useful for the ICA than for the ACA. Secondly, the probability for correctly determined distances is larger for the ICA than for the ACA. There are only few measurements with low anisotropies and large deviations (> 8 Å) being improper for the ICA. In contrast, there are several measurements with high anisotropies, but small deviations (< 8 Å) being potentially improper for the ACA.

To be on the save side, the orientation factor should be set to 2/3, while assuring dynamic averaging of the orientation factor by time-resolved anisotropy measurements. These results are proposed to be general for other proteins, if similar dyes are used. Small dyes with polarization axes parallel to the linker provide fast rotational relaxation, which should be faster than the energy transfer rate. Moreover, it has been found that the modeling of dye accessible volumes with homogeneously and isotopically distributed dye positions reflects reality very well. Thus, reducing these prior accessible volumes with a global Bayesian optimization seems to be an artificial process that increases the overall uncertainty.

The self-consistency of the large dataset for Hsp90 has been verified with several redundant distances and seems to be a more confident accuracy check than any global statistical analysis of non-redundant data. An important finding is also that at least one of 10 distances is recognized as an outlier if using the ICA and constant accessible volumes.

For the following generation of dynamic structures, the ICA brings the further advantage that the distances can be treated independently from each other (after checking their consistency).

4.3.4 Dynamic structure of a multi-domain protein

The following three-step procedure describes how the dynamic open structures of yeast Hsp90 have been generated based on the selected distance network for the open state. The distances were selected by applying the same criteria that had been derived for the distance network of the closed state.

Domain arrangement of an average structure

In the first step, an average model structure of Hsp90's N-terminal open ADP-bound state was generated based on the mean values of the measured distances. Therefore, the known x-ray structure of Hsp90's closed state was separated into individual domains and subdomains that are assumed to be globally rigid. For these substructures and for the required positions, accessible volumes with homogeneously distributed dye positions were generated with the FPS tool. According to above results, isotropic coupling between the dye dipoles was assumed. Instead of using all accessible positions for each dye, the center of mass positions of the position probability densities were calculated for the minimization procedure. This approximation is valid, because most of the distances of Hsp90's open state are larger than 50 Å, so that differences between FRET efficiency averaging and distance averaging can be neglected in the first run. After the first minimization, the resulting domain arrangement was iteratively reevaluated with the FPS tool in order to calculate potentially new accessible volumes, potentially new center of mass positions and efficiency-averaged model distances.

The minimization procedure has been implemented in Matlab. The selected domains were randomly positioned and orientated in space. A gradient-descent algorithm (*91*) with

stochastically varied starting parameters was used to minimize the deviation between the model distances R^{mod} and the experimental distances R^{exp} respecting the experimental uncertainty ΔR^{exp} using the following minimization criterion:

$$\chi^2_{single} = \sum_{i=1}^{n} \left(\frac{R_i^{exp} - R_i^{mod}}{\Delta R_i^{exp}} \right)^2 \qquad (51)$$

The experimental uncertainty mainly considered the sensitivity of the efficiency-distance relation. The evidence of other local minima arrangements was tested, too. Covalent linker restraints were added to accelerate the minimization procedure and to reduce the degrees for freedom. Intra-domain distances within the large M-domain were measured to verify preservation of the secondary structure from state to state.

The total root mean square deviation (RMSD) was iteratively improved by defining subdomains and by exchanging domains with domains from other x-ray structures (Figure 43). The final domain arrangement is based on the x-ray structure of the closed conformation of yeast Hsp90 (pdb: 2cg9), with the M-domain being divided into two subdomains and the N-domain being exchanged with another x-ray structure from the ADP-bound isolated N-domain (pdb: 2wep). This final domain arrangement fulfils the anisotropy-selected distances with an RMSD of 3.8 Å. An RMSD of 2.8 Å between model and experimental distances as obtained for the closed structure is not expected for the open state, because distance fluctuations are much larger.

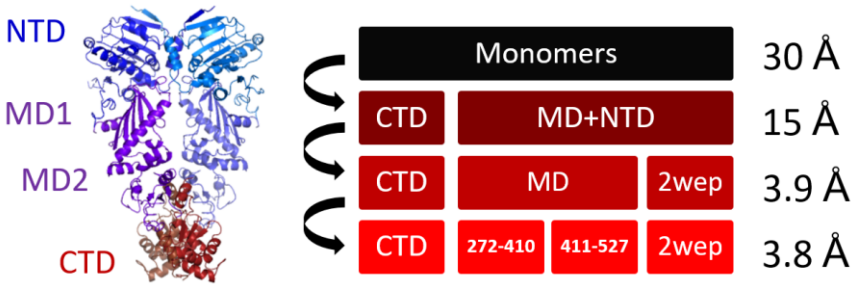


Figure 43: Domain arrangement for the open ADP-bound state of Hps90 using x-ray structures from the closed state (pdb: 2cg9) and homologues (pdb: 2web).

Deviations between the resulting average open structure and the measured distance network were analyzed concerning their consistency. The remaining maximum deviations from the mean domain arrangement for the open state were about 8 Å concerning only 2 of 82 selected mean distances. This approves the overall accuracy of FRET based distances and the modeling of the dye accessibility. When using all 130 distances for the open state (including the high-anisotropy distances), 14 % deviated with more than 8 Å. These values are similar to what has been found for the closed state. Remarkably, when arranging the domains with all distances, the resulting structure deviated only with 3.5 Å from the domain arrangement that was based on anisotropy-selected distances, which is within the estimated uncertainty. And, the high-anisotropies distances

were recognized as outliers. This means that the deviations between domain arrangement and measured distances nicely average each other out. Such an outrivaling is only possible when using several redundant distances. Additionally, distance deviations are recognized best, if prior accessible volumes according to the x-ray structure are used. A dynamic adaption of the accessible volumes as done in the NPS approach would complicate the recognition of real systematic deviations. For the following multiple structure analysis only the anisotropy-selected distances were used (green circles in Figure 44).

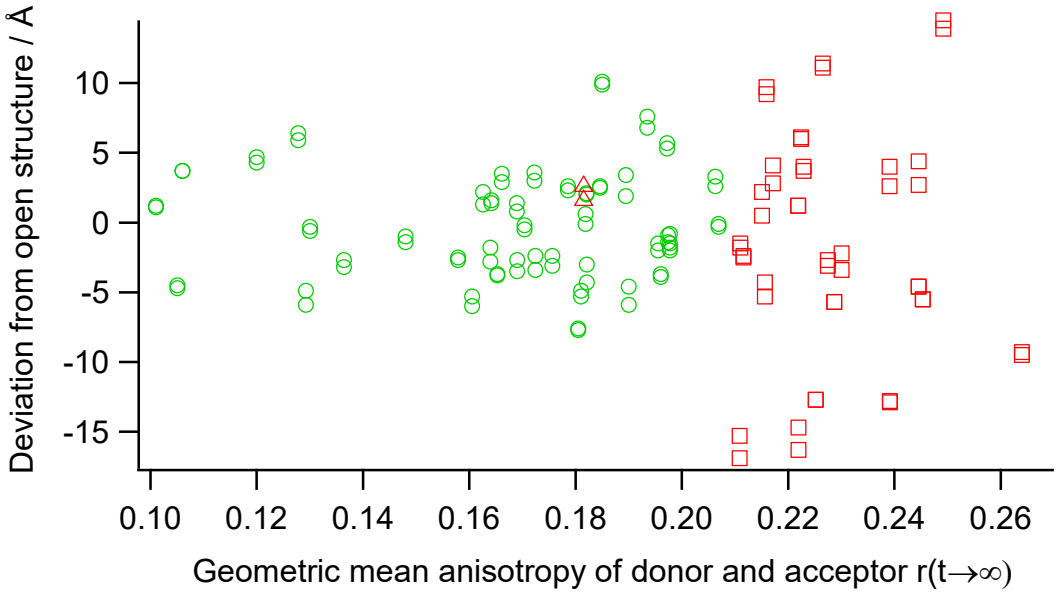


Figure 44: Deviations between measured mean distances and model distances of the average domain arrangement of Hsp90's open structure versus the geometric mean anisotropy of donor and acceptor dyes (see also Chapter 4.1.4).

Arrangement of multiple dynamic structures

In the second step, an ensemble of certain conformational structures was generated. Therefore, not only the mean distances but also the distance distributions were extracted from the FRET efficiency histograms (see Chapter 4.2) assuming single Gaussian distributions in the distance space. Notably, almost every histogram could be satisfactorily fitted with one single Gaussian distribution in the distance space. Now, it is tested whether all distance distributions can be fully described by a minimal structural model with two or more sub-conformations. Each of the n experimental Gaussian probability density functions G_i is fitted with l states, while each state is represented by an arbitrary Gaussian probability density functions G_k. This leads to the following minimization criterion:

$$\chi^2_{multi} = \sum_{i=1}^{n} \left(\frac{G_i\left(R \middle| R_i^{exp}, \sigma_i^{exp}\right) - \sum_{k=1}^{l} \rho_k G_k\left(R \middle| R_{k,i}^{mod}, \sigma_{k,i}^{mod}\right)}{\Delta R_i^{exp}} \right)^2 \qquad (52)$$

The parameter ρ_k weights each of the l states to obtain equal occupation probabilities for each of the n experimental distance distributions. To decrease the degree of freedom and to prevent overlaying Gaussians, their standard deviations were empirically constrained with: $\sigma_{k,i}^{mod} < \sigma_i^{exp} \cdot 1.5/l$. The best results were obtained when assuming one or three states. When assuming three states, the three Gaussians were constantly distributed within the respective experimental Gaussians, for most FRET pairs. Thereby, one Gaussian was located at the center and two Gaussians at the intervals (standard deviations) of the experimental Gaussians. Notably, this supports the general assumption of symmetry and congruent Gaussian intervals, but does not exclude asymmetric intermediates. That is why symmetric structures were generated from the experimental Gaussians based on the above-described domain arrangement. Then, these structures were used as rigid body restraints for full-atom MD simulations (see Chapter 3.3). The resulting structural ensemble visualizes a possible trajectory of the fluctuations (Figure 49).

It has to be mentioned that the distance fluctuations from some asymmetric FRET pairs (i.e. pairs where the positions are not the same in both monomers) were slightly but not significantly broader than those from symmetric FRET pairs. This could be indicative for one or several asymmetric structures. To test for a possible asymmetric structure, two Gaussians with equal occupation probabilities were modeled for each distance distribution that belongs to an asymmetric FRET pair. When allowing arbitrary standard deviations $\sigma_{k,i}^{mod} < \sigma_i^{exp}$, several solutions were found. The most probable one equaled the average symmetric structure found for single Gaussians. A second less likely asymmetric structure was found, which, however, only deviated with an RMSD of less than 6 Å from the symmetric structure. When artificially restricting the allowed standard deviations to $\sigma_{k,i}^{mod} < 8$ Å, the likelihood for asymmetric solutions was increased, logically. However, these solutions can be seen as asymmetric intermediates that lie within the symmetric ensemble.

Arrangement of structural ensembles

The fact that the left and right intervals (standard deviations) of the experimental distance distributions can be each assigned to a single structure, suggests that there might be a preferential trajectory with certain conformational substates. However, some distances differed from the model by more than the typical precision, possibly hinting towards a local conformational change leading to additional broadening of the FRET efficiencies or towards a remaining inconsistency of the left and right intervals.

Therefore, in the third step, an entity of structures is generated to visualize the possible disorder of global conformations. Structural ensembles were generated by positioning and orientating spatially equally distributed structures that lie within the standard deviations of the Gaussian distance distributions. Thereby, also less likely structures are considered, and, the arbitrary

restriction of classifying the data into left and right intervals is circumvented. The resulting ensemble somehow visualizes the spatial restriction of possible global conformations (Figure 45).

Notably, the average structure calculated from this structural ensemble deviates with an RMSD of 1.5 Å from the average structure that was calculated from the mean distances in the first step. Mathematically, a structure generated from mean distances is not necessarily the same as the real spatial average structure, especially for large fluctuations.

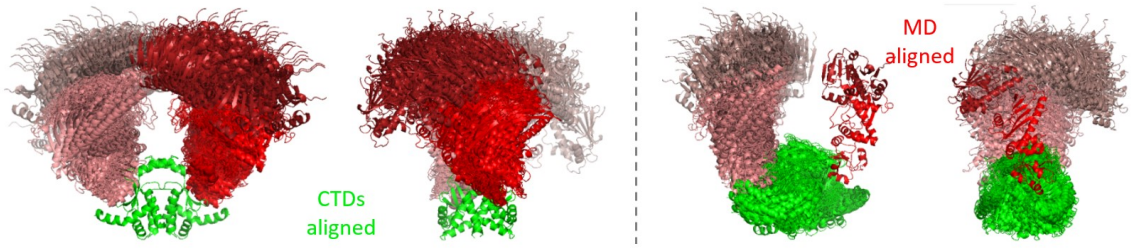


Figure 45: Ensemble of structures visualizing possible global conformations. The figure shows two side views of the ensemble that is aligned at the C-domains (left) and at the M-domain (right), respectively. The uncertainty of the positions relative to the C-domains is higher than the uncertainty of the positions relative to the M-domains.

Arrangement with global probability distribution analysis

The question arises whether the complete distance distribution can be extracted from an efficiency histogram. Figure 46 illustrates ambiguities in the probability distribution analysis. The efficiency distribution (top) can be described either with one Gaussian distance distribution (bottom, black dashed line) or better with three or more states, naturally. Assuming three states each with arbitrary Gaussian distance distribution and a minimum standard deviation of 3 Å, results in different solutions (two examples are plotted top left and right). Notably, the sum of these substates is quasi identical in the efficiency histogram for both cases (top, solid lines), but differs in the distance space (bottom, solid lines). The best procedure, to minimize these ambiguities, is a global probability distribution analysis with a global structural model. This means that the distance distributions are deconvoluted from the efficiency histograms on the fly during the domain arrangement process. The probability distribution analysis derived in Chapter 4.2 is suited for this computationally expensive task.

If there are three well populated conformational substates, they are most likely to be distinguished that way. However, if the open state is dominated by an entity of conformational disorder, there would not be the advantage of population assignment. In this case, thousands of structures would be randomly positioned and oriented in space, and then, each structure would obtain a product of probabilities for fulfillment of each measured distance distribution. Afterwards, the resulting structural ensemble would be refined, until the respective distance distributions globally fit to the efficiency histograms.

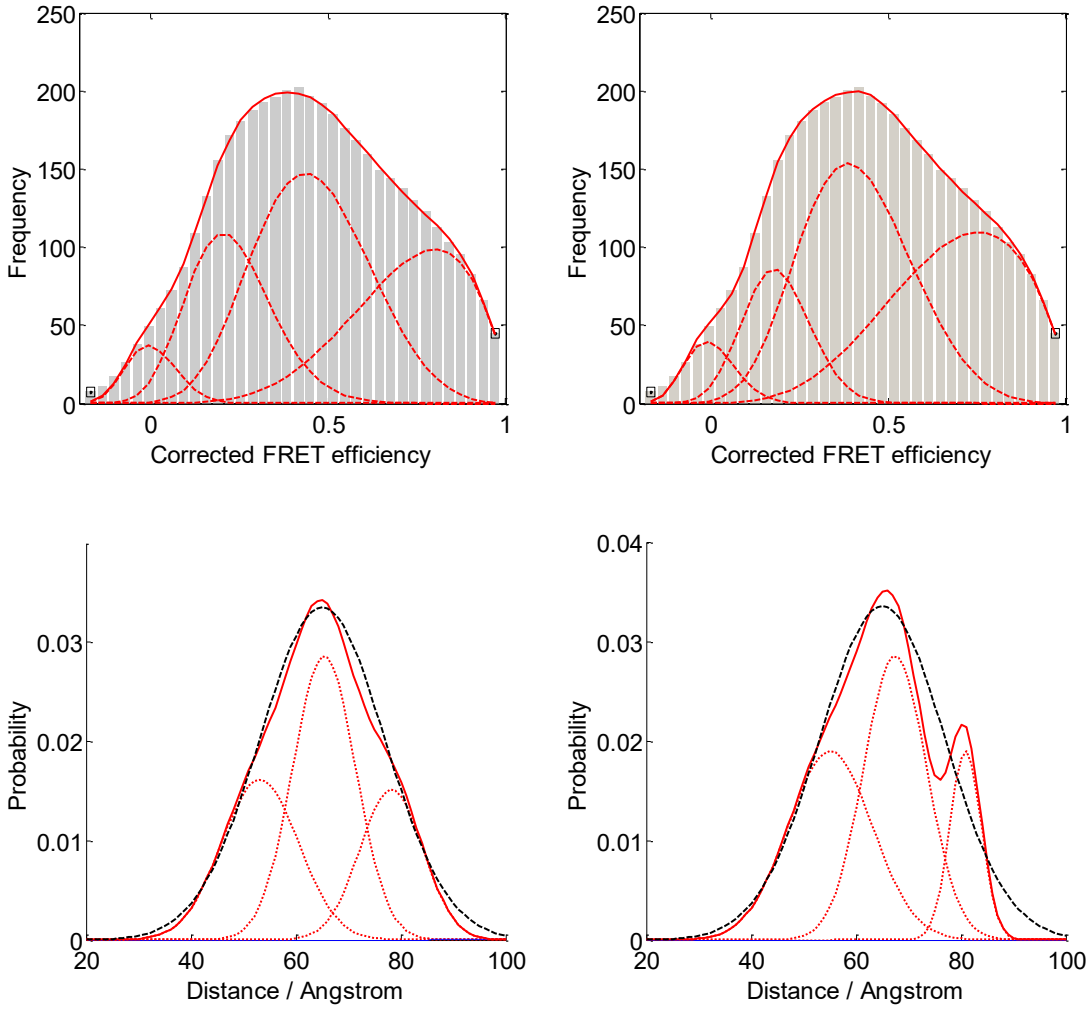


Figure 46: With the probability distribution analysis from Chapter 4.2, three states with Gaussian distributed distances are fitted to the efficiency histogram of a FRET pair located at the middle domains of Hsp90. The two most likely results are very similar in the χ^2 evaluation (top, left and right). (The small population at E=0 could be crosstalk from the donor-only population or could hint towards rare dissociation events. It is fitted with a forth distribution for the sake of completeness.) The sums of the respective distance distributions show clear deviations from each other (bottom, solid red lines), although the sums in the efficiency space are quasi equal. The black dashed line indicates the result of the probability distribution analysis when assuming only one state.

Systematic deviations and local conformational changes

The large number of distances that are connected in a self-consistent network further allowed characterizing the precision of the distances. Outliers were recognized and distorted the final structure arrangement only slightly. For example, some distances from FRET pairs with a high mean anisotropy (that were rejected by the anisotropy threshold) differ by up to 15 Å from the x-ray structure of the closed state (see Figure 48). Nevertheless, the application of an upper threshold for the anisotropy is important, especially for the arrangement of multiple dynamic

structures and for discovering flexible elements. Remaining outliers (systematic errors) with a deviation of more than 5 Å were analyzed with further FRET pairs to uncover possible conformational changes of small elements.

For example, position 452 systematically deviated from the distance network for the open state with about 5 Å, while agreeing with the distance network for the closed state within 2 Å. Additionally, the anisotropies of the dyes at this position are slightly larger in the open state than in the closed state, although the dyes at this position are outward-oriented. This indicates a small conformational change of the loop at this position and will be further analyzed. It is worth mentioning that subsets were generated by systematically excluding the distances to certain positions.

Constrained and non-constrained MD simulations

Selected single distances were integrated as constraints in MD simulations by Martin Zacharias and coworkers (see Chapter 3.3) using the x-ray structure of Hsp90's closed state as a start structure. The dyes have been modelled as a stick and a sphere, to emulate dye accessible volumes. However, the distance constraints between the dyes restricted dye mobility and shifted the dyes to preferential and artificial positions during the simulations. This was overcome by applying back-calculated distance constraints between Cα atoms. But, these force constraints led to partial unfolding of domains that are supposed to remain rigid on much longer timescales. Therefore, instead of single distances, the experimentally generated domain arrangements were integrated as constraints and then step by step loosen. The results for the open state of Hsp90 are presented in Chapter 4.4.

The next idea was to cover the experimental distance distribution during the MD simulations. However, the distribution-equivalent constraints caused long-lasting local minima of free energy during the simulation time of several 100 ns, so that only a small part of the experimental distance distributions was covered. That is why the generated model structures for the open state have been used as start structures for MD simulations without force constraints. In future, the measured distances distributions will be used to evaluate the unrestrained structures after certain simulations steps to generate copies of most likely structures in form of a Bayesian optimization.

4.4 Dynamic structure and mechanism of Hsp90

The multi-domain heat shock protein Hsp90 is a highly abundant and conserved protein in eukaryotic cells. It is essential for the activation of a large set of signal transduction and key regulatory proteins (*9-11*) and has been assigned critical functions in chaperoning oncogenic proteins (*12-14*) and evolution (*15*). The ATP-dependent functional Hsp90 dimer performs large conformational rearrangements involving its N-terminal and middle domains (*23*). The x-ray crystal structure of the C- and N-terminal dimerized closed state (*92*) of yeast Hsp90 triggered

several studies including extensive searches for inhibitors as cancer drugs (*93*). However, the dynamic nature of eukaryotic Hsp90s in the N-terminal open state, which is important for client binding, has challenged established methods and a high-resolution structure is still lacking.

Therefore, the approach for dynamic structure determination developed in Chapter 4.3, is applied to Hsp90 to provide dynamic ensembles of structures. Often several open or closed conformations of Hsp90 are proposed; here they are included as sub-conformations within the global states; namely the open and the closed state. A key aspect of the presented method is the preservation of the structure-dynamics relationship by simultaneously measuring absolute distances, distances distributions and their dynamics from micro- to milliseconds.

Based on the dynamic structure and additional kinetic analyzes, most likely mechanistic models of the multi-domain Hsp90 are discussed, followed by local conformational changes occurring in dependence on the global domain arrangements. An overall dynamic mechanism may decrease the apparent complexity of the Hsp90 machinery. Finally, physicochemical influences and modes of regulation are discussed on the basis of this mechanism and further preliminary measurement results.

4.4.1 Dynamic structure

Hsp90 was labeled with more than 100 FRET dye pairs across the entire protein. The measurements of one FRET pair are depicted exemplarily in Figure 47. Under saturating ATP conditions, yeast Hsp90 slowly interchanges on the timescale of seconds between a closed and an open state, represented by the two FRET efficiency peaks. The less populated peak equals the one for AMPPNP-bound Hsp90, i.e. the closed state. The most populated peak equals the one found for ADP-bound Hsp90, i.e. the open state. The efficiency histograms from all FRET pairs result in two networks of averaged distances, one for the closed state and one for the open state (Figure 48A). The depicted average open structure was generated using the approach developed in Chapter 4.3. The complete distance networks are listed in Table 4 in the appendix.

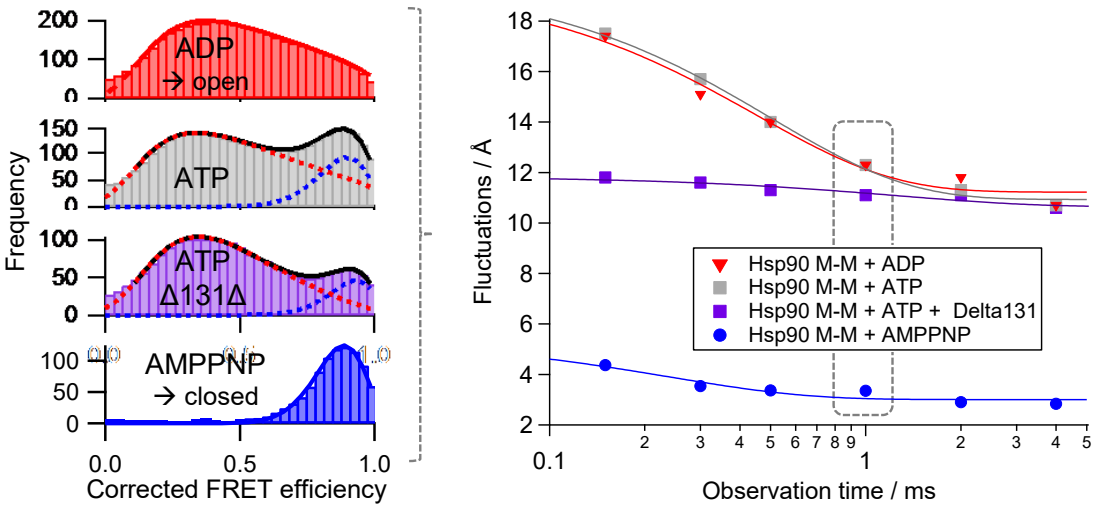


Figure 47: Measurements of one FRET dye pair spanning between the two middle domains of Hsp90 in the presence of different nucleotides and a client protein (substrate). FRET efficiencies are obtained from single molecules in a confocal microscope with alternating laser excitation and color- and polarization-sensitive detection with picosecond time resolution. Shown are efficiency histograms for a binning time of 1 ms and the results of a probability distribution analysis assuming Gaussian distance distributions as dashed lines (left). Structural dynamics of Hsp90's different states are analyzed with different fluorescence-based techniques. A time-correlated distance distribution analysis revealed the timescales of the distance fluctuations for the open state (right, red triangles), the open state in the presence of 2mM ATP (gray squares), the open state in presence of 2mM ATP and 20 µM of the model client Δ131Δ (purple squares) and for the closed state (blue dots).

Transforming FRET efficiencies into distances with Angstrom precision requires mobile dyes and knowledge of their accessible volumes and photophysical parameters. For this purpose, a correction procedure, self-consistency tests and selection criteria based on time-resolved dye anisotropies were derived in Chapter 4.1 and 4.3. Figure 48B exemplarily shows how the anisotropy criterion reduces the root mean square deviation (RMSD) between the distance networks and the average model structures. Altogether the x-ray structure of the closed conformation of yeast Hsp90 (pdb: 2cg9(*92*), resolution: 3.1 Å) could be reproduced with an RMSD of 2.8 Å (Figure 27).

In addition to the mean distances, spatial fluctuations on timescales from 100 µs to 10 ms were quantified within the closed and open state, respectively (see Figure 47 right for example data of one FRET pair). The time-correlated distance distribution analysis is developed in Chapter 4.2. Additionally, conformational changes on other timescales were analyzed with different fluorescence based methods (see Chapter 3.2 and Figure 68). As expected, the AMPPNP state is rigid with fluctuations smaller than 3 Å. Notably, the previously unresolved open ADP-bound state of Hsp90, as well as, its open ATP-bound state reveal fluctuations of more than 10 Å. Of particular note is the shift of timescale and amplitude of these fluctuations in presence of the model client Δ131Δ, a partially folded fragment of staphylococcal nuclease. The reduction of the fluctuations of the open state was independent of the present nucleotide. Separate fluorescence correlation measurements of labeled Δ131Δ in presence of Hsp90 and different nucleotides indicated that the model client preferentially binds to Hsp90's open state (Figure 75). Binding of Δ131Δ to certain residues of the M- and C- domain of HtpG had been shown with NMR and cross-linking experiments (*94*). However, timescale and mechanism had remained unclear. Here, the time-correlated distance distribution analysis allowed to estimate the timescale of interaction and the dependence of the binding affinity on the global Hsp90 state.

Figure 48C illustrates the standard deviation of the distance distributions at the timescale of one millisecond as average values per domain pair contrasting the closed AMPPNP-bound with the open ADP-bound state. Within the open state, the fluctuations increase from the C-terminal domains (C-domains) towards the N-terminal domains (N-domains). Within, the closed state the C- and M-domains are relatively rigid. The somewhat larger fluctuations of the N-domains were reduced when 10 µM of the co-chaperone p23 were added (Figure 74). This agrees well with the observation that p23 stabilizes the N-domains of the closed state (*36, 92*).

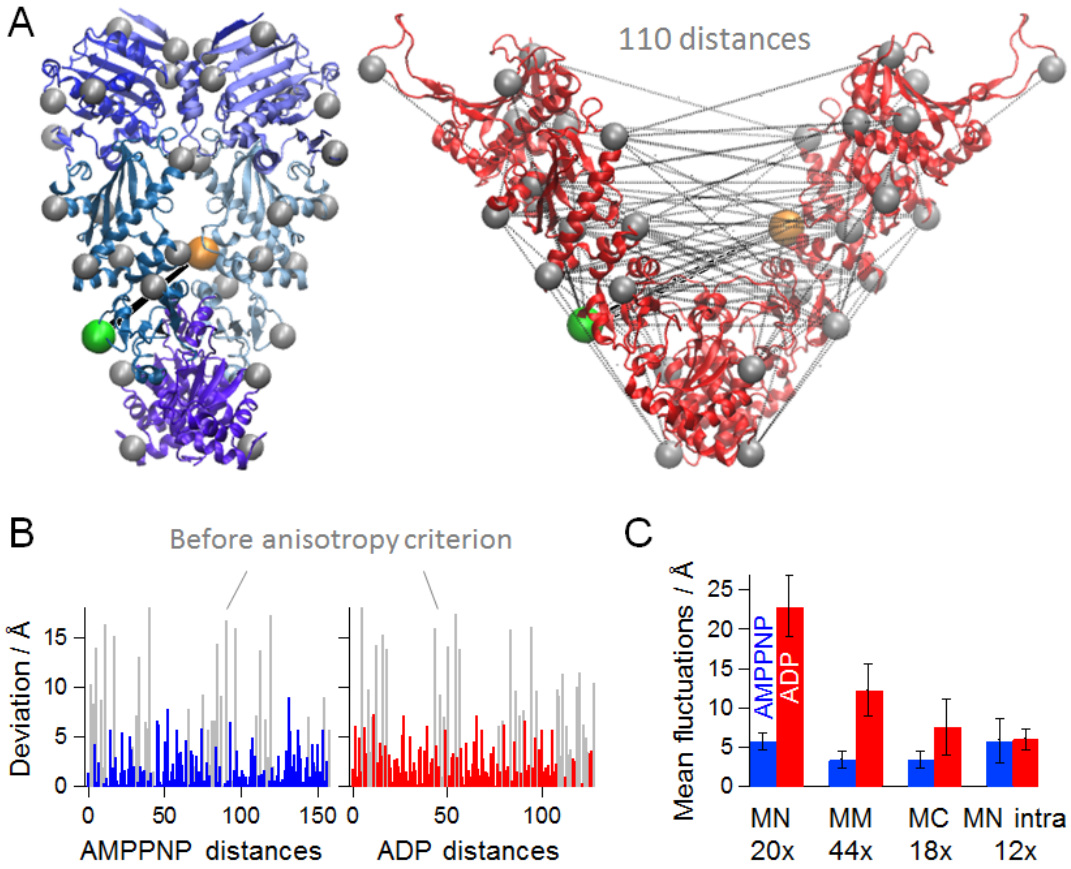


Figure 48: FRET networks reveal absolute distances and dynamics. (A) The first distance network was generated from 144 FRET dye pairs in the AMPPNP-bound state of Hsp90 and matches the x-ray structure (pdb: 2cg9) to 2.8 Å. The second distance network was generated from 110 FRET pairs in the ADP-bound state. The right picture illustrates the average open structure ignoring the distance fluctuations. (B) Deviations of the measured distances from the closed x-ray structure (blue) and from the average open structure (red) are shown with anisotropy criteria; and also without filtering data according to anisotropy criteria (gray). (C) The standard deviations of the Gaussian distance distributions were determined for a photon binning time of 1 ms for all selected FRET pairs in the presence of AMPPNP or ADP and presented as average values for the indicated domain pairs (for pairwise values see Figure 74). See Table 4 for the complete distance distribution networks.

Figure 49 left shows the final dynamic ensemble for the closed state. The average closed structure in solid colors was generated from the measured distance network for AMPPNP-bound Hsp90. The four ±1σ-snapshots (i.e. snapshots representing the standard deviation σ of the Gaussian distributed distance fluctuations) were extracted from a distance-constrained MD simulation that had been started from the average closed structure.

In contrast to the closed state, the dynamic ensemble of Hsp90's open state was generated directly from the measured distance distribution network by a global analysis, whereat MD simulations were only used for structural refinement. First, the number of rigid domains and the domain compilation is optimized based on the experimental mean distances and their uncertainty (see Chapter 4.3). Notably, the best matching solution was obtained with the N-domains from the crystal structure of an isolated ADP-bound N-domain (pdb: 2wep(22)) and not from the AMPPNP-bound closed structure (pdb: 2cg9). For further refinement and inclusion of loops, which are not present in the crystal structure, all-atom MD simulations were employed to generate structural models with bound ADP (Chapter 3.3). The final model structure matched the distance network with a χ^2<1 and an RMSD of 4 Å. The same approach was used to derive all-atom model structures for the Gaussian intervals (+2σ, +1σ, -1σ, -2σ) of the distance distributions. The resulting structural ensemble represents a possible average trajectory of the open state (Figure 49 right). The ensemble demonstrates the global fluctuations in presence of the model client Δ131Δ, because in presence of the model client, the distance fluctuations are constant for timescales faster than one millisecond.

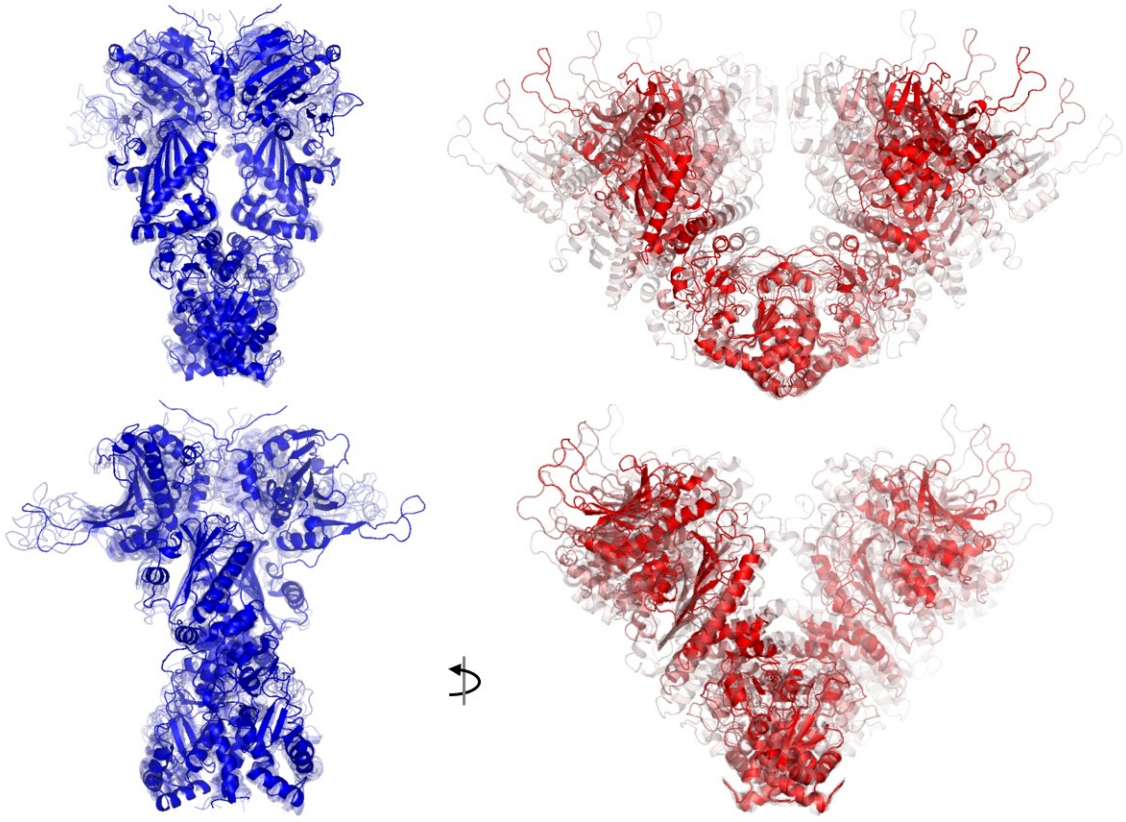


Figure 49: Dynamic ensembles of Hsp90 structures representing 1ms-snapshots. The structural ensembles were generated from the experimental distance distribution networks for the closed AMPPNP-bound state (left) for the open ADP-bound state (right). The mean structures are shown in solid color, whereas the fluctuations represented by the *σ*-structures are depicted in transparent colors. The experimentally arranged ATP-bound closed conformation equals the experimentally arranged AMPPNP-bound closed conformation within 2.0 Å. The mean and σ-structures match the experimental distance distributions with an RMSD of 4 Å or better.

In the following, the open model structures of yeast Hsp90 are compared to existing x-ray structures of Hsp90. Superposition of the average open structure with the closed x-ray structure at the C-domains shows an opening perpendicular to the N-terminal dimerization interface and an additional twist (Figure 50 top left). Prominent rearrangements are a rotation of the N-domain of 90° relative to the M-domain (see also Figure 52 bottom left) and a shift of the CM-interface (see also Figure 52 bottom right). The NM-arrangements (i.e. the conformation of N-domain and M-domain) of the average open structure matches the x-ray structures of the mitochondrial homologue Grp94 (pdb: 2o1v (*95*)) with an RMSD of 3 Å (Figure 50 top right).

Interestingly, the full-length -2σ-structure of the open state even resembles the overall Grp94 structures with an RMSD of 4.5 Å. Certain FRET pairs between the N-domains reveal a very small third FRET efficiency population that fits to the structure of ADP-bound Grp94. This Grp94 like state (further referred to as semi-closed state) is similar to the AMPPNP-bound conformation, but with rotated N-terminal domains. These findings are not very surprising as Hsp90 is very well conserved among homologues. More interestingly, this population is very sensitive to salt concentrations (see Chapter 4.4.4).

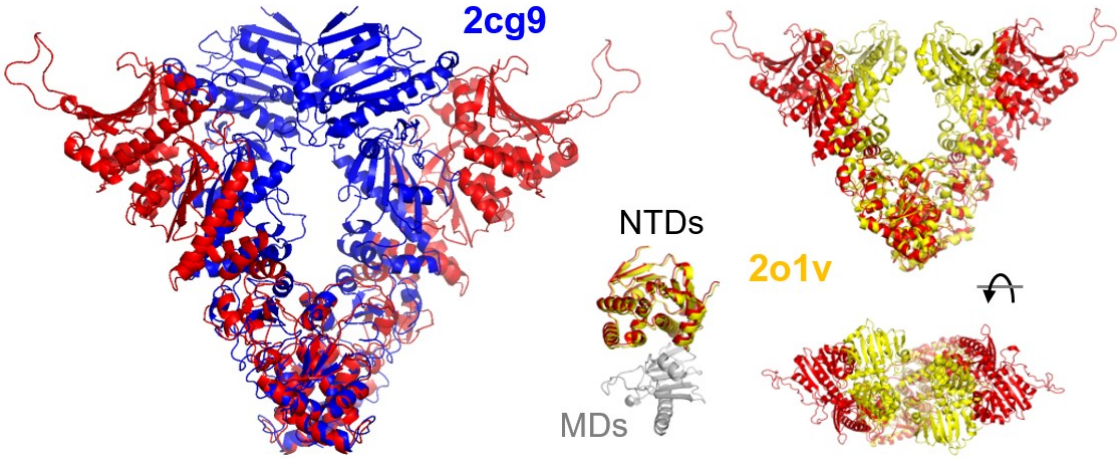


Figure 50: Left: Superposition of the average open structure of yeast Hsp90 (red) with the closed x-ray structure of yeast Hsp90 (blue) at the C-domains. Right: Superposition of the

average open structure (red) and the x-ray structure of ADP-bound Grp94 (yellow, pdb: 2o1v). The conformations between M- and N-domain match with an RMSD of 3 Å.

Fast inter-domain dynamics are determined by correlating FRET efficiencies and fluorescence lifetimes (*76, 77*). Several FRET pairs between the M- and N-domain within one monomer have been analyzed (one is depicted in Figure 51). The projection of the dynamic population of ADP-bound Hsp90 (red) to the static line (green) results in a FRET efficiency that is similar to the one of the AMPPNP-bound closed conformation (blue). A deeper analysis shows sub-millisecond dynamics involving a small population of an N-terminal rotated state (Figure 51 right). Interestingly, the overall orientation of the N-domain in this intermediate state (Figure 51 left, purple) is similar to the arrangement of the closed conformation. This observation indicates a rotation of the N-domain before closing, which has been previously suggested (*96*). Interestingly, in presence of ATP and p23, intra-monomer distances between N- and M-domain reveal an increased population of such an NM-arrangement like in the closed AMPPNP-bound conformation. Remarkably, the estimated relaxation rate in the range of 10 ms^{-1} is possibly correlated with the dynamics of the ATP-lid that were recently found to be in the range of microseconds to milliseconds with NMR relaxation experiments (*21*). Interestingly, the NM-interface was found to be even more flexible without any nucleotide than in the presence of ADP. Additionally, the open structure indicates a stabilizing contact between the open ATP-lid and the M-domain (see below). Thus, one can speculate that ADP binding induces ATP-lid opening, which in turn stabilizes the NM-interface as in the open dimer conformation. The transfer efficiencies for a FRET pair spanning between N- and M-domain within a monomer show nucleotide dependent broadening (Figure 76, appendix).

Within the closed state, the NM- and NN-arrangements are nearly rigid at the observed timescales. The CM-interface is quasi-static above 0.1 ms in the closed state, but features very fast dynamics far below 0.1 ms. This could be a reason for the asymmetry at the CM-interface observed in the x-ray structure of the homologue Trap1 (*97*). The CC-interface remains rigid for all conformations at the explored timescales, which is indicated by equal static distances between both C-domains for all measured nucleotides.

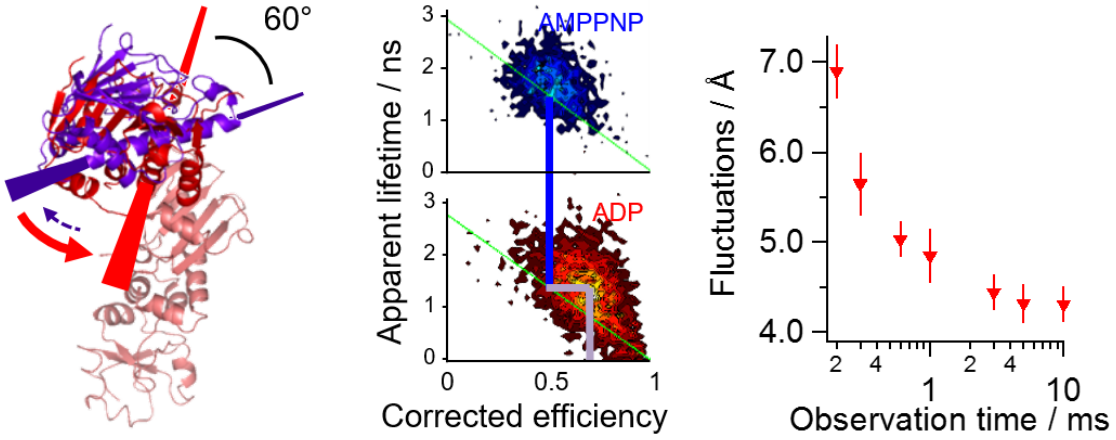


Figure 51: A dynamic intermediate state of the N-domain relative to the M-domain (left, purple) has been uncovered. Apparent donor fluorescence lifetimes were correlated with efficiencies for several FRET pairs spanning between the N- and M-domain within one monomer. Within the open ADP-bound state, a dynamic population (middle, red) can be observed that deviates from the static line (green). This population involves kinetics in the range of 100 µs indicated by a time-correlated distance distribution analysis (right). When the measurements of all intra-monomer FRET pairs were combined, a dynamic intermediate state with the N-domain being orientated similarly to the AMPPNP-bound closed conformation was obtained (left, purple).

To test the stability and dynamics of the open Hsp90 structure, unrestrained MD simulations in explicit solvent were performed by Martin Zacharias and Florian Kandzia, starting with the experimentally arranged average open structure. The RMSD between the simulated open structure and the start structure reached a plateau regime of 6-8 Å (Figure 77, appendix). The structures sampled during the simulations were compared to the distance network with respect to the dye accessible volumes resulting in a maximum χ^2 of 1.05 indicating good compatibility with the FRET data. Root mean square fluctuations (RMSF) are calculated and color-coded in Figure 52 top right. Notable flexible elements are the α-helix 7 acting as a small hinge within the middle domain, the α-helices 13+14 and the α-helix 15. The complete N-domain reveals fast rotational fluctuations while maintaining its mean position over longer time periods.

Analogous MD simulations were performed for the closed conformation and resulted in an overall rigidness (Figure 52 top left), but a slight tension was observed at the CC-interface, hinting towards a less stable dimerization at the CC-interface within the closed AMPPNP-bound state.

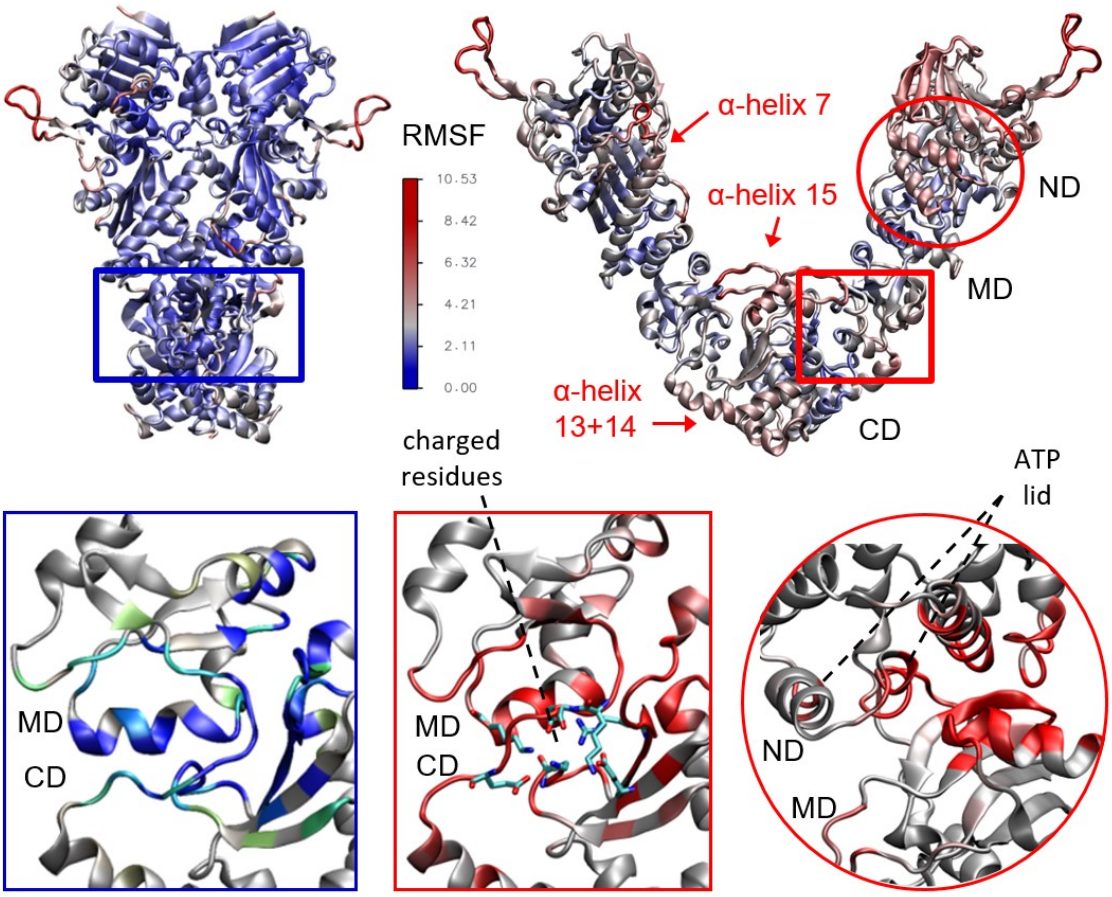


Figure 52: Top: Root mean square fluctuations were calculated from MD simulations for the closed state (top left) and the open state (top right) of Hsp90 indicating flexible elements that are color-coded from blue (no fluctuations) to red (large fluctuations). Bottom: Dynamic regions were further quantified by calculating the number of contributions per residue to the buried surface area (i.e. the contact probability) for the CM-interface in the closed state (bottom left) and in the open state (bottom middle) and for the MN-interface in the open state (bottom right) and are color-coded bluish for the closed state and reddish for the open state. Interfacial charged residues forming salt-bridges are highlighted as sticks.

The interfacial flexibility within Hsp90's open state likely enable the large conformational changes of this multi-domain protein. Therefore, the buried surface area were calculated for the CM-interface in the closed state (Figure 52 bottom left) and in the open state (Figure 52 bottom middle), as well as, for the NM-interface in the open ADP-bound state (Figure 52 bottom right) by Florian Kandzia. The normalized buried surface areas per atom are shown in the appendix in Figure 78 for the open and closed state.

The NM-interface of the open dimer state completely differs from that of closed dimer state. The ATP-lid remains in its open conformation while contacting the M-domain (Figure 52 bottom left). Interestingly, the closed conformation of the ATP-lid is sterically not accessible in the open dimer state (Figure 53 right). This agrees well with the previous suggestion that the ATP-lid might

mediate the NM-arrangement (*98*). Additionally, the open ATP-lid conformation is sterically not accessible in the closed dimer state (Figure 53 left). As ATP binds also to the open dimer state (but with less affinity than in the closed state and much less affinity than for ADP (*99*)), one can speculate that ATP-biding in the open dimer state facilitates the rotation of the N-domain before closing. Future FRET measurements at the ATP-lid will dissect such a possible mechanism.

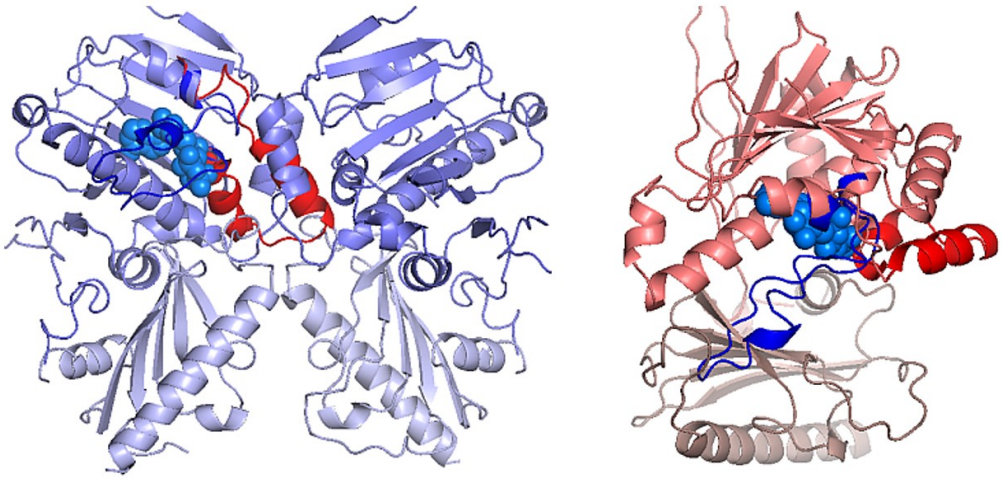


Figure 53: Left: Superposition of the closed Hsp90 dimer (bluish) with the N-terminal domain from the open Hsp90 dimer (violet + red) indicating that the open ATP-lid conformation as in the open dimer (red) is sterically hindered within the closed dimer. Right: Superposition of the open Hsp90 dimer (reddish) with the N-terminal domain from the closed Hsp90 dimer (light red + blue) indicating that the closed ATP-lid conformation from the closed dimer (blue) is sterically hindered within the open dimer state (right).

At the CM-interface a clear shift of interface contacts from the closed to the open state is observed. The α-helix 10, the loop 568-575 and the loop 524-533 contribute mostly to the CM-interface, which conforms well to hydrogen exchange experiments (*25*). Remarkably, the CM-interface in the open state seems to be stabilized by (complex) salt bridges between charged residues (highlighted in Figure 52 bottom middle) that are not present in the closed state. During the MD simulation of the open structure, several charged residues that are conserved among homologues capture interatomic distances of 2.9 Å (Figure 54). This is characteristic for ionic or hydrogen bonds (*100, 101*). Most of these charged residues show (anti-) correlated switching suggesting a locally flexible, but globally stable interface. Such a fuzzy region can explain the experimentally observed inter-monomer fluctuations of the Hsp90 dimer. Point mutations of those charged residues may change these fluctuations and also the overall equilibrium.

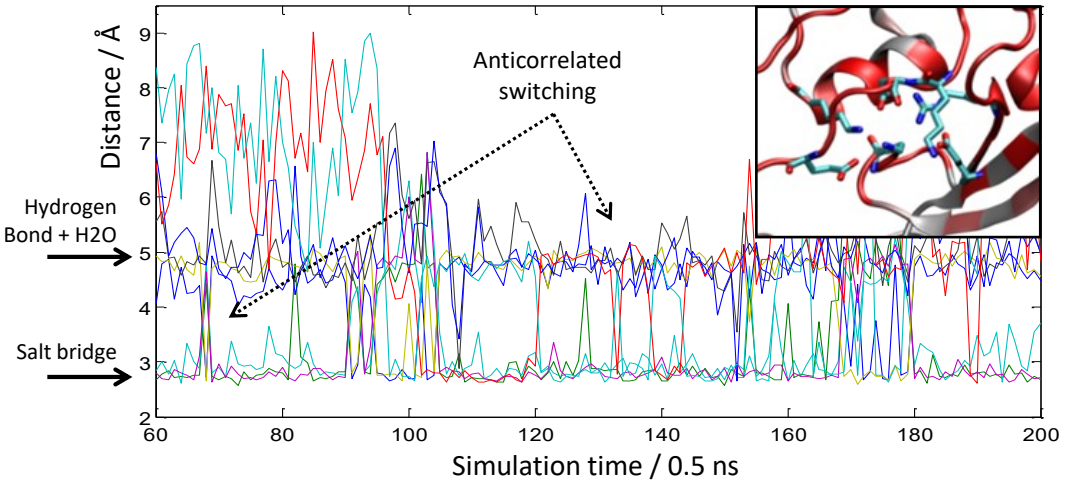


Figure 54: Distances between charged residues that are located at the CM-interface (see also Figure 52 bottom middle) are plotted versus simulation time to test for salt bridges. During the MD simulations, several charged residues captured relative distances of 2.9 Å, which is characteristic for ionic or hydrogen bonds. These bonds are stable for long time periods of several 10 ns and seem to be protected from water. Some salt bridges interchange in an anti-correlated pattern. A second plateau at 4.5-5 Å could be explained by hydrogen bonds interacting with an enclosed water molecule.

Figure 55 highlights interesting surface patterns that were sampled during the non-constrained MD simulations for the closed state (left) and the open state (right) of Hsp90. The upper plots indicate negative (magenta) and positive (cyan) surface charges. The open structure features a contiguous pattern of positive charges across C- and M-domains. The lower plots indicate hydrophobic residues color-coded in reddish. The closed structure and particularly the open structure feature a contiguous pattern of hydrophobic residues at the inner protein surface spanning C- and M-domains. These patterns are probably important for client binding. See Figure 79 for more snapshots and comparison with homologues.

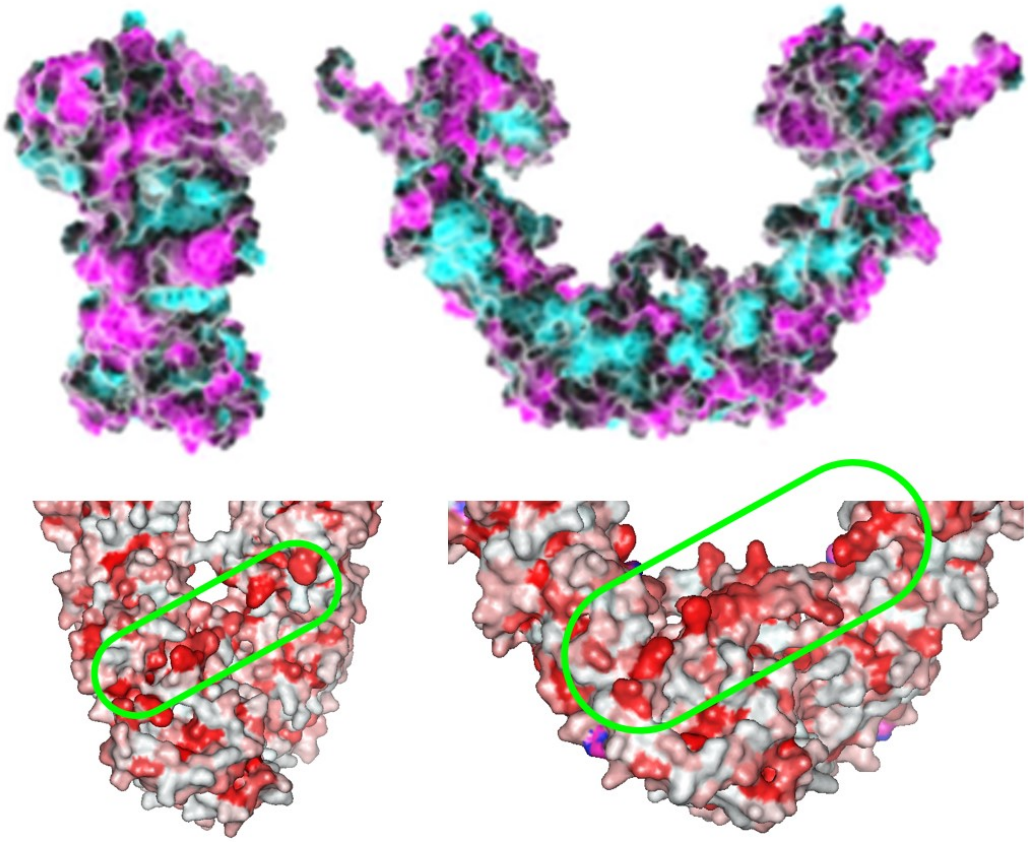


Figure 55: Surface plots from MD simulations for Hsp90's closed (left) and open structure (right) indicating charged residues (top) and hydrophobic residues (bottom). Negative charges are color-coded in magenta, positive charges in cyan and hydrophobicity in red.

4.4.2 Mechanistic model

Figure 56 summarizes the overall mechanism based on experimentally determined dynamics. The closed state (blue) is mainly stabilized by direct contact between the N-domains. The open state (red) is more flexible and stabilized by NM- and CM-interfaces that significantly differ from the closed state. The CC-interface remains globally unchanged during the global conformational changes. The large inter-monomer fluctuations within the open dimer state are presumably caused by the dynamic CM-interface. Large dwell times of minutes to reach the AMPPNP-bound, or rather ATP-bound, closed conformation (thus representing the hydrolysis competent conformation) are likely caused by a small probability of both N-terminal domains being in close proximity at the correct orientation. Equilibration kinetics of Hsp90 after the addition of nucleotides were observed with confocal-based single-molecule FRET because of its ability of detecting small conformational changes (see Chapter 3.2.6). When considering all possible open conformations as one global state, the average closing times was 8±4 minutes for AMP-PNP and 3±2 minutes for ATP (Figure 57).

The stabilizing effect of AMPPNP is visualized by plotting the Gibbs free energies (Figure 56 right) using equation (10). Compared to ATP, in presence of AMPPNP, the global equilibrium is

clearly shifted towards the closed state (blue). By contrast, for Δ131Δ a different mechanism is observed. In presence of Δ131Δ the fluctuations within the open state (red) are significantly reduced on fast timescales (see also Figure 47). Therefore, the energy barriers within the open state are reduced resulting in a more defined conformation (purple). The open-close equilibrium is hardly affected by Δ131Δ, probably because the dissociation rate of Δ131Δ is faster than the transition rates of the open-close equilibrium. The observed timescale of Hsp90's fluctuations and client interaction is similar to the timescale that has been observed for enzyme domain motions with NMR relaxation experiments (*102*).

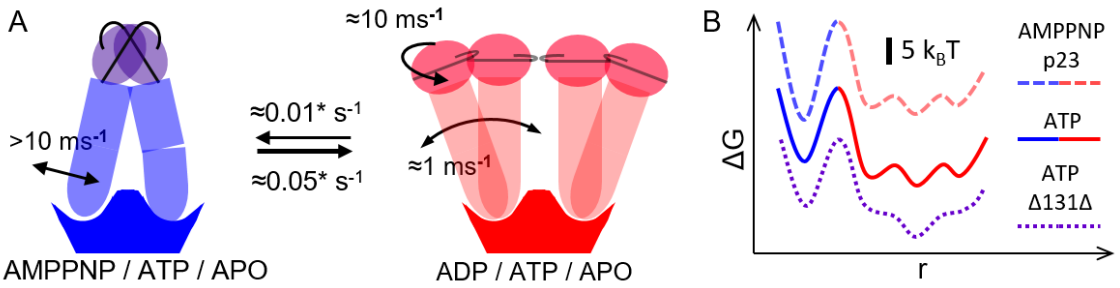


Figure 56: Nucleotide and client dependent inter-domain dynamics of Hsp90. The closed state (blue) is mainly stabilized by direct contact between the N-domains. The open state (red) is much more flexible, mainly stabilized by the CM-interface, which significantly differs from the closed conformation. The N-terminal beta-sheet strand is depicted in black. The determined rates are presented as orders of magnitudes. The transition rates between open and closed states hold for ATP. The states are accessible with the indicated nucleotides. (B) Gibbs free energies are estimated for Hsp90 in the presence of ATP (solid line) or p23 + AMPPNP (dashed line) or Δ131Δ + ATP (dotted line). In presence of Δ131Δ the global equilibrium remains nearly unchanged, but the fluctuations within the open state are reduced resulting in a preferred conformation.

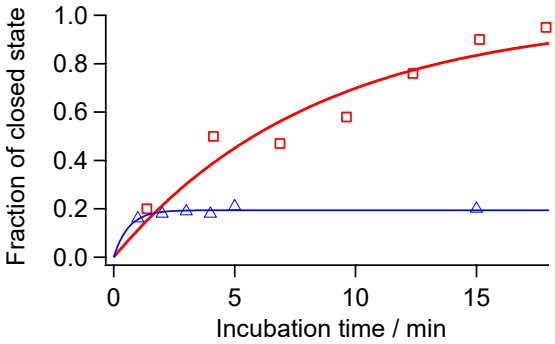


Figure 57: Equilibration kinetics upon nucleotide addition starting from the open APO state of Hsp90 observed with single-molecule FRET. The single-exponential fits revealed relaxation times of 8±3 min for AMPPNP (red) and 0.6±0.3 min for ATP (blue) and an equilibrium population of the closed state of 0.95±0.1 for AMPPNP and 0.2±0.02 for ATP.

Inter-monomer fluctuations

The inter-domain motions that have been measured with the time-resolved distance distribution analysis (Figure 47 right) might agree with some slower modes of motions that have been partially observed within the MD simulations implemented for Hsp90's open state. In addition, several complementary fluorescence based methods were applied and refined (see Chapter 3.2) to further scrutinize the inter-monomer fluctuations. Figure 58 exemplarily shows three complementary kinetic analyses for a FRET pair spanning both M-domains of Hsp90 in the closed state (blue), in the open state (red) and in the open state in presence of 20 µM of the client protein Δ131Δ (purple). Several control measurements have been done that exclude photophysics as a source for the observed fluctuations. Additionally, implausibility of such artifacts is intrinsically given because of the large kinetic differences between Hsp90's closed and open state. The dominating relaxation rates obtained by the modified FCCS analysis agree well with the apparent relaxation rates estimated by the novel time-correlated distance distribution analysis (Figure 47). This novel analysis can be seen as complementary to the FCCS analysis, because it is not affected by intensity fluctuations. And, it reveals additional information about the amplitude of the distance fluctuations.

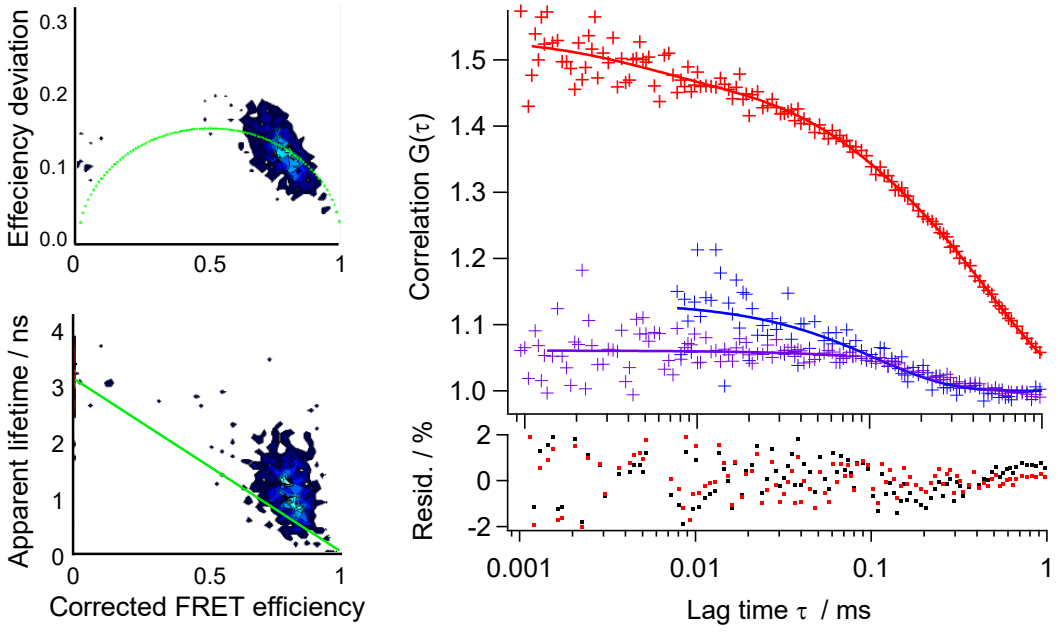


Figure 58: Selection of complementary fluorescence-based analyses for obtaining conformational kinetics in the range from micro- to milliseconds. Burst variance analysis (top left), multi-parameter analysis (bottom left) and fluorescence cross-correlation spectroscopy (right) for a FRET pair spanning the middle domains of the Hsp90 dimer within the open ADP-bound state (red) and the closed AMPPNP-bound state (blue). The burst variance analysis reveals an average relaxation rate between 0.5-5 ms^{-1} for the open state, while closed state is quasi static at that timescale. The correlation between apparent

donor lifetimes and corrected FRET efficiencies reveals rates faster than 10 ms^{-1} for the closed state. Cross correlation and division of certain fluorescence intensity traces (see Chapter 3.2.6) reveals a small populated average relaxation rate of about 10 ms^{-1} for the closed state. For the open state, a three-term exponential fit reveals a relaxation rate of about 2 ms^{-1} and much smaller populated relaxation rates of about 20 ms^{-1} and 200 ms^{-1} (red curve), while a two-term exponential fit reveals a dominant relaxation rate of about 2.5 ms^{-1} and a small populated relaxation rate of about 50 ms^{-1}. Residuals between measurement and fit are shown for the open state for the three-term fit (red) and for a two-term fit (black). A power law does not fit to the complete time range, but to correlation times smaller than 100 µs. For the open state in presence of 20 µM of the client protein Δ131Δ, a single-term exponential fit reveals a small populated rate of 2 ms^{-1}, which is about 10 times smaller populated than in absence of client protein.

If one assumed that the two Hps90 monomers moved independently from each other, one would obtain a mean square displacement of about $<x^2>$=100 Å² at 100 ns according to Stokes-Einstein relation. This would be a good assumption if the monomers were connected by a flexible linker. However, the mean square displacement would be averaged at the timescale of observation (>100 µs), unless the energy is somehow damped due to viscosity or internal friction of the protein.

In order to answer the question whether the inter-domain fluctuations within the open state originate mainly from substates and/or subdiffuion, they were included in Monte Carlo simulations (see Chapter 4.2.2). Based on different kinetic models, time-correlated FRET efficiency histograms have been simulated for comparison with the experimental data and for finding most likely mechanisms.

Figure 59 illustrates the two most likely models that well reflect the observed inter-monomer fluctuations and their time dependence. Figure 59 A fits to model D describing a continuous random walk due to a high-viscous diffusive process within a harmonic potential. It can be seen as a local Gaussian process of equilibrium fluctuations covering a broad range of timescales.

Figure 59 B fits to model C that describes two independent sources of Gaussian distributed fluctuations occurring at two different timescales. The transferred mechanism could be the following: Each monomer independently switches between two states on slow timescales (0.1 ms^{-1}) while being rigid in one of both states and quickly fluctuating (10 ms^{-1}) in the other state. This would involve asymmetric intermediates but an overall symmetry on larger timescales. An alternative mechanism for model C would be a continuous asymmetry meaning that one monomer quickly fluctuates and the other one is more rigid, while they interchange their behavior on slower timescales. Such a coordinated asymmetry could be mediated by the flexible alpha helix 15 (see Figure 52B top left) or another asymmetric local conformational change of the C-domains.

The simulations demonstrated that the observed broad distance distributions and their time-correlation cannot be explained by two or three rigid substates. The distributions are described better by single Gaussian processes. One may ask whether these distance fluctuations at the timescale of 100 µs can be described with a (single) diffusive process. The fact that the protein's motion is shifted to much larger timescales than thermal fluctuations has been described with the Langevin equation and fractional Brownian motion (*103*). The Rouse or Zimm model used in polymer physics describe a similar behavior with a diffusion coefficient that is anti-proportional the number of substructures (*104*). In contrast to elastic network models, one might consider recently often discussed glass-liquid or jamming transition at the protein-interface that can explain very high viscosities (*105, 106*). Appropriately, it has been proposed that ligand induced conformational changes operate in a viscoelastic regime (*107*).

Alternatively, the microscopic origin of the fluctuations in Hsp90's open state is caused by real structural rearrangements represented by multiple energetically equivalent (irreversible) conformational substates. According to Arrhenius, the distance fluctuations observed at a timescale of about 100 µs, would correspond to an (fluctuating) energy barrier of more than 5 k_BT.

At the end, the measured distance distributions likely represent a mixture of substates and subdiffuison. FCCS revealed several rates on a broad range of timescales (1 µs – 1 ms), while the smaller timescales below 100 µs can be fitted with a power law, too. This scenario could be explained by superposition of Markovian fluctuations at larger timescales and subdiffusion at smaller timescales, or, by a random walk through a fractal-like structure. It seems likely that the open dimer interchanges between local energy minima that are probably caused by structural rearrangements at the CM-interface (possibly due to switching salt-bridges or structural ambiguity), whereat faster subdiffusive inter- and intra-domain motions occur around each conformational substate. In this case, the Gaussian-like shape of the distance distributions might be caused by the interchange between one optimal conformation of the CM-interface and several surrounding conformations with decreasing affinities (due to loss of interface contacts) and decreasing dwell times. This is illustrated by the half gearwheel in Figure 59A.

Temperature-dependent measurements may clarify this issue (see Chapter 4.4.4). In the future, Hsp90 will be immobilized in a confocal setup to cover the contiguous time range between 100 µs and 1 s, and to further encircle the real model.

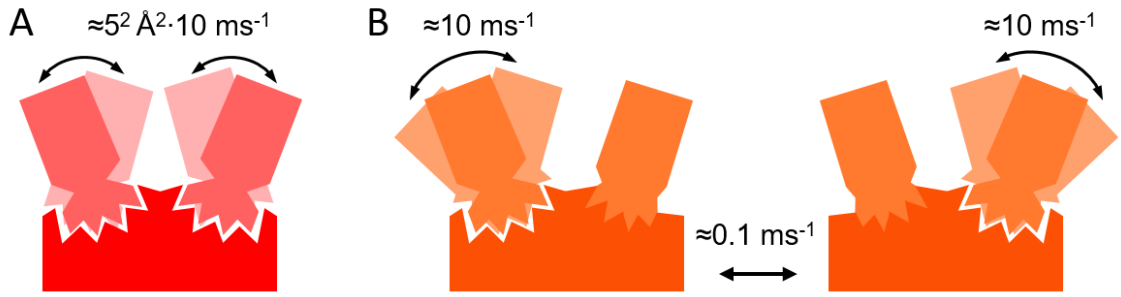


Figure 59: Two likely mechanisms for Hsp90's open state based on the generated structural ensemble and on the comparison of measured time-resolved distance fluctuations with simulated kinetical models that include experimental kinetics. C-domains are colored in dark red and M-domains in light red. A: Fast fluctuation between several microstates at the CM-interface. B: Slow (coordinated) switching between two globally identical asymmetric states due to a local asymmetric conformational change at the C-domains, whereat each monomer switches between a rigid and fast fluctuating state.

One could also think of transient dissociation at the CM-interface. Some specific measurements between C- and M-domain might hint a fast but infrequent dissociation. However, such a mechanism cannot be solely responsible for the observed distance distributions and there is no increasing probability for low FRET efficiencies with decreasing binning times (see Figure 35). Nevertheless, one could think of a very short-lived dissociation followed by rebinding to another but energetically similar position (imagining a leash between C- and M-domains in Figure 59A). Further FRET measurements with an inserted linker between C- and M-domain and increased viscosity could shed light on this vague presumption. Additionally, optical or magnetic tweezers based force distances measurements could help to narrow the possible binding free energy.

A particularly interesting aspect of the observed state-dependent fluctuations is that they are independent from nucleotide states. Nucleotides and co-chaperone probably change the overall rates only by inducing rearrangements between NN- and NM-domains. The CM-interface could function as a basic oscillator quickly fluctuating between a very open and nearly closed conformation. In other words, it samples a huge conformational space on a fast timescale, while the N-domains mediate the limiting rate for the large conformational changes (and probably also hydrolysis). Furthermore, such a mechanistic concept can explain communications between domains that are far away from each other – generally called allosteric effects. This could be realized by changing the flexibility of the CM-interface, for example, due to co-chaperone binding, which in turn shifts the overall equilibrium.

Finally, in the future, the results might be directly compared with other dynamic structure determination methods. The overall dynamic structure could be compared with data from small angle x-ray scattering and the local flexibility could be verified with data from nuclear magnetic resonance spectroscopy. Hydrogen exchange experiments on Hsp90 (*25*) gave some hints to inter-

domain flexibility agreeing with the results presented here. One could perform more specific measurements not only in dependence of nucleotides but also in dependence of client proteins, while keeping an eye on the CM-interface. The dissociation rates of the Hsp90 dimer and an Hsp90 complex have been analyzed with microfluidic measurements (see Chapter 6.3).

4.4.3 Local conformational changes

After having solved the global structures, the distance networks allow local fluctuations of small elements to be detected. This enables investigating the temporal and functional correlation between these local conformational changes and the global states.

The movement of the N-terminal beta-sheet strand (amino acids 1-9) was examined by selecting the position 9 and determining the distance to the position 452 that had been analyzed within the distance network. While forming a stable cross-monomer contact in the closed AMPPNP state, the N-terminal beta-sheet strand binds to its own domain in the open ADP-bound state (Figure 60 left). Interestingly, it reveals faster dynamics than the neighboring beta sheet (right), probably due to ADP-induced docking and undocking of the N-terminal strand. This is in line with previously reported conformational states of the isolated N-domain (*98*). One can speculate that ADP prevents the formation of the cross monomer contact by increasing the binding affinity to the beta-sheet of its own N-domain.

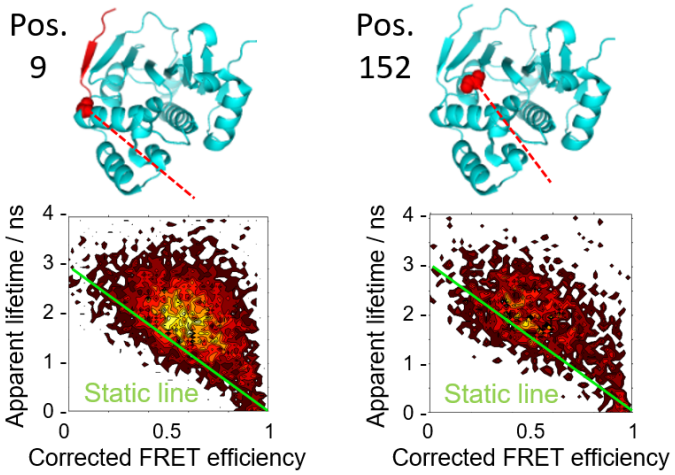


Figure 60: Position and dynamics of flexible elements can be determined by selecting suitable positions from the distance network. The N-terminal beta-sheet strand (left) was found to be more dynamical than the neighboring position within the beta-sheet (right).

As indicated with structural superpositions (Figure 53), the open ATP-lid conformation is sterically hindered within the closed Hsp90 dimer and the closed ATP-lid conformation is sterically hindered within the open dimer. ADP binding may induce ATP-lid opening, which in turn may stabilize the NM-interface as in the open dimer conformation. This is indicated by nucleotide dependent dynamics between N- and M-domain (see Figure 76, appendix) and by MD

simulations of the open structure (see Figure 52 bottom right). In presence of ADP, the hydrolysis competent closed state is most likely not accessible, because ADP binding proabably opens the ATP-lid and thus sterically hinders the dimerization of the N-terminal domains. After hydrolysis, the remaining ADP might cause opening of the ATP-lid and thereby mediates, inter alia, the opening of the dimer. Such correlated conformational changes and possible intermediates will be dissected with a further dye at the ATP-lid.

A further interesting local element is the alpha helix 15 (see Figure 52 top right). The later was found to be very flexible during MD simulations and captured several interesting positions. It could cause an asymmetry of the homo-dimer by binding to the opposite monomer and thereby hindering the other alpha helix 15 of the opposite monomer to capture the same and thus symmetric orientation. Alternatively, it could bind to the M- or N-domain of the same monomer and thereby stabilize the open conformation. Possible binding position will be measured with a further dye label at the alpha helix 15, in the near future.

The distance networks are perfectly suited to address further issues in the future, for example, detecting local influences of and real-time interactions with a client protein.

4.4.4 Physicochemical influences

The dynamic structural ensemble and the mechanistic models that have been derived and discussed in the previous section answered open questions and raised also new ones. For a better understanding of the overall conformational stability and the observed fluctuations, the physical and chemical environment of the protein has been systematically varied.

Salt concentration

The population for Hsp90's semi-closed state (see above for its definition) is very sensitive to salt concentration as shown for the FRET positions 61-61 in Figure 61. The efficiency population at E=0.1 represents the open ADP-bound state of Hsp90 and the population at E=0.2 belongs to the closed AMPPNP-bound closed state. When lowering the concentration of potassium chloride from 150 mM (left) to 50 mM (middle), a third efficiency population at E=0.5 became clearly occupied, representing the semi-closed state. Several FRET pairs indicate a close match with the x-ray structure of Grp94 (Hsp90 in the endoplasmic reticulum), which is similar to the closed structure, but with rotated N-terminal domains. In this semi-closed state, the positions 61-61 are vis-à-vis and the average dye-separation is reduced to 60 Å.

The observed conformational shift fits surprisingly well to the fact that there is more potassium chloride in the cytosol than in the endoplasmic reticulum, meaning that conformational changes could be additionally promoted by physiological conditions. However, it has to be noted that the salt conditions during crystallization were probably different. Another interesting observation is that this population vanished, when adding 10 µM p23 – a cochaperone that preferably binds to

the AMPPNP-bound closed state – to the low-salt sample, indicating that the semi-closed state is quite flexible (Figure 61 right).

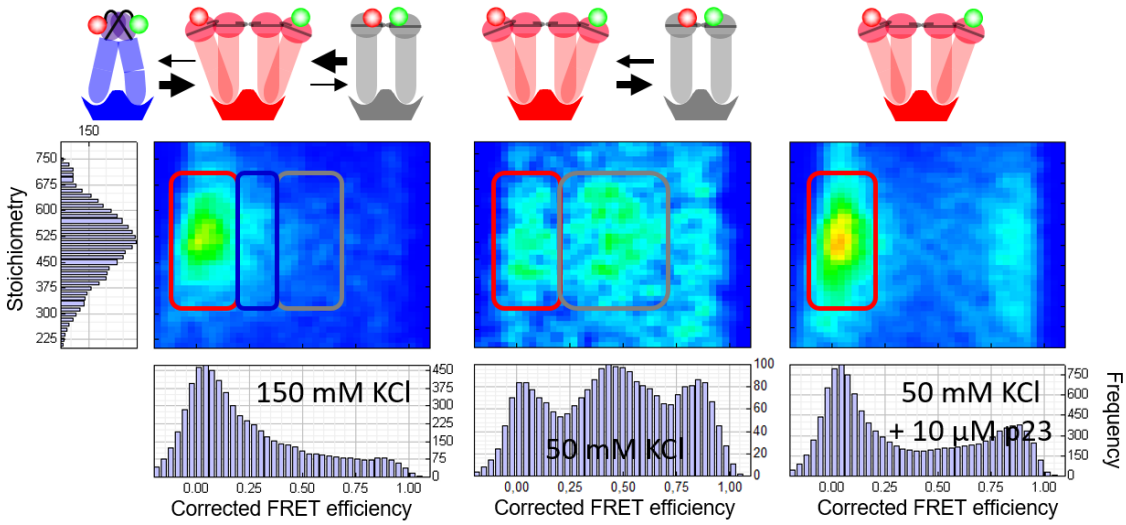


Figure 61: Changing salt concentrations shifted the overall equilibrium of Hsp90. Left: At standard buffer conditions, mainly low FRET efficiencies are observed for a FRET pair located between both N-domains at position 61. These low efficiencies belong to the AMPPNP-bound closed (blue pictogram) and ADP-bound open state (red pictogram). The rare events observed at medium efficiency belong to a semi-closed state (gray pictogram). Middle: This population strongly increased when lowering potassium chloride from 150 mM to 50 mM. In this way, the semi-closed state has been found, the structure of which is similar to the x-ray structure of Grp94. Right: Upon further addition of 10 µM p23, the population vanishes.

Preliminary data revealed such a population shift also due to temperature changes. One can speculate that changes in the population of this state may be the reason for the salt and temperature influences on the ATPase function (*108*) by shifting the overall equilibrium.

Temperature

The fluctuations of the open state were further analyzed by varying the temperature (Figure 62). The fluctuation amplitudes for a certain observation time decrease with increasing temperature. On the first view, this seems contradictory as one would expect a larger conformational space to be sampled with increasing temperature. However, as discussed before, the inter-monomer movements are most likely restricted in a confined potential, so that faster fluctuations lead to stronger averaging within the time window of measurement. If the fluctuations were caused only by a subdiffusion process, a temperature increase of 10 K would accelerate the fluctuations only by a factor of about 1.2 due to the temperature-dependent viscosity of water. But, the measured rate increment factor of 3-4 per 10 K is much larger and suggests an Arrhenius-like increase under the assumption that the activation free energy and the Arrhenius pre-factor are largely

temperature-independent (Figure 62). Assuming a single state, this factor may correspond to an activation free energy of about 80-90 kJ/mol, calculated with equation (11). This might indicate that the inter-monomer fluctuations of Hsp90's open state are not predominantly driven by conformational entropy. Conformational entropy would probably be less temperature dependent.

The relative high activation free energy corresponds to a dwell time of several seconds according to Arrhenius and therefore seems to be in contradiction to the relative fast inter-monomer fluctuations observed at the micro- to millisecond timescale. Even if there were additional rates at much larger timescales in the range of the ATPase rate, they could not lead to the observed time-dependency of the distance distributions at 100 µs. In other words, when assuming an Arrhenius frequency factor of 10^8 s^{-1}, the observed rates would equal an energy of about 15-25 kJ/mol. A frequency factor of about 10^{15} s^{-1} could explain the different energies, but such a high value seems impossible. This seeming contradiction might hint some interesting aspects of the investigated protein, as explained in the following.

Cooperated exchange between local micro states could be seen as a global stable state with an effective activation free energy. Such cooperation might be realized by interchanging complex salt-bridges. A simultaneous increase of several (similar) forward rates k_{1+} for N successive microstates due to temperature increase by $\Delta T=T2-T1$ might potentiate the rate changes in the following manner:

$$f_k(\Delta T) = \frac{k_{eff}(T2)}{k_{eff}(T1)} = \frac{k_{1+}(T2) \cdot (k_{1+}(T2)/k_{1-})^{N-2}}{k_{1+}(T1) \cdot (k_{1+}(T1)/k_{1-})^{N-2}} = \left(\frac{k_{1+}(T2)}{k_{1+}(T1)}\right)^{N-1} \qquad (53)$$

The presumption that the rates in one direction are more affected by temperature than the rates in the other direction could be explained by dissolving bonds, for example, of salt bridges in one direction and friction in the other direction. Remarkably, the activation free energy calculated for the fluctuations of Hsp90's open state is close to the activation free energy of 91 kJ/mol that has been measured for the ATPase rate of yeast Hsp90 (*108*). Additionally, it fits quite well to the slow rates measured for the transition from the open to the closed state of Hsp90 (see above). These observations might underline the speculation that it is not a single activation free energy that is ATPase rate limiting, but rather several dynamic intermediates, represented by conformational substates within the open state. The probability, to overcome several small energy barriers between several substates, can be similar to that of overcoming a single larger energy barrier. In other words, the open state of Hsp90 seems to be globally stable but locally very flexible.

As an alternative to the above described point, the observed time-dependent fluctuations could originate from subdiffusion according to model D (see above). If subdiffusion is originated from viscoelastic rather than elastic behavior of the protein, the time-shift of the subdiffusive process

would be temperature-dependent. The viscous flow in amorphous materials can strongly deviate from the Arrhenius-like behavior, mainly because of the temperature-dependent activation energy (*109*).

It has to be mentioned that temperature may affect the dye mobility, and as a consequence the orientation factor and the Förster radius. However, the dye fluctuations are completely averaged at timescales above 1 µs and thus cannot affect the observed distance fluctuations. Additionally, the time-resolved residual anisotropy remains constant at all measured temperatures, so that eventually longer-lasting dye conformations can be excluded, too. Depolarization effects due to torsional vibrations are quasi temperature-independent within this temperature range.

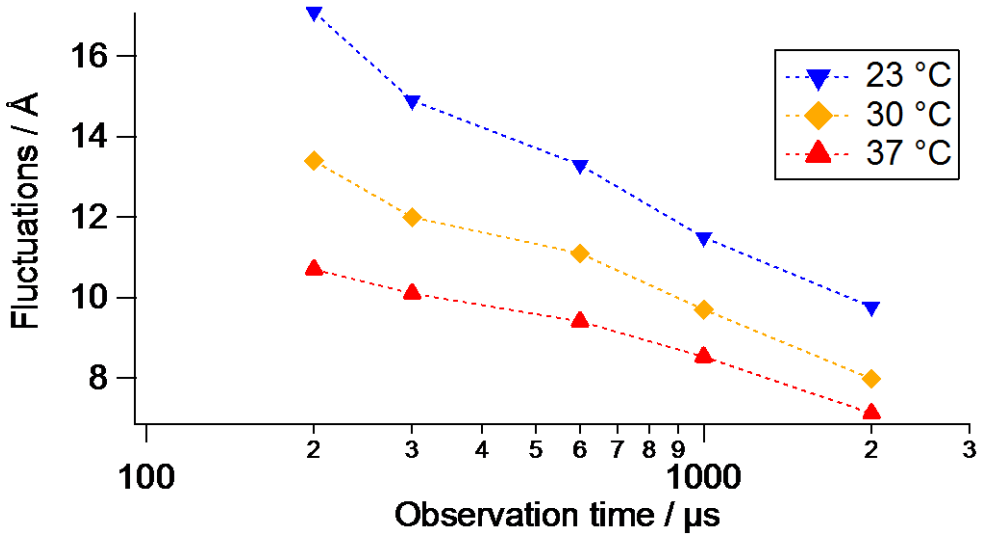


Figure 62: Standard deviation of Gaussian distance distribution (i.e. fluctuation amplitudes) versus photon binning time (i.e. observation time) for different temperatures. For a fixed fluctuation amplitude, the observation time decreases by a factor of 2-3 per 7 K. Increasing temperature leads to faster inter-monomer fluctuations and thus to averaging at smaller observation times.

4.4.5 State observables and distance networks

By comparing the measured distance networks of the closed and open state of Hsp90 with the measured transfer efficiency histograms, a small populated semi-closed state has been found (see Chapter 4.4.1). Its population was strongly increased when changing the salt concentration (see Chapter 4.4.4).

The question arises whether this semi-closed state represents an off-pathway or a committed state to the hydrolysis competent closed state. For the moment, it seems to be off-pathway, because the closed state is sterically difficult to access form the semi-closed state. Transition rates at the timescale of seconds, which were observed for immobilized Hsp90 molecules (*23*), possibly fit

to the transition between the open and the semi-closed state, as the observed efficiency values of the used FRET pair (positions 61-385) were nearly identical for the closed and semi-closed state.

Figure 63 illustrates the open question. The hydrolysis competent state (state 3) may be reached from the semi-closed state (transition state 2 in model E) or from a substate within the flexible open state (transition state 1 in model F). Distinguishing the possible transitions via hidden Markov modeling of fluorescence trajectories is difficult if not impossible if only the two efficiency observables 1 and 2∨3 are given. Both models can generate nearly identical dynamic patterns for kinetic transitions between the two observables 1 and 2∨3. This is because the rate network of model E can be mathematically substituted with the rate network of model F, so that the expectation values of dwell times are completely identical for both models. Only the distribution of dwell times might be slightly broader in model E than in model F for the observable 2∨3. Remarkably, this additional variety of model E will probably steer any maximum likelihood optimization in a wrong direction. This example demonstrates that additional observables are mandatory. These observables are given by the three-dimensional distance networks that were determined for the entire multi-domain protein.

The distance networks are excellent suited to find a FRET pair that allows direct observation of the three states with well-separated expected transfer efficiencies. In this way, the succession of these states can be measured deterministically with single-molecule fluorescence trajectories. Further clarity can be gained, for example, by measuring the ATPase rate of a mutant with a cross-linker between the middle domains that fits to the semi-closed state.

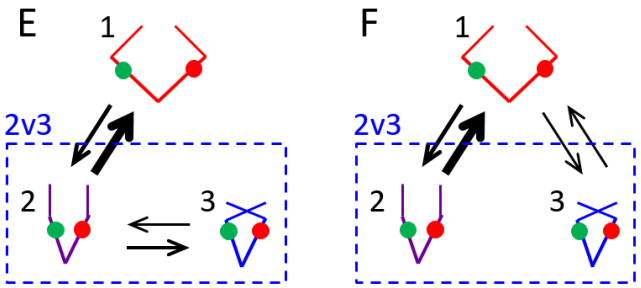


Figure 63: Schematics illustrating two kinetic models for three Hsp90 states observed with two FRET efficiencies. The M-domains are indicated by thick lines and the N-domain by thin lines. States with similar observables might be distinguished kinetically. However, both models can generate a nearly identical dynamic pattern. Substitution of the rate network of model F with the rate network of model E can result in nearly identical dwell times and transition sequence for the two observables 1 and 2∨3. Another observable is required to identify the most likely model and is gained from the distance networks.

The following global mechanism might bring all observations together (see Figure 64). Hsp90 quickly fluctuates within the open state (red) while sampling a large conformational space. Each 10-100 ms both monomers come close to the semi-closed state (gray) with the N-domains not

oriented as in the closed state, indicated by preliminary measurements of immobilized single Hsp90 molecules (with a FRET pair with three efficiency observables) in the confocal setup (see Figure 80). The probability that both N-domains are correctly oriented is relative small, so that the dimer closes several times without N-terminal closing. These conditional probabilities could be the reason why the transition rate to the closed hydrolysis competent state is so small. The indicated rates (for the transitions from the open state to the closed states) might be average values of distribution of rates, instead of single rates, because of the very flexible open state.

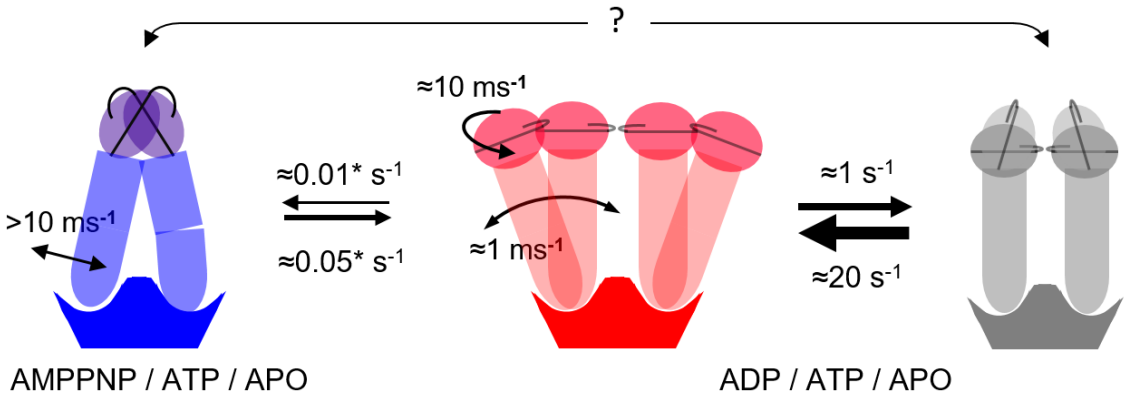


Figure 64: Schematics of probable conformations and dynamics of yeast Hsp90. The N-terminally stabilized closed state (blue) slowly interchanges with the flexible open state (red). An intermediate semi-closed state (gray), which resembles the x-ray structure of ADP-bound Grp94, interchanges on much faster timescales with the open state. The determined rates are presented as orders of magnitudes. The rates between open and closed state (marked with a star) hold for ATP. The states are each accessible with the indicated nucleotides.

4.4.6 Modes of regulation

To demonstrate the applicability of the derived mechanistic model and learn more about functional aspects, preliminary measurements in the presence of different co-chaperones, on the one hand, and in the presence of competing nucleotides, on the other hand, have been done.

The flexible CM-interface likely facilitates a fast sampling of a huge structural space allowing to adapt to certain conformation for binding to co-chaperones and to change the limiting step of the ATPase rate. To concretize this idea, the influence of AHA1 and p23 on the conformational equilibrium of yeast Hsp90 was determined (Figure 65).

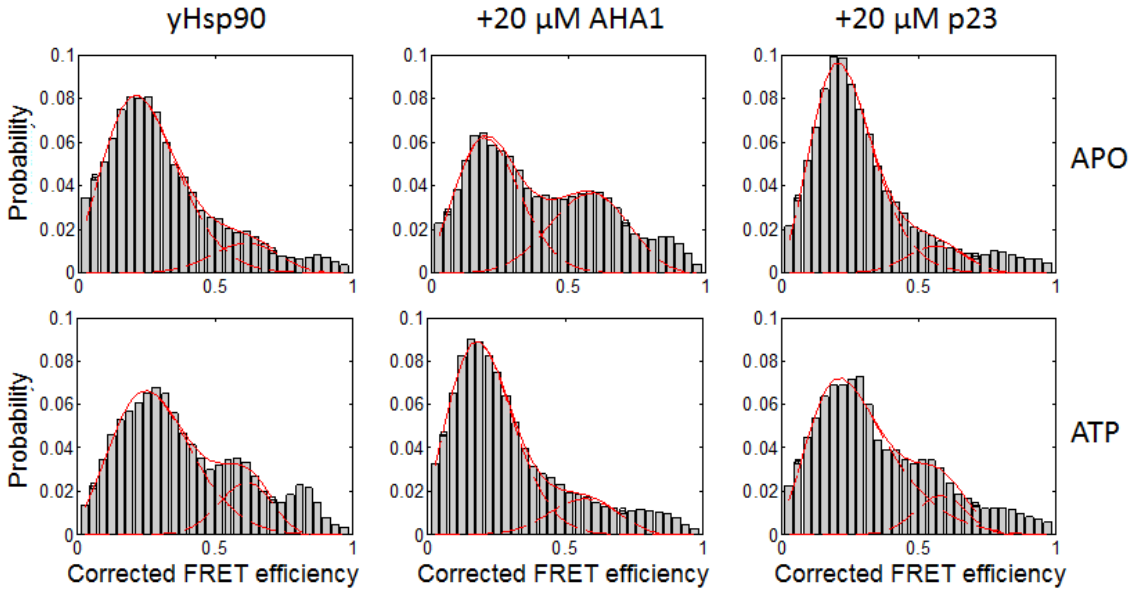


Figure 65: FRET efficiency histograms for yeast Hsp90 labeled at positions 452-452 in presence of the co-chaperones Aha1 and p23. AHA1 increases the closed population (resembling either the closed or semi-closed state) in the absence of ATP and decreases the closed population in the presence of ATP. With p23, however, the equilibrium remains nearly unchanged.

Hsp90 and AHA1

The addition of AHA1 to APO Hsp90 led to a 5-fold increase of the closed population under normal salt conditions. But, in presence of additional ATP, the closed population was reduced compared to the APO state. It seems that AHA1 stabilizes a closed conformation that is quickly depopulated in the presence of ATP but not in APO. Whether it stabilizes the closed conformation or a semi-closed conformation (see Figure 64 for the semi-closed state), has to be dissected with a suited FRET distance. At the moment, it seems to be the semi-closed state, because of the additional fact that the increased binding affinity of AHA1 under low salt conditions could be caused by an equilibrium shift of Hsp90 itself towards the semi-closed state. As the population of this state almost vanishes in presence of ATP and AHA1, it is likely a committing state to hydrolysis, while hydrolysis itself might lead to opening and complex dissociation. Such a commitment can be explained by a local or global mechanism. Both mechanisms presuppose a fast open-close rate of the dimer in presence of only ATP that is at least as fast as the ATPase rate of Hsp90+AHA1.

As a local mechanism, AHA1 could stabilize a conformation of the N- and M-domains that is similar to the conformation of the closed state and thereby increase the transition probability to the hydrolysis competent closed state. Alternatively, it may destabilize the semi-closed state if it is an off-pathway, or (de)stabilize the conformation of a small element such as the ATP-lid.

As a possible global mechanism, AHA1 could simply increase the population of a short-lived semi-closed state and thereby increase the transition rate to the hydrolysis component state if this semi-closed state is on the pathway to hydrolysis. The decrease of the semi-closed population in presence of AHA1+ATP compared to only ATP, could be explained by more than one event that are required for the transition from the semi-closed to the hydrolysis competent state – for example the binding of two ATPs. This is because the probability that two events occur within a time window of one second, for example, can be larger than 5 times the probability that these events occur in 200 milliseconds.

The hypotheses are supported by the fact that the increased ATPase rate in presence of AHA1 is similar to the transition rates previously observed between an open and a (semi-)closed state (*23*) (Figure 64). One could imagine a cascading energy landscape with an AHA1-ATP-Hsp90 complex representing the global minimum. In other words, AHA might stabilize a certain conformation (probably the semi-closed state) and thereby shift the equilibrium towards this state, from which the hydrolysis competent state is reached more often. This way, a flux would be generated without the need of direct activation of the hydrolysis through AHA1. In addition or alternatively, hydrolysis might be coupled to the dissociation of AHA1.

A future hidden Markov analysis of immobilized Hsp90 molecules might narrow down the possible mechanism. Future measurements with labeled AMPPNP will probably clarify whether AMPPNP binds before or after AHA1 association and whether AHA1 facilitates the binding of one or two nucleotides.

Hsp90 and p23

Upon addition of p23 to APO Hsp90, the equilibrium remained unchanged. Yet it is not clear why the steady state ATPase rate is decreased in presence of p23. Interestingly, there was found an equal single-turnover ATPase rate for Hsp90 in absence and presence of p23 (*25*), indicating that p23 binding does not influence the steps leading to hydrolysis. At first glance, one would suggest a delayed ADP release after hydrolysis, for example, because p23 blocks the ATP-lid. This in turn would lead to a much larger closed population than observed for p23 and ATP. On the other side, the affinity of p23 to Hsp90 in presence of ADP is much smaller than in presence of ATP as measured with FCS within this work (*36*). Thus, one possibility that could explain an ATPase limiting step after – not before – hydrolysis would be a stable Hsp90-p23-ADP(-P) complex, which forms (only) due to hydrolysis. However, the rate limiting step for the slow steady state rate could also occur before hydrolysis. It might be possible that the higher ATP concentration in the steady state measurement increases an off-pathway in presence of p23, for instance, ATP and p23 could bind to the open state of Hsp90 in a manner that sterically hinders the closing of the Hsp90 dimer. Future FRET measurements at the ATP-lid could dissect such a possible mechanism.

Hsp90 as an ADP sensor

Hsp90 might be a sophisticated ADP sensor. The overall binding affinity for ADP is about 10 times larger than the affinity for ATP, whereas the state-dependent affinity can extremely differ. It was shown that the affinity for ADP is much higher in the open state than in the closed state and higher than ATP in both states; and the off-rates for ATP in the closed state are much faster than hydrolysis (*99*). One might ask why the affinity to AMPPNP is so large, although apparently, a huge amount of AMPPMP should dissociate before hydrolysis (i.e. before 60 seconds) which would lead to an opening, because the equilibrium of APO Hsp90 lies strongly on the open state. The average closing time of Hsp90 is about 8 minutes upon addition of AMPPNP (see above), while the equilibrium is nearly completely shifted to the closed state (at least 90 %), which would mean an average opening time of at least 80 minutes. Now, to further understand these apparent contradictions, some competition experiments have been done between AMPPNP and ADP. The non-hydrolysable ATP analog AMPPNP is used to mimic all effects happening before hydrolysis. Because the affinities might differ between ATP and AMPPNP, the following argumentation holds only for AMPPNP, but is probably transferable to ATP.

Hsp90 is nearly completely open in presence of ADP (according to an expected efficiency of 0.25) and nearly completely closed in presence of AMPPNP (according to an expected efficiency of 0.6), each under saturating conditions (Figure 66 top). To the AMPPNP-incubated sample twice the amount of ADP is added. After one hour and more, Hsp90 remains unchanged in the closed state, which fits to the huge dwell time of the Hsp90-AMPPNP-complex. When adding AMPPNP to the ADP-incubated sample, a very small but significant closed population appears. The interesting point here is that this population does not further increase. The time of this increase equals the relaxation time when adding AMPPNP to APO Hsp90 (8 minutes). Thus, the small equilibrium population of the closed state is probably not caused (predominantly) by a competition between ADP and AMPPNP before closing. Instead, it seems that ADP competes with AMPPNP within the closed state, but before hydrolysis. And, the first experiment demonstrated that addition of ADP to the AMPPNP-preincubated sample does not lead to an opening after one hour, probably due to solvent inaccessibility in the hydrolysis competent state. All these findings suggest that this hydrolysis competent state is much less accessible in presence of additional ADP. One may speculate that already the binding of one AMPPNP to one monomer can lead to a closing of the dimer, whereat for a probable succession to the hydrolysis competent state, the other monomer must either be in the APO state or must have committed a second AMPPNP (but no ADP); otherwise, the dimer will probably open again (see Figure 67). To test this assumption, future experiments with labeled AMPPNP will be done.

The above findings could indicate that the presence of ADP hinders hydrolysis. One could imagine a mechanism, where holding or folding of client proteins becomes necessary when there is lack of ATP in the cell.

In may be mentioned in this context that the high efficiency peak located at E=0.85 might fit to the structure of the homologue Trap1 (pdb: 41pe), which is similar to the AMPPNP-bound closed structure (pdb: 2cg9), but with a dissociated CM-interface in one monomer.

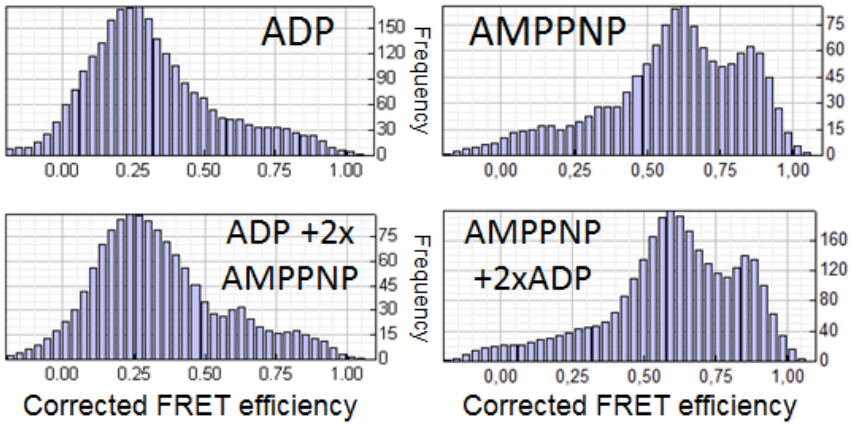


Figure 66: Hsp90 was incubated with 1 mM ADP (top left) and 1 mM AMPPNP (top right) for one hour. Then 2 mM AMPPNP was added to the ADP-incubated sample (bottom left) and 2 mM ADP to the AMPPNP-incubated sample (bottom right) and incubated another hour. The respective efficiencies are 0.25 for the open state and 0.65 for the (semi-) closed state.

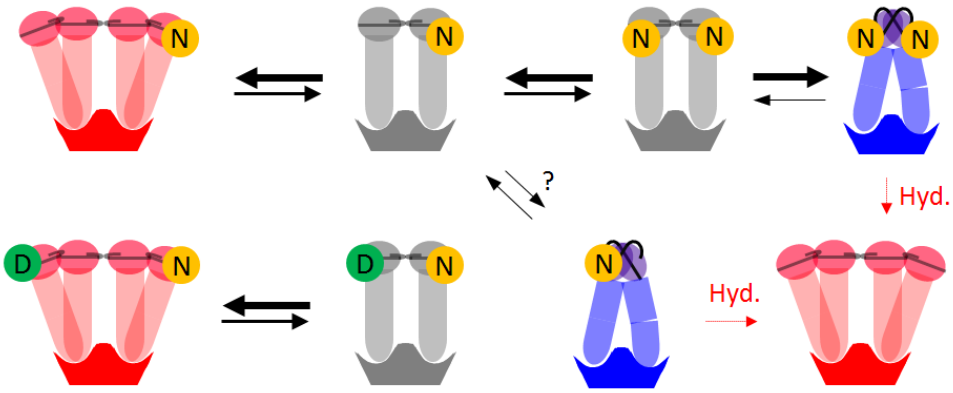


Figure 67: Idea for an asymmetric mechanistic model in presence of AMPPNP / ATP (N) alone (top) as well as in presence of AMPPNP / ATP (N) and ADP (D) (bottom). The hydrolysis competent state (blue) seems to be less accessible when ADP is present before hydrolysis. The arrow thicknesses indicate relative transition probabilities.

5 CONCLUSION

Within this work, several fluorescence based methods were improved or developed with particular focus on their application to the multi-domain protein Hsp90. The results show that Hsp90's open state is highly dynamic on a broad range of timescales. The micro- to millisecond dynamics have a direct and indirect impact on the overall mechanism and function.

The state-dependent global analysis of time-correlated distance distributions has proven to be a powerful approach to gain a comprehensive picture of large-scale and small-scale conformational changes. Slow large-scale conformational changes depend on several small-scale conformational changes and vice versa. While the dwell times of the global open and closed state of Hsp90 most likely depend on at least two local reorganizations in form of conditional probabilities, local inter-domain fluctuations and flexible elements in turn depend on the global state. This approach is most likely applicable to any large multi-domain protein consisting of at least two partially rigid domains.

The success of the experiments can be attributed to optimizations regarding the optical setup and the analysis of fluorescence parameters. Optimal measurement parameters and detection modes were substantiated with an unprecedented large data set of FRET-based intra-molecular distances. Additionally, a microfluidic device capable measuring single low-affinity protein complexes was developed. Thanks to this, structure and dynamics of the Hsp90 dimer were measured under nearly native conditions.

Overall, the complexity of the FRET-based analysis of structural dynamics has been reduced through methodological improvements in the early phase of the experiments, appropriate model simplifications, and a deeper investigation of unexpected outcomes in all phases.

Figure 68 lists the fluorescence based methods utilized and advanced within this work, their respective accessible timescales and the relevant uncovered conformational dynamics of the Hsp90 system. Here are some examples worth mentioning:

i. Very slow transition rates from the open state to the closed hydrolysis competent state were measured with time-resolved single-molecule counting.

ii. Dissociation between both C-domains occurring at the timescale of several ten seconds was measured with a novel microfluidic device.

iii. Major steps ahead in the understanding of the overall mechanism is the determination of fast local and fast inter-domain dynamics in dependence on the global states.

iv. Fast inter-domain dynamics on the (sub-) millisecond timescale were detected within Hsp90's open state between the C- and M-domain – which were found to be mostly independent of the nucleotide state –, and between the N- and M-domain – which were found to depend on the nucleotide state.

v. Correlated local motions of nucleotide binding, ATP-lid opening and N-terminal beta-sheet strand movement on the sub-millisecond timescale were suspected.

vi. Very fast motions on the timescale of nanoseconds were indicated by MD simulations implemented by Martin Zacharias and coworkers using the experimental average structure of the open state as a start configuration.

vii. Fast interactions between Hsp90 and a client protein were detected and seem to be interrelated to Hsp90's broad-band dynamics.

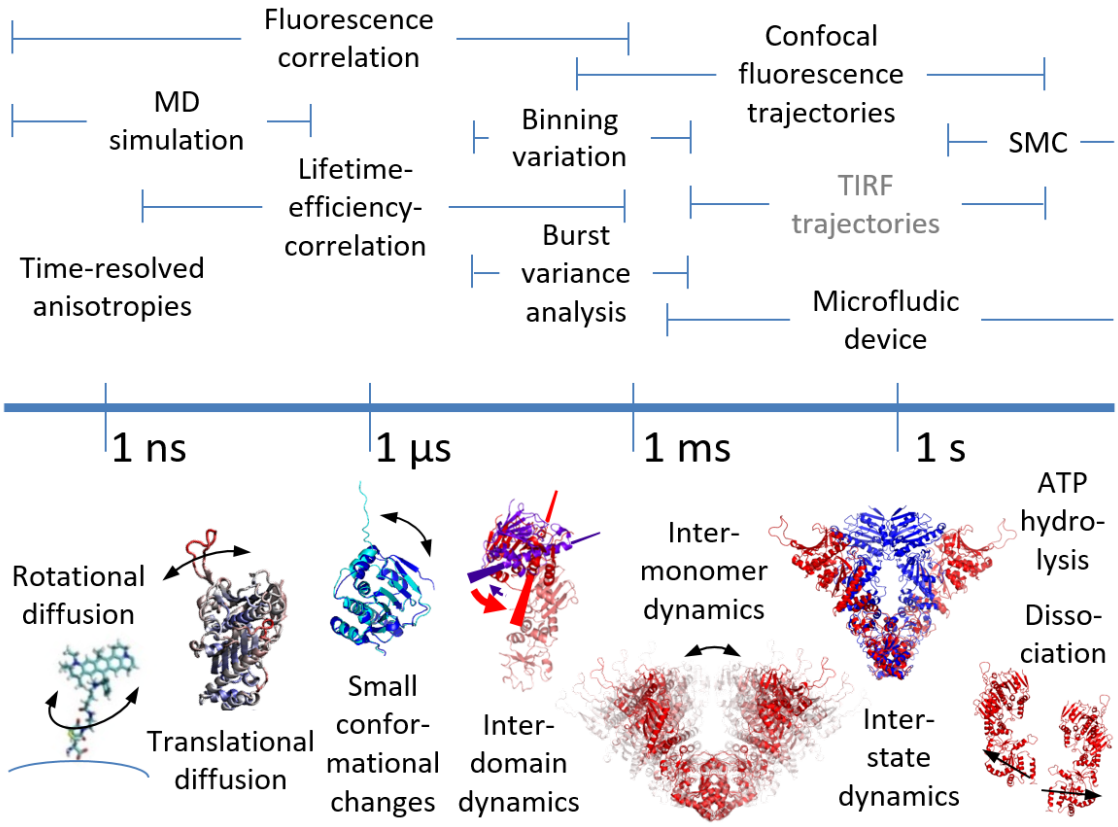


Figure 68: Timescale of Hsp90's conformational dynamics that were analyzed within this work (bottom) and of accessible methods that were advanced and applied within this work (top).

Self-consistent distance networks in combination with time-resolved distance distributions enabled the determination of dynamic structures for the open state of yeast Hsp90. This structural ensemble can be seen as the accessible ensemble in presence of the model client Δ131Δ – a partially folded fraction of staphylococcal nuclease – and demonstrates strong deviations from the conservative static picture. The flexible interface between C- and M-domain features dynamics that enable large fluctuations of the dimer, but simultaneously stabilize the global state for several ten seconds. The combined analysis of the derived distance networks and the time-correlated distance distributions with a Monte Carlo simulation led to self-consistent mechanistic models. The observations are best described by a subdiffusive process around several conformational substates interchanging at the lower millisecond timescale. This might be seen as a random walk through a fractal-like structure. The conformational fluctuations within the open dimer state are mostly originated by a meta-stability at the CM-interface. The CM-interface within the open state consists of partially disordered flexible loops that comprise several charged residues. As indicated by MD simulations, they form interchangeable complex salt-bridges. Notably, these salt-bridges are not accessible within the closed state. The overall interfacial contacts clearly differ between the closed to the open state. Next, the orientation between the N- and M-domain even completely changes from the closed to the open state. Several intra-monomer

FRET pairs between N- and M-domain, as well as, independent MD simulations revealed fast fluctuations of the N-domain within the open state and suggest that the orientation as within the closed dimer state can be quickly sampled within the open dimer state.

The flexible domain interfaces of Hsp90 facilitate fast sampling of a huge conformational space. One can consider the open Hsp90 dimer as a highly dynamic but stable oscillator quickly fluctuating between a very open and nearly closed conformation (Figure 69 left). At the same time, the N-domains feature fast rotational fluctuations relative to the M-domains within the open state. It appears evident that several conditions have to be met within the open state to enable the transition to the hydrolysis competent closed state. The conditional probability that both N-terminal domains are correctly oriented – which is probably facilitated by ATP binding to the open dimer state – while both monomers are in closed proximity is likely small and can explain the slow ATPase rate (Figure 69 right).

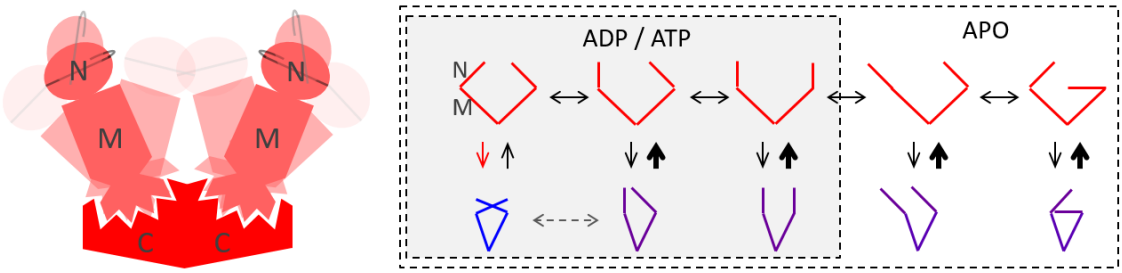


Figure 69: Left: Mechanistic model of Hsp90's open state. The dimer fluctuations are originated by a flexible interface between C- and M-domain. Right: Schematics of a proposed mechanistic subcycle illustrating the interdependency between global states (open state in red, closed state in blue and semi-closed sate in violet) and local relative orientations between M-domains (thick lines) and N-domains (thin lines). ADP or ATP binding reduces flexibility between N- and M-domains. The closed state is not notably accessible in presence of ADP as indicated with a red arrow. The dashed arrow indicates an unlikely transition. The arrow thicknesses indicate the relative transition probabilities. Several possible subconformations likely shift the equilibrium away from the closed state. Co-chaperones can stabilize certain subconformations and thus conveniently shift the overall balance.

The inter-domain dynamics occur at the sub-millisecond timescale and therefore offer an elegant adaption mechanism. One interesting aspect of the observed state-dependent inter-monomer fluctuations is that they are largely independent from nucleotides. Nucleotide binding and hydrolysis probably change the overall rates only by inducing rearrangements between N- and M-domain and thus changing the binding probabilities between the N-domains, while the inter-monomer fluctuations occur on much faster timescales and are not directly influenced by the nucleotides. Similarly, co-chaperones may stabilize a certain global conformation and thereby shift the overall equilibrium and change the limiting step of the ATPase rate. Co-chaperones might

either keep the monomers in closed proximity – i.e. increasing the dwell time in the semi-closed state – or modify the orientation of the N-domains. In other words, the enzymatic function of the Hsp90-co-chaperone-complex could be simply "controlled" by the binding affinity between Hsp90 and the co-chaperone without the need of an exclusive local binding site. Furthermore, such a mechanistic concept can explain communication between domains (here the N-domains) that are far away from each other, for example, in the context of allosteric regulation.

Bringing the observed local conformational changes together, one could speculate the following correlated motion. After hydrolysis, the bound ADP might favor the open ATP-lid conformation and, at the same time, increase the binding affinity of the N-terminal lid – that forms a cross monomer contact in the closed state – to its own domain. Both effects in parallel or in succession, could mediate N-terminal dissociation leading to a rotation of the N-terminal domains and to an opening of the dimer. Within the open state, the ADP induced ATP-lid opening likely stabilizes the NM-interface. This is indicated by nucleotide dependent dynamics between N- and M-domain, by the open structure arrangement and by MD simulations of the open structure. A similar correlated motion could be suggested for the closing mechanism of the dimer, but is probably more complicated, as indicated by competition experiments between different nucleotides.

The measured distance networks were further utilized to investigate (sub)states of yeast Hsp90. It was shown that certain conformations are difficult to distinguish by analyzing only one-dimensional conformational trajectories with otherwise powerful Hidden Markov modeling. Instead, by comparing the determined structural ensembles and the three-dimensional distance networks previously unknown conformations can be uncovered. First, a very small populated semi-closed state was found, which resembles the ADP-bound structure of Grp94, the endoplasmic reticulum Hsp90. Second, dissociation at very short timescales (< 100 µs) has been observed at the CM-interface of the closed state of yeast Hsp90. A structure with such an asymmetric dissociation likely agrees with the x-ray structure of Trap1, the mitochondrial Hsp90. Third, he open x-ray structure of HtpG, the bacterial homologue, lies within the open ensemble of yeast Hsp90. It seems likely that each Hsp90 homologue has a very similar conformational landscape, but with strongly shifted conformational equilibrium. Figure 70 illustrates the global conformational states that have been observed for cytosolic Hsp90. Interestingly, decreasing the salt concentration for yeast Hsp90 shifted the conformational equilibrium towards the semi-closed Grp94-like state, which fits to the fact that there is less potassium chloride in the endoplasmic reticulum than in the cytosol. Changing the equilibrium by the physiological environment might be advantageous for meeting environment specific client proteins.

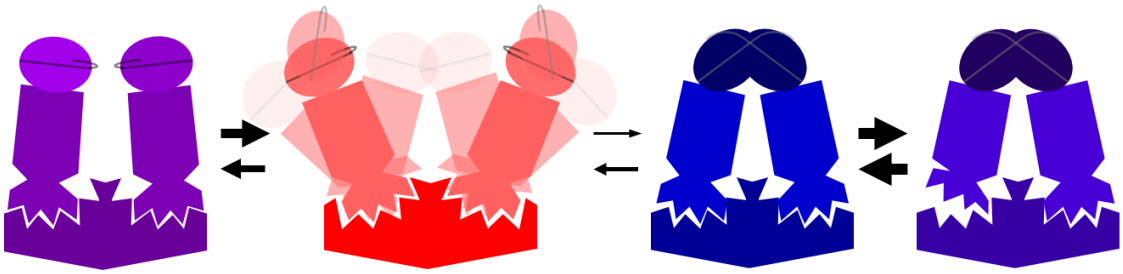


Figure 70: Global conformations of the highly dynamic and ubiquitous Hsp90 machinery. The large-scale dynamic behavior of the measured global states are color-coded from blue (more rigid) to red (more flexible) and the small-scale dynamic behavior of the local domains is color-coded from dark (more rigid) to light colors (more flexible). The four global states were observed for cytosolic Hsp90, and their structures resemble the structures from homologues. From left to right: semi-closed state (resembling the closed Grp94 structure), open state (similar to the open HtpG structure), closed state (2cg9) and asymmetric closed state (similar to the closed Trap1 structure). The arrow thickness is proportional to the estimated order of magnitude of the transition rate (ranging from 10 ms^{-1} to 1 min^{-1}) when ATP is present. Under standard physiological conditions, the equilibrium of yeast Hsp90 is clearly shifted towards the open state.

The dynamic model can explain contradictious observations that have been made by different groups with different methods. They suggest either a strong or a weak coupling between ATP hydrolysis and conformational changes. This contradiction can be simply explained by two substates with similar observables, namely the closed and semi-closed state. It seems likely that the open state frequently interchanges with the semi-closed state, while these transitions are not coupled to hydrolysis. On the other side, there are less frequent transitions between the hydrolysis competent closed state and an open state, while these transitions are probably coupled to hydrolysis. This coupling might be overseen, because of similar observables for the closed and the semi-closed state. In the end, single-molecule studies are more sensitive to frequent transitions and thereby state a weak coupling, while ensemble studies are more sensitive to larger dwell times and thereby state a strong coupling.

The temperature dependency of the fluctuations has been measured to learn more about the underlying mechanism. The observed non-Arrhenius temperature dependence can be explained by temperature-dependent transition probabilities between substates within Hsp90's open state (Figure 71). The transition probabilities from an average open structure to less populated surrounding structures representing local energy minima, might increase with increasing temperature, for example, due to the dissolving of salt-bridges located at the CM-interface. If these substates are on the pathway towards closing, they would macroscopically appear as a single open state with large activation energy for the transition to the closed state. A sample calculation for four successive microstates illustrates this effect. When assuming that a temperature increase

by 10 K increases the forward rates by just 50 % while not affecting the backward rates, the total transfer rate would increase by 340 %. In this case, the macroscopic activation energy for the open-close transition would be similar to the activation energy of Hsp90's ATPase rate, indicating that the open state is ATPase rate limiting. Finally, the observed strong temperature dependence of the fluctuations in the open state might be advantageous for accelerating the adaption to size and shape of client proteins. Such temperature dependence might be part of a heat shock mechanism.

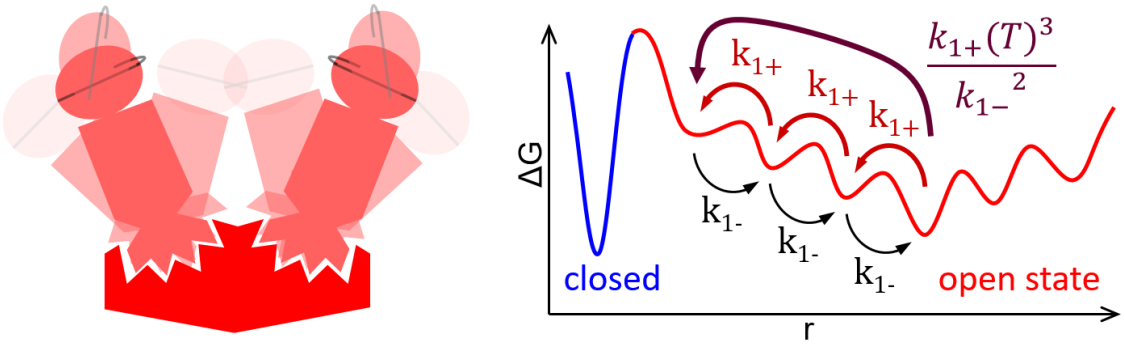


Figure 71: Left: Mechanistic model of Hsp90's open state reflecting the inter-domain fluctuations occurring at the micro- to millisecond timescale. Right: Free energy landscape representing the closed state (blue) and the open state divided into substates (red). Temperature-dependent successive transitions between microstates can lead to a strong temperature dependence of the dwell time of the global state, in spite of relatively fast single transition rates. In this model, the forward rates k_{1+} increase with temperature (reddish arrows), but the backward rates k_{1-} do not (black arrows).

Transient interaction between Hsp90 and a client protein

The dynamic nature of Hsp90 allows the adaption to size and shape of hundreds of client proteins. The conformational flexibility does not only offer steric advantages. The partially disordered CM-interface and MN-interface that could be seen as a molten globule state with several intermediate conformations might permit overlapping binding motifs, similar to what has been suggested for partially disordered regions in protein complexes (*110, 111*). Such disordered regions not only explain the overall dynamic behavior of the multi-domain protein, but could additionally enable transient binding.

According to the current opinion, co-chaperones of Hsp90 control the specificity of interaction with clients. Here, it was shown that a client protein itself reduces the sampled conformational space by dynamic interaction. A partially folded fragment of staphylococcal nuclease, Δ131Δ, was taken as client protein.

Two complementary experiments – distance fluctuation analysis of labeled Hsp90 as well as FCS of labeled Δ131Δ – indicated that Δ131Δ binds preferentially to the open state of yeast Hsp90.

The fast inter-monomer fluctuations on the (sub-) millisecond timescale within Hsp90's open state were reduced in presence of the client protein hinting towards a fast and transient binding. Such a fast progress is beneficial to sample several conformational intermediates of the client and to guarantee a fast release, so that the global conformations are left unchanged. Thus, transient interactions could serve as an interaction level that acts orthogonal to the global cycle of the chaperone machinery (Figure 72). Furthermore, the experiments indicate that protein-protein interactions are not only defined by regulating the free energy of activation, but also by altering the fluctuations (energy landscape roughness). This can be seen as a new thermodynamic mode of action (see also Figure 6).

At first glance, the slow ATPase rate suggests that a large energy barrier rules the chaperone activity. However, the experiments indicate that the transient interactions between Hsp90 and the model client occur on much faster timescales, namely in the range of the fast inter-monomer fluctuations. The global equilibrium between open and closed state(s) were not affected at all. This suggests that client binding does not directly depend on ATP hydrolysis. Instead, ATP hydrolysis might be necessary for the processing of co-chaperones or for forcing the Hsp90 dimer or a complexed state to open. Co-chaperones are supposed to transport client proteins to Hsp90 or to specify the process of client folding. The mutual interaction between chaperone and client could be additionally regulated by co-chaperones. Alternatively, client binding could represent a dynamic signature that can facilitate the binding of certain co-chaperones that then act on the client protein.

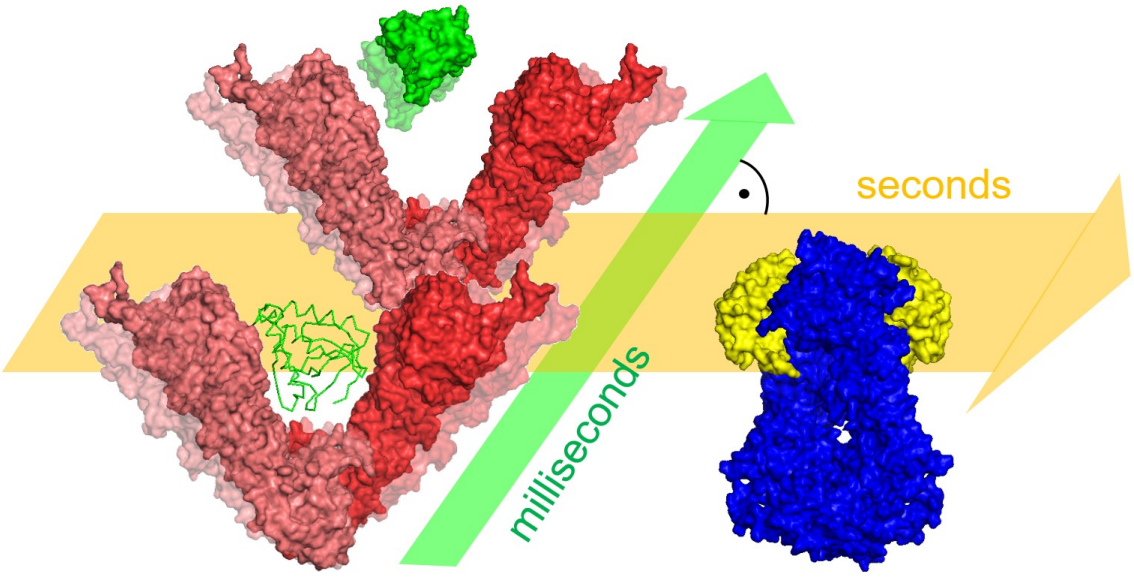


Figure 72: Schematics of an orthogonal hierarchic regulation mechanism. The structural ensembles of Hsp90's open state (red) and closed state (blue) have been determined by the novel approach for dynamic structures presented within this work. The transparent structures represent standard deviations of the measured distance distributions for an observation time of one tenth of a millisecond. Client proteins (green) bind to Hsp90's open

state. Transient binding on the millisecond timescale can be assisted by a co-chaperone (yellow). Fast client processing guarantees minimum inhibition, so that the overall molecular cycle can be dominated by the global co-chaperone pathway.

The following considerations lead to the speculation that Hsp90 enables folding of the partially folded client protein followed by a quick release. As shown, ATP hydrolysis is probably not necessary for the processing of the client protein. A stabilizing effect is also improbable, as Hsp90 does not notably bind to completely folded client proteins. Thus, Hsp90 seems to have (only) a catalytic effect on the client protein. The partially folded client might transiently bind to Hsp90 due to hydrophobic residues that are surface accessible. The transient binding likely manipulates the conformation of the client enabling the client to refold. In contrast to the slower inter-monomer fluctuations of Hsp90 that are probably necessary for size adaption, the faster fluctuations of the alpha-helix 15 and other flexible elements involving randomly oriented residues (as indicated by MD simulations of the open start structure) are probably useful to quickly sample different structural patterns. Dynamic interactions with the partially folded client – induced in this manner – might reduce several different energy barriers and thus facilitate a faster sampling of client conformations accelerating the process of finding its native state.

Another advantage of Hsp90's broadband dynamics could be the reduction of the dwell time of client binding, which is necessary to prevent product inhibition. Client binding likely leads to alignment of randomly oriented residues, so that the entropy and thereby the activation free energy for dissociation is reduced and the dissociation rate is increased. Contrary to expectations, such transient interactions seem to be specific more due to dynamic patterns than due to structural patterns.

In summary, a more detailed understanding of FRET-based experiments, methodological advancements and a comprehensive application of complementary and interdisciplinary methods resulted in a better understanding of Hsp90's function. It is anticipated that the presented methods and insights prepare the ground for the investigation of many more dynamic multi-domain proteins and multi-component protein assemblies.

6 APPENDIX

6.1 Figures, tables and protocols

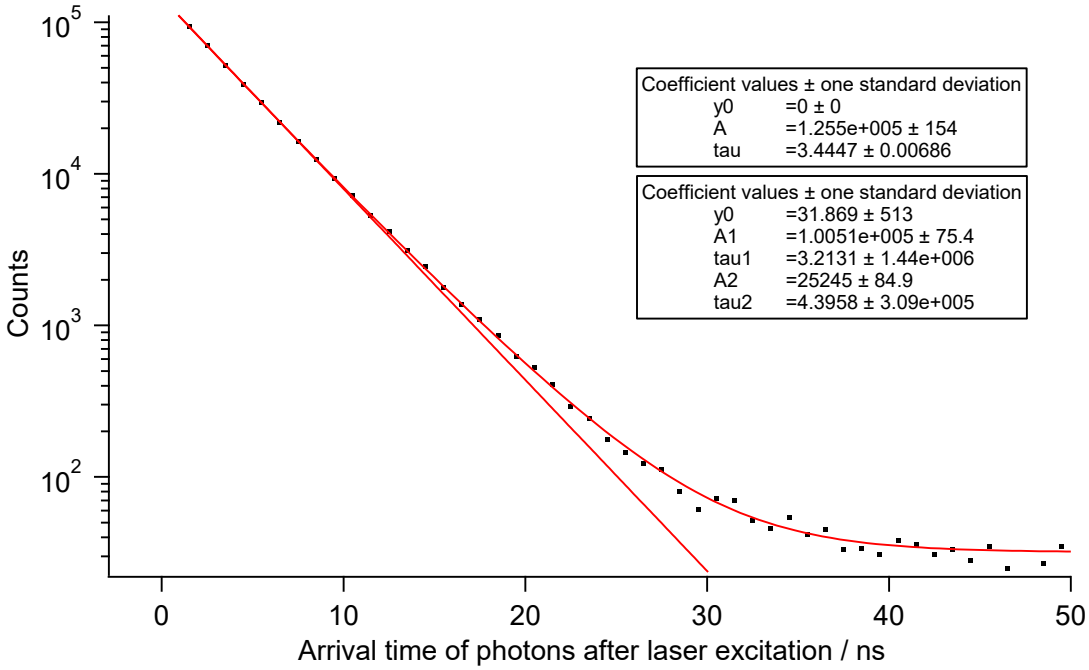


Figure 73: Time-correlated single photon counting histogram for 500.000 simulated exponential distributed photons with expectation values of either 3 ns or 4 ns (weighted by 49.5 % / 49.5 %) and 1 % Poison distributed background photons. The results were fitted with a single exponential without offset or with a double exponential with offset. The single exponential fit reveals an average lifetime of 3.44 ns. The double exponential fit reveals average lifetimes of 3.2 and 4.4 ns, thus deviating about 10 % from the expected values.

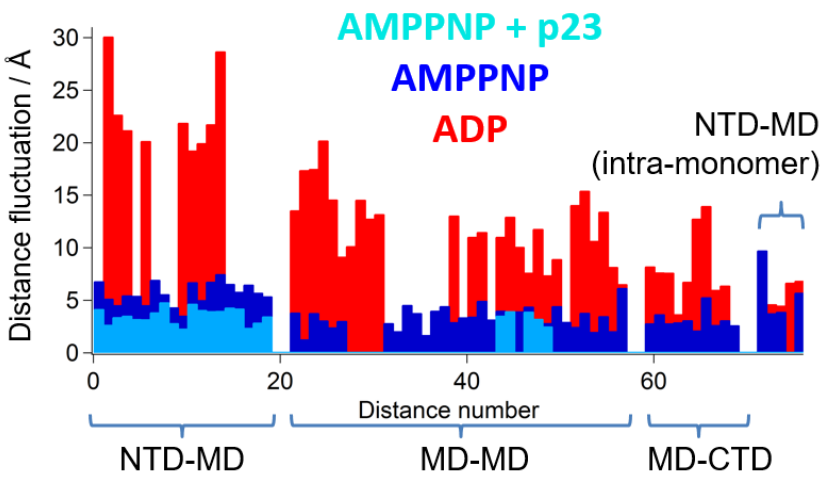


Figure 74: Standard deviation of the Gaussian distributed distance fluctuations were determined for all selected inter- and intra-domain distances in presence of AMPPNP or AMPPNP + P23 or ADP with a probability distribution analysis of the FRET efficiencies for a photon binning time of 1ms. In presence of AMPPNP + P23 the overall structure exhibits average fluctuations of less than 3 Å. The open ADP-bound state shows much larger fluctuations with maximum values of more than 20 Å.

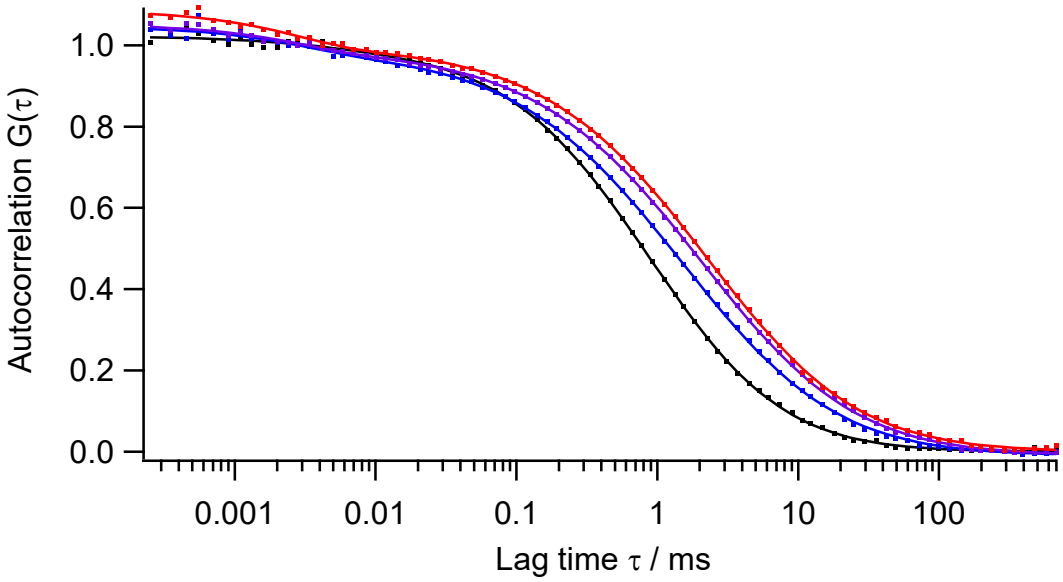


Figure 75: Autocorrelation of fluorescence intensities for the model client protein Δ131Δ labeled with Atto647N, in absence of Hsp90 (black), in presence of 10 µM Hsp90 and 2 mM AMPPNP (blue), in presence of 10 µM Hsp90 and 2 mM ATP (violet) and in presence of 10 µM Hsp90 and 2 mM ADP (red). The fits include kinetic terms for diffusion and intersystem crossing. All curves were normalized to an average number of one molecule per confocal volume. The fits revealed the following average diffusion times: 0.98 ms, 1.67 ms, 2.3 ms and 2.5 ms.

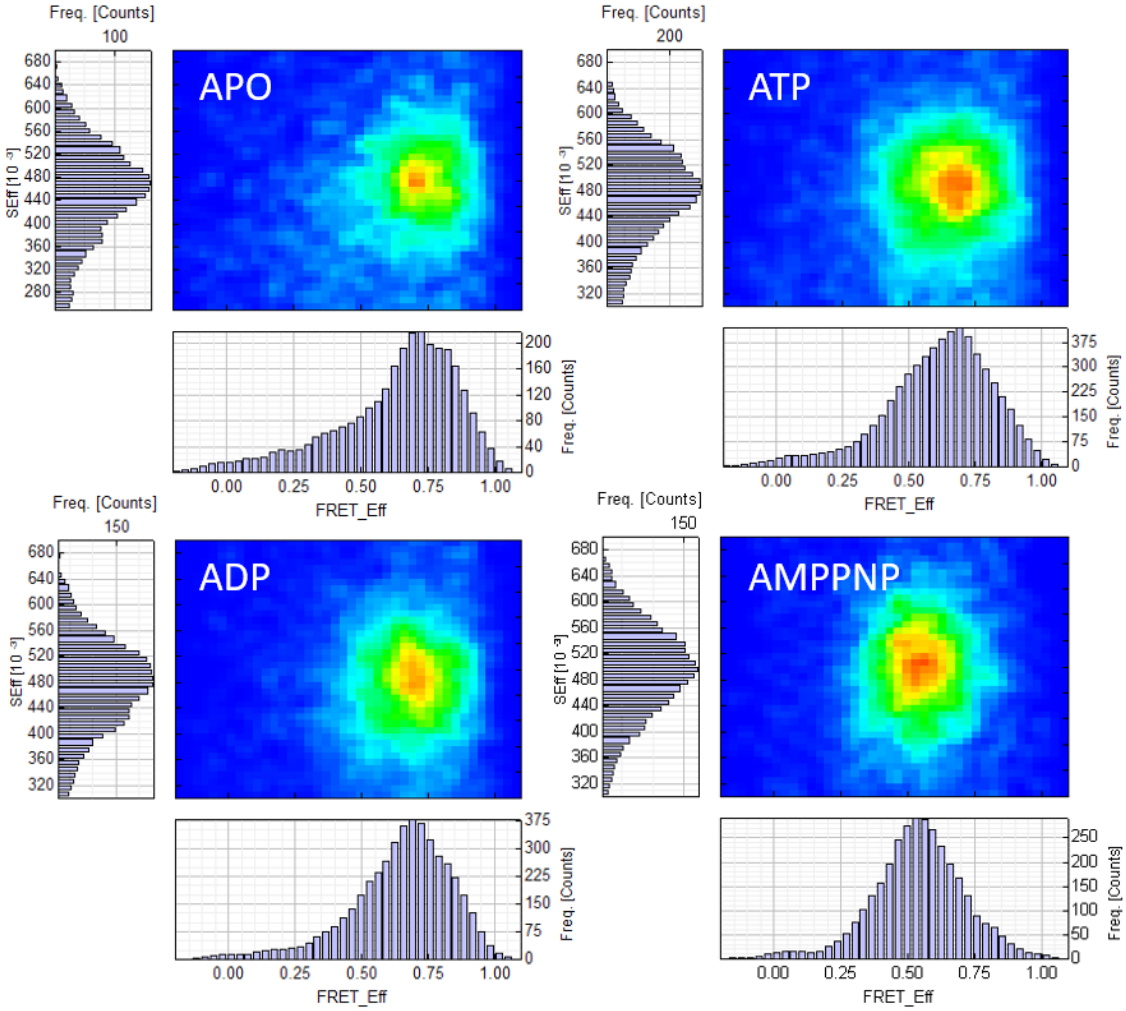


Figure 76: Stoichiometry-efficiency scatter plots obtained from a FRET pair spanning between N- and M-domain within a monomer of Hsp90 in presence of different nucleotides. The binning time of the photon traces was 1 ms. Smaller binning times increased the broadening of the transfer efficiencies, especially for APO, ADP and ATP.

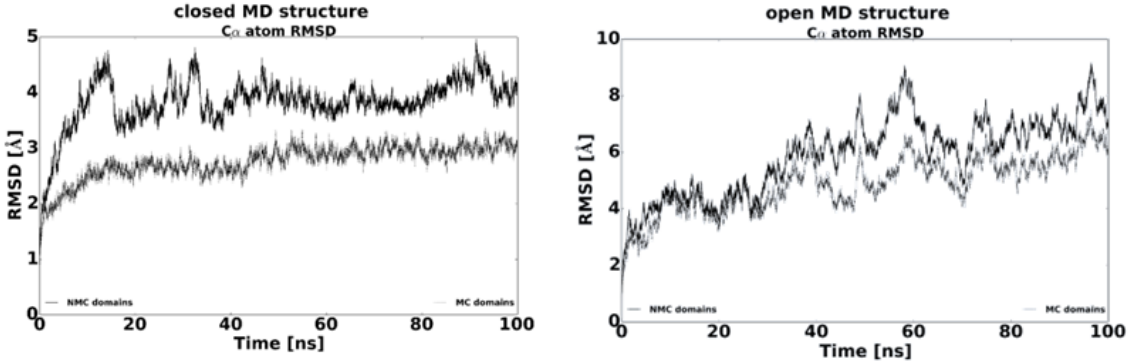


Figure 77: Root mean square deviations (RMSD) between simulated structure and start structure calculated from full-atom MD simulations in explicit solvent that were started from the closed structure (left) or the average open structure (right) of yeast Hsp90. The

RMSDs are calculated for the aligned full-length structure (black) and for the aligned structure without N-domains (gray).

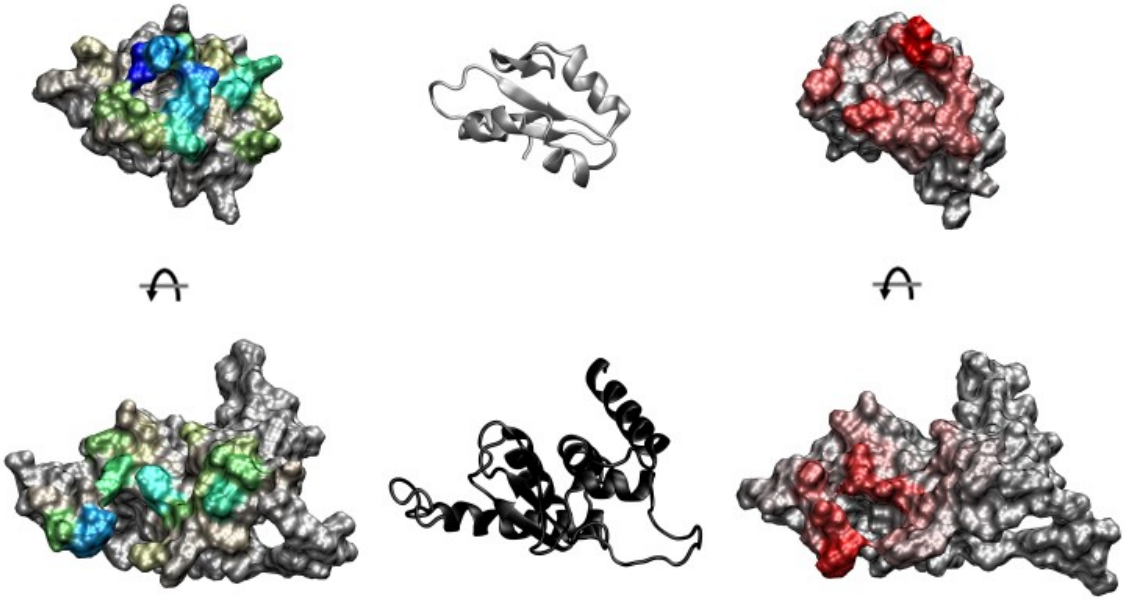


Figure 78: Normalized buried surface area per atom for the interface between M-domain (top) and C-domain (bottom) of Hsp90 generated by Florian Kandzia. The values are color-coded for the closed state (blue) and for the open state (red).

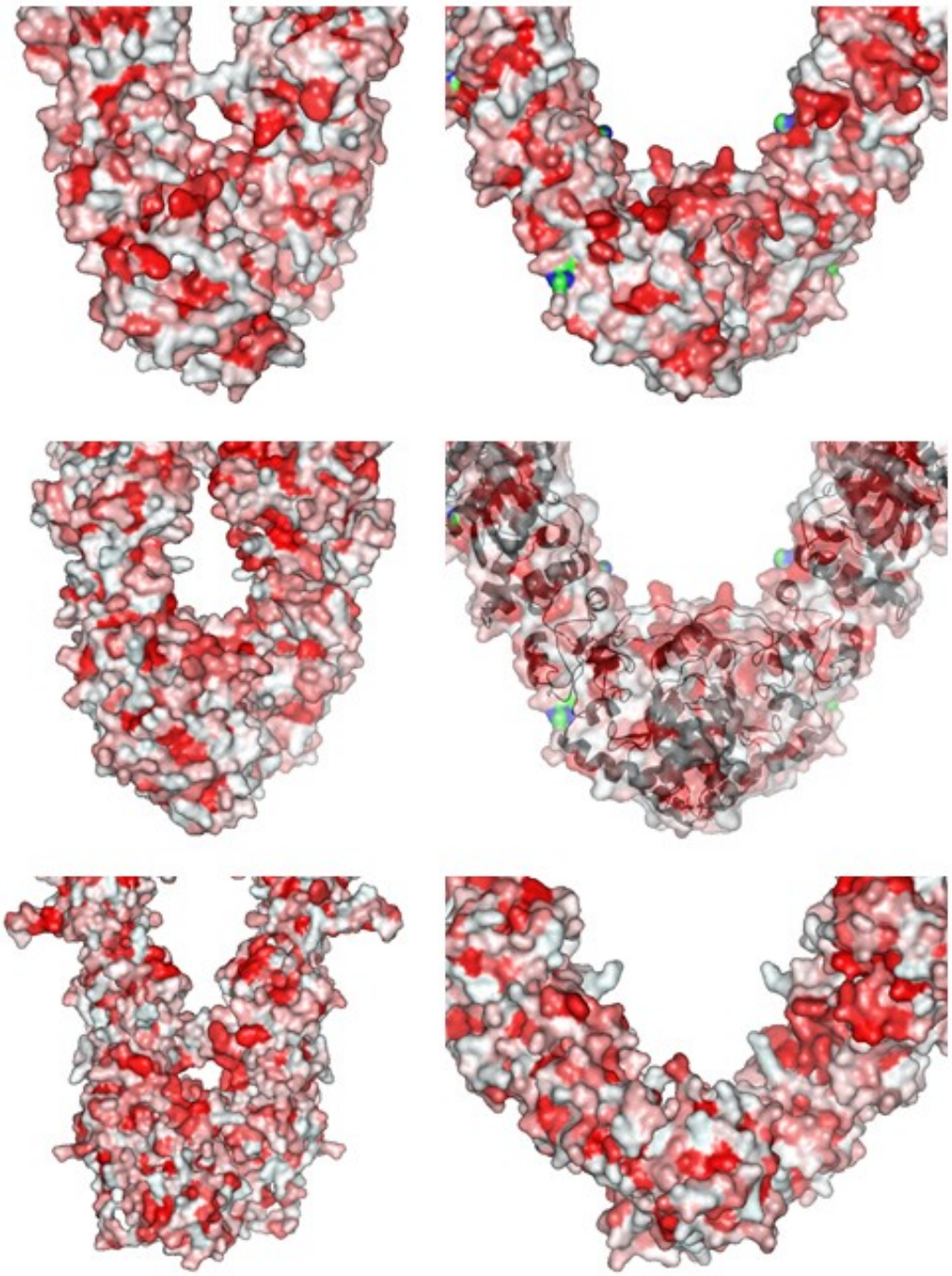


Figure 79: Hydrophobic surface of different structures color-coded in reddish. Top left: Closed AMPPNP-bound structure of yeast Hsp90 (pdb: 2cg9). Top right and middle right: Average open structure obtained from the measured distance network. Middle left: ADP-bound structure of Grp94, the endoplasmic reticulum Hsp90. Bottom left: Closed AMPPNP-bound structure of the bacterial homologue HtpG. Bottom right: Open APO structure of HtpG. The latter structure lacks of some flexible loops that might increase hydrophobicity.

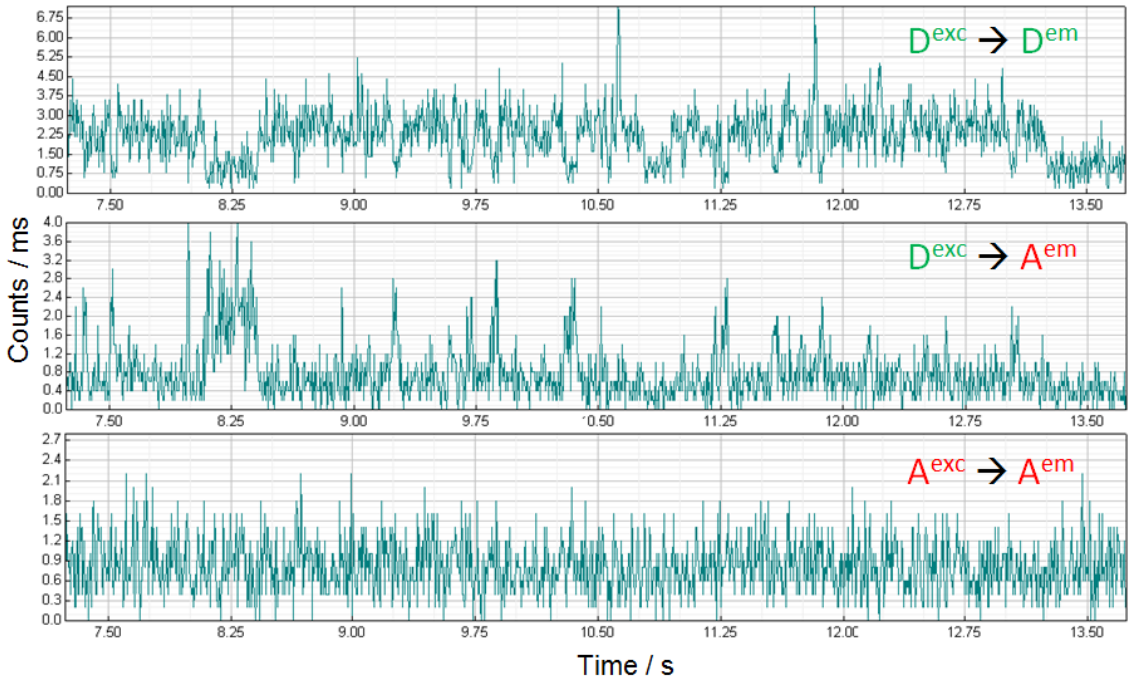


Figure 80: Fluorescence trajectories of Hsp90 molecules that were immobilized to a functionalized surface and measured with the confocal setup with picosecond time resolution. The fluorescence intensity traces were binned to 5 ms for donor emission after donor excitation (top), acceptor emission after donor excitation (middle) and acceptor emission after acceptor excitation (bottom). A clear anticorrelation between donor and acceptor channel is observed. The short-lived high efficiency events likely correspond to the semi-closed state.

		Closed state of Hsp90					Open state of Hsp90				
Pos.1	Pos.2	⟨R⟩ / Å	σ_R / Å	R_M-R	r(t→∞)	select	⟨R⟩ / Å	σ_R / Å	R_M-R	r(t→∞)	select
A_061	B_285	85.0	4.2	5.4	0.184						
A_061	B_298	73.0	2.7	0.3	0.228	A	93.8	30.1	4.1	0.217	A
A_061	B_327	66.2	3.4	12.2	0.253	A	90.8	22.7	11.4	0.227	A
A_061	B_385	56.2	3.6	-7.8	0.220	A	91.2	21.2	1.2	0.222	A
A_061	B_409	82.8	3.2	4.6	0.203						
A_061	B_431	63.6	3.2	-12.7	0.248	A	80.3	20.2	6.1	0.222	A
A_061	B_452	90.4	5.8	5.1	0.190						
A_061	B_517	77.4	7.4	-2.1	0.212						
A_152	B_285	56.9	2.9	-0.6	0.202						
A_152	B_298	61.9	2.4	-5.7	0.237	A	86.5	21.9	4.0	0.239	A
A_152	B_327	61.4	2.3	4.2	0.250	A	89.8	19.3	4.0	0.223	A
A_152	B_385	42.3	n.a.	-18.2	0.178	D	73.3	21.4	-2.7	0.227	A
A_152	B_409	57.9	4.1	0.1	0.186		101.9	11.1	2.2	0.163	
A_152	B_431	58.2	4.0	-15.1	0.265	A					
A_152	B_452	73.6	4.0	3.7	0.186		92.9	28.7	9.7	0.216	A
A_159	B_285	73.9	5.4	3.3	0.177						

A_159	B_298	75.1	4.3	-1.2	0.206						
A_159	B_327	76.1	2.4	7.4	0.211						
A_159	B_385	49.0	n.a.	-12.7	0.148	D					
A_159	B_409	75.3	n.a.	2.1	0.159						
A_159	B_431	68.9	2.9	-12.8	0.229	A					
A_159	B_452	91.6	3.5	5.0	0.166						
A_298	B_285	63.8	2.7	0.1	0.184		95.9	10.2	1.9	0.189	
A_298	B_298	53.4	1.3	1.3	0.219		77.2	18.0	-5.7	0.229	A
A_298	B_327	44.1	2.9	5.3	0.191	D	70.9	17.5	0.5	0.215	A
A_298	B_385	41.4	2.0	-6.8	0.174	D	69.2	20.2	-3.4	0.230	A
A_298	B_409	57.1	0.8	-0.1	0.175		88.1	14.6	1.6	0.181	G
A_298	B_431	43.4		-22.6	0.131	D	70.6	18.0	-16.9	0.211	A
A_298	B_452	60.3	1.8	2.7	0.178		85.5	9.2	-5.9	0.190	
A_298	B_503	47.3	n.a.	2.2	0.197						
A_298	B_517	48.2	n.a.	-4.4	0.131		80.3	14.5	-3.4	0.172	
A_327	B_285						85.7	10.2	10.1	0.185	
A_327	B_298						68.7	14.6	4.4	0.245	A
A_327	B_327						73.2	9.9	-4.6	0.245	A
A_327	B_409						77.7	12.8	7.6	0.194	
A_327	B_431						66.4	13.5	-12.9	0.239	A
A_327	B_452						69.4	13.2	3.3	0.206	
A_409	B_285	79.8	2.9	5.8	0.106						
A_409	B_298	57.7	2.1	0.7	0.144						
A_409	B_327	52.4	4.6	4.7	0.175		88.0	9.7	-3.5	0.169	
A_409	B_385	68.5	n.a.	-4.8	0.147						
A_409	B_409	81.2	3.8	1.6	0.106						
A_409	B_431	48.7	1.7	0.3	0.146						
A_431	B_285	51.3	4.1	-3.1	0.172		87.4	14.0	-3.7	0.165	
A_431	B_298	44.8	n.a.	-23.2	0.150	D	70.0	19.6	-14.7	0.222	A
A_431	B_385	51.1	4.5	-6.3	0.235	A	74.1	15.6	-9.3	0.264	A
A_431	B_409	50.9	3.0	-0.9	0.157		75.3	13.1	3.6	0.172	
A_431	B_431	42.2	3.4	-9.8	0.257	A	61.9	13.2	-12.7	0.225	A
A_431	B_452	46.5	3.5	0.4	0.150		66.5	12.5	-3.7	0.196	
A_431	B_503	41.0	n.a.	-13.1	0.239	A					
A_431	B_517	50.6	5.0	2.6	0.139		71.2	11.5	1.4	0.169	
A_452	B_285	79.0	0.4	3.8	0.146		101.7	12.4	-1.8	0.211	A
A_452	B_298	61.4	4.1	2.2	0.191		83.9	11.1	-3.0	0.182	
A_452	B_327	49.3	2.2	1.3	0.201		73.5	13.0	-1.5	0.198	
A_452	B_385	66.9	4.0	0.5	0.180		84.5	10.1	-2.5	0.212	A
A_452	B_409	73.7	4.4	0.9	0.169		90.2	7.7	-2.7	0.136	
A_452	B_431	47.7	2.9	-1.5	0.215		67.8	11.8	-5.3	0.181	
A_452	B_452	61.1	2.9	2.8	0.157		78.6	7.4	-7.6	0.181	
A_452	B_503						64.5	12.3	-1.0	0.148	
A_452	B_517	65.7	4.5	-1.0	0.119		76.6	9.0	-2.5	0.158	
A_517	B_285	75.3	3.0	0.1	0.090		99.6	11.3	4.3	0.120	

A_517	B_298	51.7	0.8	-6.7	0.164		79.7	14.1	-1.8	0.164	
A_517	B_327	49.2	3.1	-2.2	0.197		72.5	15.5	-1.4	0.197	
A_517	B_409	75.6	2.0	0.7	0.101		96.1	10.7	1.2	0.101	
A_517	B_431	49.5	2.7	2.8	0.182		72.0	13.5	-0.1	0.182	
A_517	B_452	65.6	2.1	-0.1	0.120		78.6	8.2	-4.7	0.105	
A_517	B_503	49.1	n.a.	-1.4	0.170		64.8	12.2	-0.3	0.130	
A_517	B_517	76.0	5.0	0.6	0.106		87.5	6.6	3.7	0.106	
A_560	B_285	50.2	2.9	0.1	0.190	L	72.0	8.3	-3.1	0.176	L
A_560	B_298	43.4	3.7	-2.6	0.209	L	60.7	7.7	-5.3	0.161	L
A_560	B_409	49.3	2.8	-1.6	0.196	L	60.9	7.7	1.4	0.164	L
A_560	B_431	45.5	2.9	7.3	0.188	LD					
A_560	B_517	51.0	3.2	4.7	0.177	L	57.4	6.8	5.9	0.128	L
A_560	B_637						51.1	4.1	2.6	0.179	L
A_637	B_285	75.7	n.a.	-4.7	0.220	A	86.3	12.8	-5.3	0.216	A
A_637	B_385						82.4	14.0	13.9	0.249	A
A_409	B_637	62.7	n.a.	0.8	0.171		70.2	5.5	-0.2	0.170	
A_452	B_637	43.2	n.a.	-3.1	0.205		49.1	3.4	-1.8	0.198	
A_517	B_637	62.2	n.a.	4.1	0.176		65.1	6.4	3.5	0.166	
A_637	B_637	42.8	n.a.	-1.7	0.252	A	42.7	2.1	-5.5	0.245	A
A_009	A_452	86.1	9.6	8.5	0.139		55.2	7.4	2.1	0.182	
A_061	A_431	61.7	3.9	-0.3	0.218		54.7	4.7	-0.1	0.207	
A_061	A_452	64.8	3.9	-1.6	0.197		64.4	4.5	5.7	0.197	
A_152	A_285						52.1	6.4	2.5	0.185	
A_152	A_452	77.1	7.1	4.9	0.196		62.7	6.9	-1.5	0.196	
A_298	A_452	46.0	n.a.	-1.4	0.146		46.7	4.0	-5.9	0.129	
A_159	B_061	75.4	n.a.	-5.7	0.207						
A_152	B_061	59.6	n.a.	-8.9	0.223	A					
A_061	B_061	87.8	10.0	-0.7	0.216						
A_061	B_159	72.2	10.0	-2.2	0.180						
A_152	B_159	55.1	10.0	-5.5	0.183						
B_061	A_285	85.0	4.2	5.3	0.184						
B_061	A_298	73.0	2.7	-1.4	0.228	A	93.8	30.1	2.8	0.217	A
B_061	A_327	66.2	3.4	8.7	0.253	A	90.8	22.7	11.1	0.227	A
B_061	A_385	56.2	3.6	-6.2	0.220	A	91.2	21.2	1.2	0.222	A
B_061	A_409	82.8	3.2	4.5	0.203						
B_061	A_431	63.6	3.2	-13.0	0.248	A	80.3	20.2	6.0	0.222	A
B_061	A_452	90.4	5.8	4.1	0.190						
B_061	A_517	77.4	7.4	-2.4	0.212						
B_152	A_285	56.9	2.9	-0.2	0.202						
B_152	A_298	61.9	2.4	-6.0	0.237	A	86.5	21.9	2.6	0.239	A
B_152	A_327	61.4	2.3	2.4	0.250	A	89.8	19.3	3.7	0.223	A
B_152	A_385	42.3	n.a.	-28.4	0.178	D	73.3	21.4	-3.1	0.227	A
B_152	A_409	57.9	4.1	0.0	0.186		101.9	11.1	1.3	0.163	
B_152	A_431	58.2	4.0	-15.3	0.265	A					
B_152	A_452	73.6	4.0	3.4	0.186		92.9	28.7	9.2	0.216	A

B_159	A_285	73.9	5.4	3.7	0.177						
B_159	A_298	75.1	4.3	-1.8	0.206						
B_159	A_327	76.1	2.4	4.9	0.211						
B_159	A_385	49.0	n.a.	-17.2	0.148	D					
B_159	A_409	75.3	n.a.	2.0	0.159						
B_159	A_431	68.9	2.9	-13.6	0.229	A					
B_159	A_452	91.6	3.5	4.4	0.166						
B_298	A_285	63.8	2.7	1.5	0.184		95.9	10.2	3.4	0.189	
B_298	A_298	53.4	1.3	1.3	0.219		77.2	18.0	-5.7	0.229	A
B_298	A_327	44.1	2.9	5.5	0.191	D	70.9	17.5	2.2	0.215	A
B_298	A_385	41.4	2.0	-3.9	0.174	D	69.2	20.2	-2.2	0.230	A
B_298	A_409	57.1	0.8	1.2	0.175		88.1	14.6	2.6	0.181	G
B_298	A_431	43.4	n.a.	-21.8	0.131	D	70.6	18.0	-15.3	0.211	A
B_298	A_452	60.3	1.8	3.3	0.178		85.5	9.2	-4.6	0.190	
B_298	A_503	47.3	n.a.	2.7	0.197						
B_298	A_517	48.2	n.a.	-3.2	0.131		80.3	14.5	-2.4	0.172	
A_327	B_285						85.7	10.2	10.1	0.185	A
A_327	B_298						68.7	14.6	4.4	0.245	A
A_327	B_327						73.2	9.9	-4.6	0.245	A
A_327	B_409						77.7	12.8	7.6	0.194	A
A_327	B_431						66.4	13.5	-12.9	0.239	A
A_327	B_452						69.4	13.2	3.3	0.206	A
B_409	A_285	79.8	2.9	5.2	0.106						
B_409	A_298	57.7	2.1	-0.7	0.144						
B_409	A_327	52.4	4.6	1.2	0.175		88.0	9.7	-2.7	0.169	
B_409	A_385	68.5	n.a.	-0.2	0.147						
B_409	A_409	81.2	3.8	1.6	0.106						
B_409	A_431	48.7	1.7	1.3	0.146						
B_431	A_285	51.3	4.1	-4.3	0.172		87.4	14.0	-3.8	0.165	
B_431	A_298	44.8	n.a.	-23.9	0.150	D	70.0	19.6	-16.3	0.222	A
B_431	A_385	51.1	4.5	-5.2	0.235	A	74.1	15.6	-9.5	0.264	A
B_431	A_409	50.9	3.0	-1.9	0.157		75.3	13.1	3.0	0.172	
B_431	A_431	42.2	3.4	-9.8	0.257	A	61.9	13.2	-12.7	0.225	A
B_431	A_452	46.5	3.5	-0.2	0.150		66.5	12.5	-3.9	0.196	
B_431	A_503	41.0	n.a.	-13.7	0.239	A					
B_431	A_517	50.6	5.0	1.7	0.139		71.2	11.5	0.8	0.169	
B_452	A_285	79.0	0.4	3.9	0.146		101.7	12.4	-1.5	0.211	A
B_452	A_298	61.4	4.1	1.6	0.191		83.9	11.1	-4.3	0.182	
B_452	A_327	49.3	2.2	1.4	0.201		73.5	13.0	-0.8	0.198	
B_452	A_385	66.9	4.0	6.0	0.180		84.5	10.1	-2.4	0.212	A
B_452	A_409	73.7	4.4	1.4	0.169		90.2	7.7	-3.2	0.136	
B_452	A_431	47.7	2.9	-0.8	0.215		67.8	11.8	-4.9	0.181	
B_452	A_452	61.1	2.9	2.8	0.157		78.6	7.4	-7.7	0.181	
B_452	A_503						64.5	12.3	-1.4	0.148	
B_452	A_517	65.7	4.5	-0.2	0.119		76.6	9.0	-2.7	0.158	

B_517	A_285	75.3	3.0	-0.9	0.090		99.6	11.3	4.7	0.120	
B_517	A_298	51.7	0.8	-7.9	0.164		79.7	14.1	-2.8	0.164	
B_517	A_327	49.2	3.1	-7.3	0.197		72.5	15.5	-0.9	0.197	
B_517	A_409	75.6	2.0	0.2	0.101		96.1	10.7	1.1	0.101	
B_517	A_431	49.5	2.7	3.7	0.182		72.0	13.5	0.6	0.182	
B_517	A_452	65.6	2.1	-0.9	0.120		78.6	8.2	-4.5	0.105	
B_517	A_503	49.1	n.a.	-1.8	0.170		64.8	12.2	-0.6	0.130	
B_517	A_517	76.0	5.0	0.6	0.106		87.5	6.6	3.7	0.106	
B_560	A_285	50.2	2.9	-1.6	0.190	L	72.0	8.3	-2.4	0.176	L
B_560	A_298	43.4	3.7	-1.8	0.209	L	60.7	7.7	-6.0	0.161	L
B_560	A_409	49.3	2.8	-3.5	0.196	L	60.9	7.7	1.6	0.164	L
B_560	A_431	45.5	2.9	6.7	0.188	LD					
B_560	A_517	51.0	3.2	3.0	0.177	L	57.4	6.8	6.4	0.128	L
B_560	A_637						51.1	4.1	2.3	0.179	L
B_637	A_285	75.7	n.a.	-4.0	0.220	A	86.3	12.8	-4.3	0.216	A
B_637	A_385						82.4	14.0	14.5	0.249	A
B_409	A_637	62.7	n.a.	0.0	0.171		70.2	5.5	-0.5	0.170	
B_452	A_637	43.2	n.a.	-3.8	0.205		49.1	3.4	-2.0	0.198	
B_517	A_637	62.2	n.a.	3.3	0.176		65.1	6.4	2.9	0.166	
B_637	A_637	42.8	n.a.	-1.8	0.252	A	42.7	2.1	-5.5	0.245	A
B_009	B_452	86.1	9.6	8.5	0.139		55.2	7.4	2.0	0.182	
B_061	B_431	61.7	3.9	-1.3	0.218		54.7	4.7	-0.3	0.207	
B_061	B_452	64.8	3.9	-2.8	0.197		64.4	4.5	5.3	0.197	
B_152	B_285						52.1	6.4	2.6	0.185	
B_152	B_452	77.1	7.1	4.9	0.196		62.7	6.9	-2.0	0.196	
B_298	B_452	46.0	n.a.	-1.9	0.146		46.7	4.0	-4.9	0.129	
B_159	A_061	75.4	n.a.	-5.4	0.207						
B_152	A_061	59.6	n.a.	-8.7	0.223	A					
B_061	A_061	87.8	10.0	-0.7	0.216						
B_061	A_159	72.2	10.0	-2.5	0.180						
B_152	A_159	55.1	10.0	-5.6	0.183						

Table 4: Intra-molecular distance distributions were measured between numerous positions of chain A and B of yeast Hsp90 (A-B or B-A: inter-monomer distances; A-A or B-B: intra-monomer distances). Mean distances $\langle R \rangle$ and standard deviations σ_R of Gaussian distance distributions were extracted from the measured efficiency histograms with a distance distribution analysis (see Chapter 4.2). The deviations between the measured distances R and respective model distances R_M take into account the accessible volumes of the dyes and FRET-averaged distances (see Chapter 4.1 and 4.3). The x-ray crystal structure (pdb: 2cg9) and the average open structure served as model structures. The geometric mean of the time-resolved residual dye anisotropies $r(t\rightarrow\infty)$ was taken as a selection criterion: 0.22 for the closed structure and 0.21 for the open structure (see Chapter 4.1.4 for the derivation). The distances that were sorted out are marked with an A for high anisotropy, L for relative low

ATPase activity, D for a position-specific deviation of the anisotropy or G for a deviating global gamma-factor (see Chapter 4.3.2 for further explanations).

Measurement chamber with surface functionalization

First step, cleaning of 10 glass slides (170µm, Carl Roth):

- Prepare Piranha solution: 1 part H_2O_2 (50ml, 30%), 3 parts (150ml) H_2SO_4.
- Prepare RCA solution: 1 part H_2O_2 (30ml, 30%), 5 parts ddH_2O (150ml), 1 part NH_3 (30ml, 33%).
- Put slides in 2% Hellmanex (Hellmanex analytics) in ultrasonic bath for 10 min.
- Put slides in ddH_2O in ultrasonic bath two times for 10 min.
- Clean slides for 2 h at 55°C in piranha solution.
- Clean slides with ddH_2O.
- Put slides in 2% Hellmanex in ultrasonic bath two times for 10 min.
- Clean slides with ddH_2O.
- Put slides in ddH_2O in ultrasonic bath two times for 10 min.
- Clean slides for 2 h at 55°C in RCA solution.
- Clean slides with ddH_2O.
- Put slides in 2% Hellmanex in ultrasonic bath.
- Put slides in ddH_2O in ultrasonic bath for 10 min.
- Storage in ddH_2O.

Second step, functionalization of cleaned glass slides:

- Solve 80 mg mono PEG 5 kDA (Rapp Polymere) in 600 µL 100mM $NaHCO_3$ at pH 8.5; shake it and keep it cool.
- Dry slides with compressed pure nitrogen and put them in pure acetone.
- Incubate slides 5-10 min in 200 ml acetone with 1 ml Vectrabond (Axxora).
- Swivel slides in ddH_2O and place each one on a petri dish.
- Bring 70 µL of PEG solution on each slide and cover it with an untreated slide.
- Keep everything dark, cold (4°C) and humid with water-soaked wipes.
- After 3 hours: Swivel slides in ddH_2O to wash off unbound PEG.
- Dry slides with nitrogen; storage them dark, cold, ideally in a vacuum chamber.

Third step, chamber construction:

- Cut 5mm-holes into a Nescofilm with a laser cutter (3 rows, 8 columns).
- Drill holes into a 1mm-thick Teflon block.
- Bring the Nescofilm between a functionalized slide and the Teflon block.
- Burden the chamber on a heat plate (70°C) for 20 min with the slide side down.

- Finally, use each hole for 1-2 measurement days (see Figure 81).
- Reuse the Teflon block after cleaning with Hellmanex and ddH_2O.

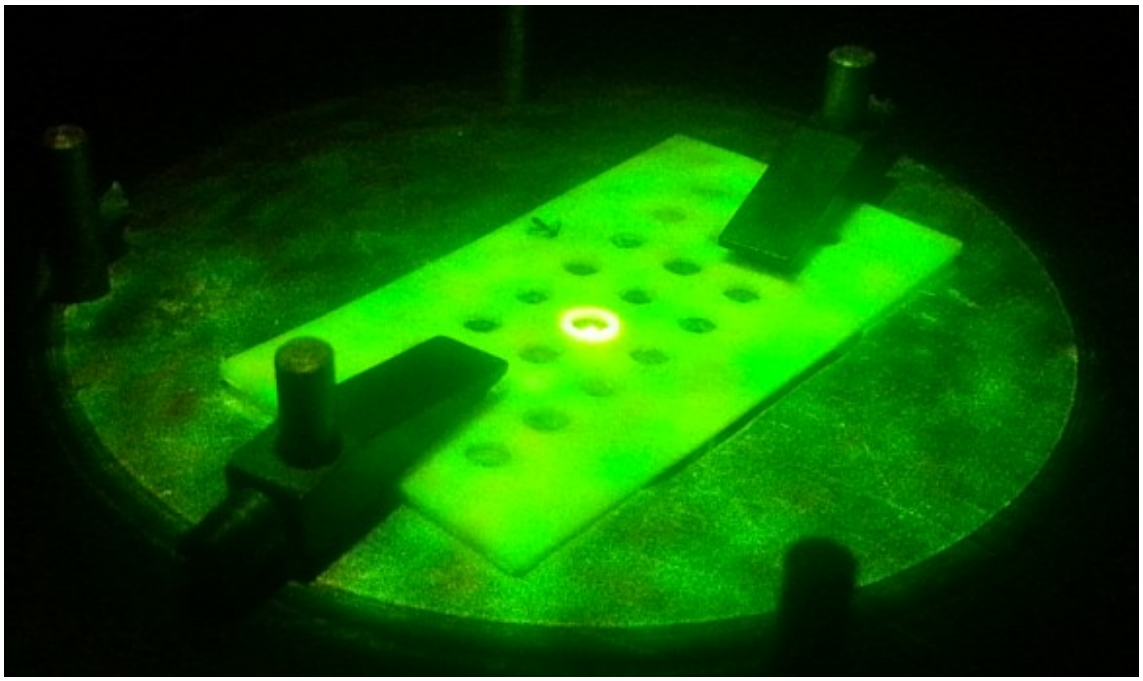


Figure 81: Measurement chamber in action

6.2 Fluorogami

6.2.1 Intensity proportional oligomer size determination

The knowledge of the size distribution of oligomeric proteins is essential for understanding the oligomerization process. It has been reported that certain intrinsically disordered proteins can form oligomeric intermediates, and for some of these proteins such as α-synuclein, it has been suggested that oligomer formation represents a pre-stage to amyloid formation. The oligomeric size distribution and its temporal evolution is averaged and therefore lost when observing an ensemble of such oligomeric molecules.

Therefore, several single-molecule fluorescence methods have been developed to access such size distributions. If the oligomers can be immobilized to a surface, their size distribution can be roughly estimated by counting the bleaching steps of the dye labels. Estimating the number by plain fluorescence intensity is nearly impossible without further tricks because of inhomogeneous excitation with total internal reflection and confocal-based microscopy. By fitting theoretical models to time-correlated single photon counting histograms, to fluorescence bursts intensity distributions or to fluorescence intensity correlations, the distribution of monomers, dimers and eventually trimers can be distinguished (*112-114*). Alternatively, by combining confocal detection and sample scanning, the size of the biggest oligomers (brightest species) can be detected. However, these semi single-molecule methods are insensitive to small populations

because of temporal averaging and intensity heterogeneities. The stoichiometry of a two-color sample (50% donor and 50% acceptor) allows distinguishing between monomers, dimers and trimers (*115*) for each single molecule. However, it is limited to trimers because of shot-noise broadening, heterogeneities in the confocal laser profiles and a fixed intensity threshold. Additionally, dimers are not distinguishable from tetramers, as well as, trimers from hexamers and so on.

In this work, a method has been developed that simultaneously distinguishes between monomers, dimers, trimers, tetramers and between 10mers, 20mers and so on. The main idea is the introduction of a third selection laser (510 nm), which is positioned at the center of the confocal volumes of the standard lasers (532 and 640 nm). The selection laser is focused at diffraction limit, while the other confocal volumes are eightfold enlarged (Figure 82, left). Now, a threshold for the fluorescence intensity of the selection laser (blue) selects only molecules that diffuse through the centers of the standard lasers (green), and the center of a Gaussian intensity distribution can be approximated with a constant intensity. In this manner, the broadening of the intensity distribution is dramatically reduced and the intensity after excitation with the standard laser is nearly proportional to the number of labeled molecules (here number of green dyes). Scanning the sample with constant velocity additionally decreases the chance of partial diffusion outside the threshold region of the selection laser.

Still, increasing fluorescence intensities due to several dyes per molecule, leads to a broadening of the selected intensity region after green excitation (Figure 82, middle). When the amount of green dyes per molecule exceeds a certain number, the remaining inhomogeneity of the fluorescence intensity after green excitation starts to increase rapidly. To overcome this problem, the target molecules are labeled with red dyes instead of green dyes and excited with a red laser. And a much smaller fraction of target molecules labeled with green dyes is added that is excited only by the selection laser (Figure 82, right). Like this, the fluorescence intensity profile upon excitation with the selection laser is nearly independent from the number of red dyes to be analyzed. Now, the selected intensity region becomes nearly independent from the number of red labeled molecules, so that the average intensity of each single photon burst becomes directly proportional to the number of labeled molecules and the relative populations are not biased by the threshold.

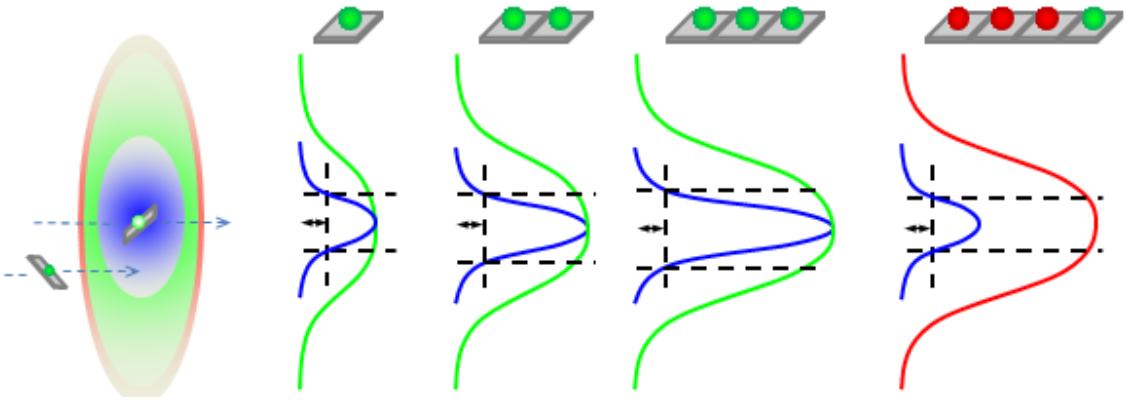


Figure 82: Principle of the novel selection method for intensity proportional oligomer size determination. The blue selection laser (510 nm) is focused at diffraction limit and the green and red lasers (532 nm, 640 nm) are eightfold enlarged (left). The fluorescence intensity profiles (meaning the convolution of excitation and detection profiles) are illustrated with Gaussian distributions. A threshold for the fluorescence intensity after excitation with the selection laser leads to a near-homogeneous intensity distribution after green excitation and therefore to an oligomer-size proportional intensity signal (middle). Further spectral separation of the selection laser and the measurement laser reduces the rest inhomogeneity of the selected fluorescence intensity region (right).

As a test sample a DNA origami was designed and prepared by Thomas Martin (group of Hendrik Dietz). This construct allows to link any desired dye combination to the positions that are marked with a star in Figure 83. At these positions, single stranded DNA molecules with specific sequences point outwards, to which complementary DNA strands labeled with a dye, can bind. With this construct oligomeric species of arbitrary size were mimicked in a controlled way.

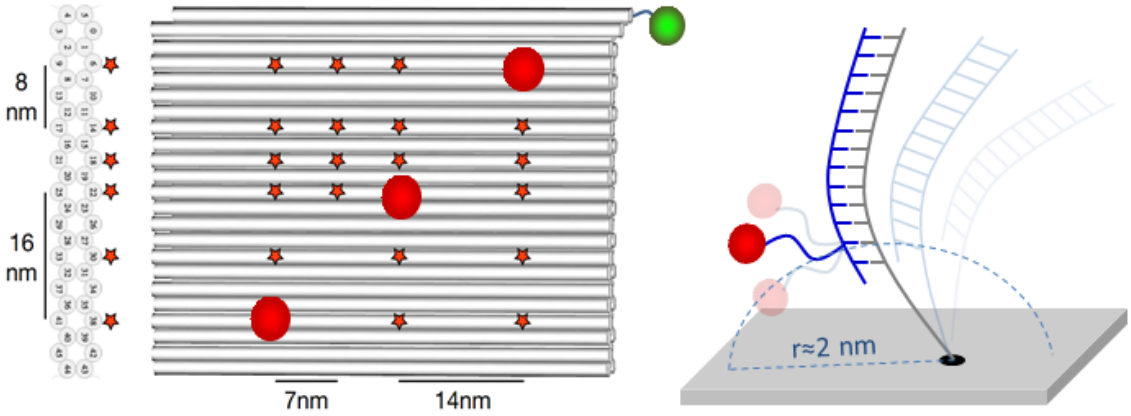


Figure 83: Fluorogami pad mimicking a tetrameric molecular assembly. Schematics of a DNA origami construct in form of a stable nanometer-sized bread board provided with selected dyes at defined positions from the side view (left), top view (middle) and zoomed-in side view (right). The stars indicate possible positions of single DNA strands hanging out, to which sequence-specific dye-labeled DNA strands can be attached. The construct has been

built by Thomas Martin (group of Hendrik Dietz). The dyes are labeled to thymidines, 7 base-pairs away from the pad (right), to avoid pad-contact and enable sufficient dye mobility and, at the same time, avoid inter-dye contact.

In the following, three red dyes were combined with one green dye mimicking a tetramer. The green dye enables direct comparison between the standard stoichiometry method (in this case dividing red intensity by green intensity) and the novel selection method (Figure 84). The coefficients of variation of the Gaussian fits strongly increase from 26 % for one red dye, over 41 % for two red dyes, up to 69 % for three red dyes, when using the standard stoichiometry method. In contrast, with the selection method, the coefficients of variation increase only from 22 % for one red dye, over 40 % for two red dyes, up to 40 % for three red dyes, meaning a decreasing increase of the coefficients of variation with increasing number of dyes. (It should be mentioned that the standard deviation is larger than shot noise limitation.) This has not only the advantage of distinguishing larger oligomers. Another advantage is that single molecules of smaller oligomers can be directly assigned to their oligomer size without the need of statistics. This goes beyond multi-parameter correlation, because each single molecule can now be assigned directly to oligomer size, transfer efficiency and time evolution. Thus, transient oligomer formations can be tracked. Notably, when applying the novel selection method, the mean intensity of each selected molecule is a perfect multiple of the mean intensity of the single-labeled species.

It is worth noting that in case of randomly assembled molecules there is a certain probability to find an oligomer with two green dyes (i.e. two selection dyes), although the number of red labeled molecules is chosen to be much larger than the number of green labeled molecules. In this case, the simple stoichiometry method cannot separate dimers from tetramers and trimers from hexamers and so on. In contrast, the novel method is able to separate these species; only the oligomer size proportional average intensity is slightly reduced.

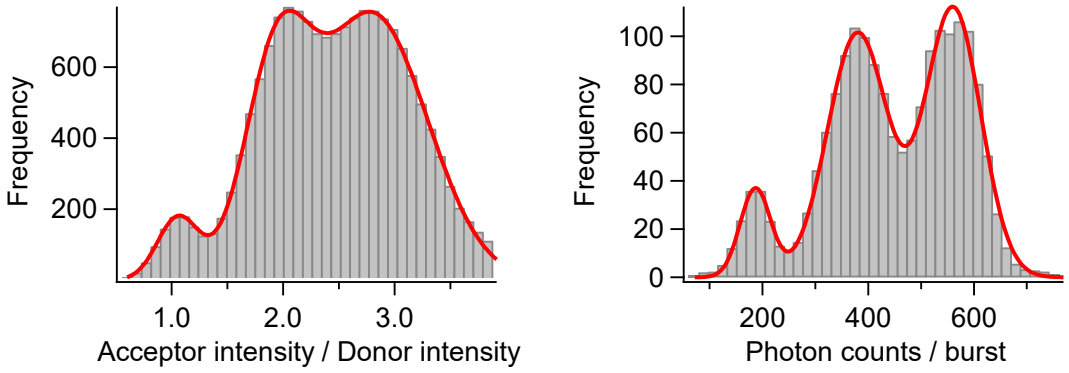


Figure 84: Intensity distributions of a tetramer sample (3x Atto647N and 1x Atto532) analyzed with the standard stoichiometry method (left) and the novel selection method presented herein (right).

From the three populations of the tetramer sample, the binding efficiency of the red dyes can be estimated. The occupation probabilities derived from the integrated Gaussian fits, fit fairly well to a binomial distribution and the average binding efficiency is calculated to be about 78 %. This total efficiency considers the efficiency of DNA hybridization, the efficiency dye labeling to the DNA strands and the probability that the dye is not yet bleached. Figure 85 shows the intensity distributions for a sample with 15 Atto647N dyes, analyzed with the selection method. Notably, the measured average intensity is smaller than expected if assuming a binomial distribution with a binding efficiency of 78 %. Saturation effects of the setup such as detector saturation can be excluded, as well as, electron transfer. Instead, it turned out that the intensity loss is caused by dipole-dipole coupling between proximate identical dyes. This observation is in contradiction to the unquenched ensemble fluorescence lifetimes and has therefore been analyzed in more detail, to develop a photophysical model (to be continued).

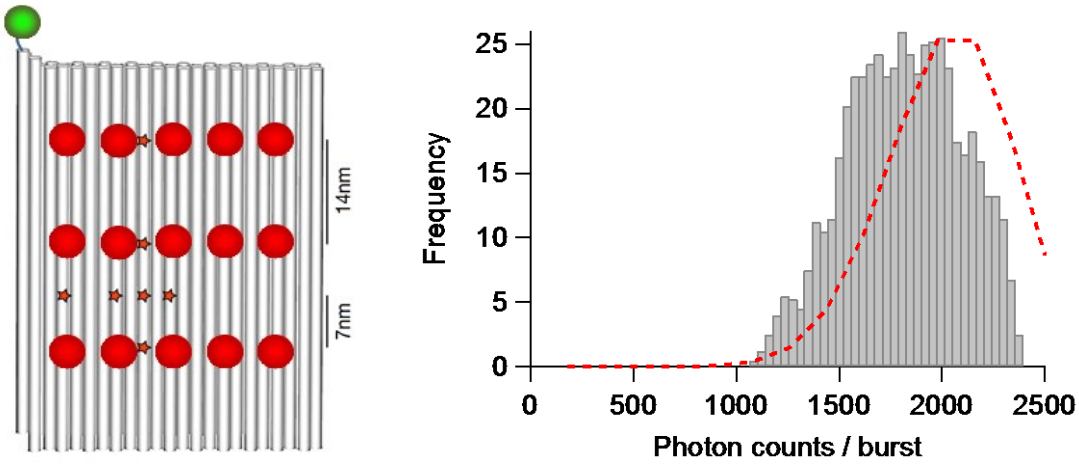


Figure 85: Oligomer sample with 15x Atto647N and 1x Atto532 (left) analyzed with the novel selection method (right). The dashed red line represents the expected binomial distribution considering the binding probability of 78 % that has been calculated by means of the trimer sample.

6.2.2 Multiple energy transfer

The origami pads and the novel selection method presented in Chapter 6.2.1 were further utilized to examine FRET between three dyes out of two species, Atto532 and Atto647N. Two different exactly defined dye arrangements have been used, a triangle geometry with 8 nm dye separation (Figure 86 left) and a linear geometry with 7 nm dye separation (Figure 86 right). A mixture of 50% Atto532 (green) and 50% Atto647N (red) was incubated together with the origami pad (prepared for one of both geometries) resulting in every possible combination. All these combinations can be perfectly distinguished by applying the novel selection method and by correlating intensity ratios with FRET efficiencies. The intensity ratio is defined as the number of

all detected photons after red excitation divided by the number of all detected photons after green excitation.

The triangle geometry has the advantage that the distances are equal between all three dyes, and therefore, fewer combinations are possible. At a distance of 8 nm the intensity decrease due to homo energy transfer between both Atto532 dyes is negligible as indicated by separate measurements (data not shown), so that population B is not measurably shifted to higher efficiencies. What is more interesting here is population D, which has a significant larger efficiency than population C, clearly demonstrating that the presence of two acceptors lead to a competition between energy transfer rates. This observation experimentally confirms competition between transfer rates for the week coupling regime, so that equation (7) can be extended to the following relation:

$$E_{ET,1} = \frac{k_{ET,1}}{k_r + k_{nr} + \sum_{i=1}^{n} k_{ET,i}} \qquad (54)$$

In the linear geometry, the populations B, C, D and E are the most interesting ones. The efficiency of population C equals the efficiency of population E. The efficiency of population B should be the average value of the efficiencies of populations D and E. However, it is slightly shifted towards the efficiency of population E. The observation can be explained by reversible energy transfer between both donor dyes and irreversible energy transfer from the middle donor dye to the acceptor and is a known effect in photosynthesis. The energy from the outer donor is transferred with a certain probability to the middle donor, which in turn can transfer the energy irreversibly to the acceptor. Like this, the energy can be transferred over long distances, while the acceptor is acting as a sink that emits the energy with highest probability. One has to consider that the intensity loss due to dipole coupling between identical dyes can be large.

Conclusively, it was shown experimentally that the transfer efficiency in presence of two acceptor dyes can be simply described by adding the transfer rates. This linear competition most likely applies also to the presence of more than two acceptor dyes as long as being in the weak coupling regime. The latter requirement is fulfilled with a high degree of probability if all inter-dye distances are larger than about 50 Å.

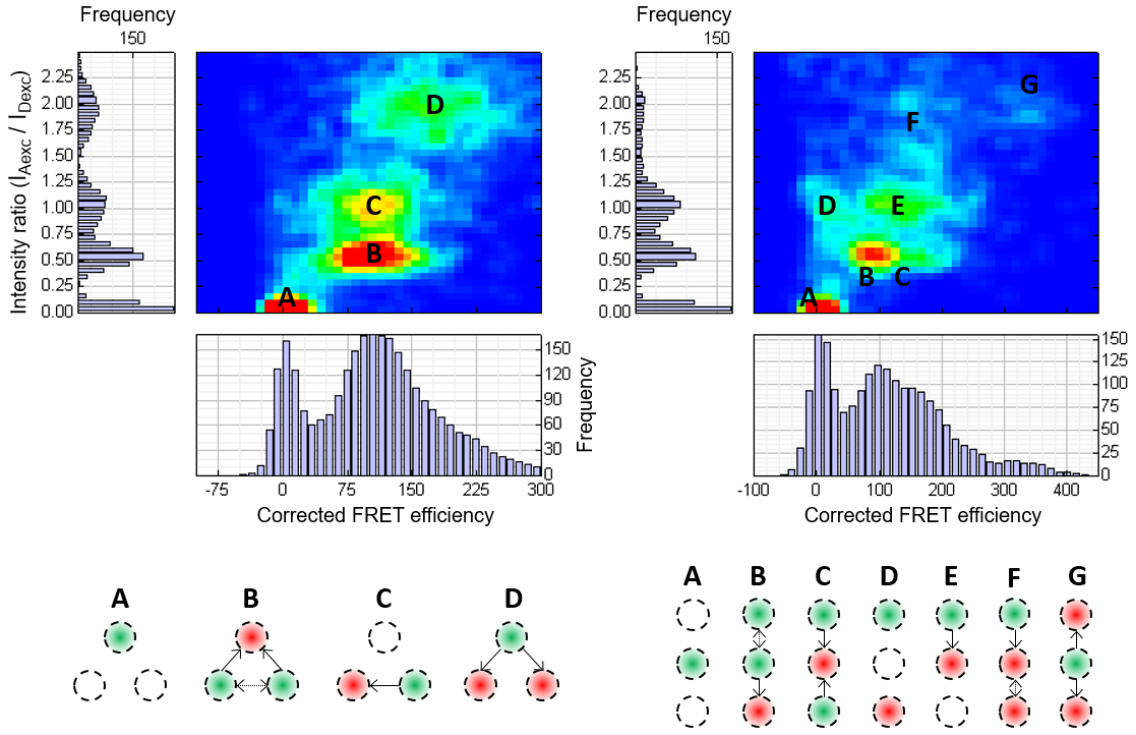


Figure 86: Intensity ratio versus FRET efficiency plots for three dyes out of two species, Atto532 (green) and Atto647N (red). Two different dye arrangements have been used, a triangle geometry with 8 nm dye separation (left) and a linear geometry with 7 nm dye separation (right).

6.3 Microfluidics for dissociation rate and stoichiometry

A microfluidic device was built to directly measure the dissociation rate of the Hsp90 dimer and of protein complexes in general. In contrast to other approaches (*116-118*), this completely new device enables 10000-fold dilution of a concentrated sample and, at the same time, an ultra-fast read out due to a single dilution step. The device allows confocal excitation and readout at a stable single-molecule concentration and at a broad range of timescales spanning millisecond to minutes. Furthermore, a sophisticated protocol enabled overall passivation of the glass slide and the PDMS channels with PEG molecules. Green fluorescing PDMS cross linker molecules were largely eliminated, to obtain acceptable low background fluorescence intensity.

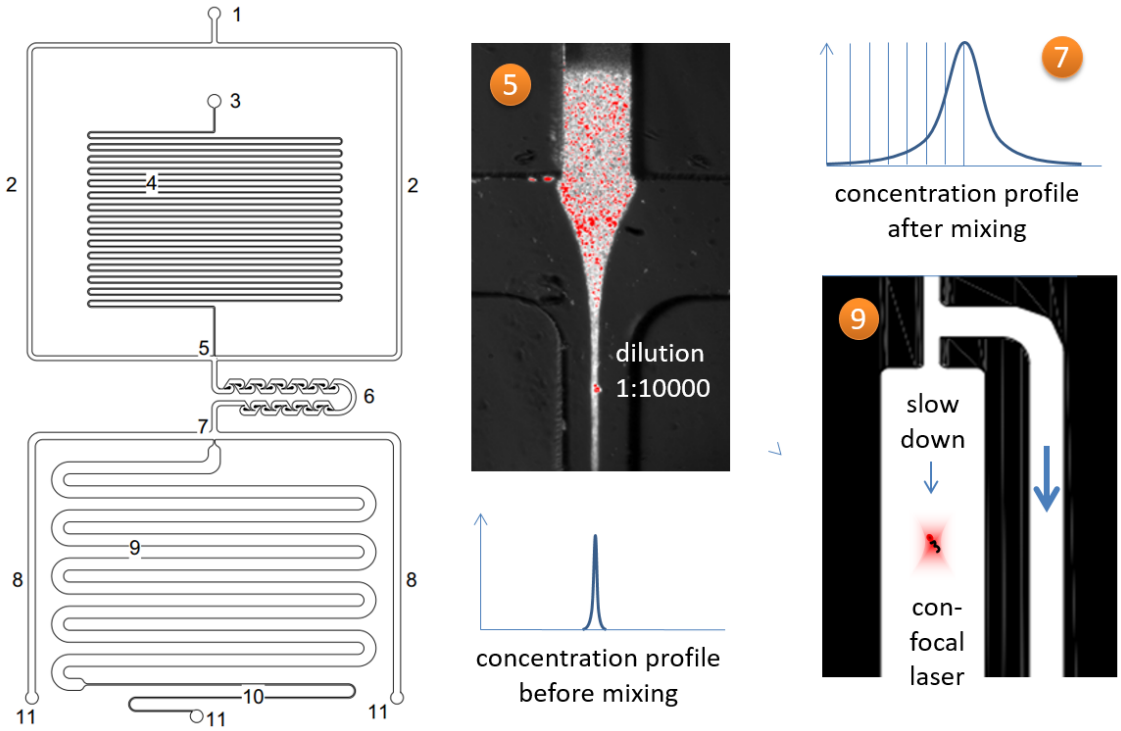


Figure 87: Design of the microfluidic system developed within this work: (1) puffer inlet (2) puffer channels with low hydrodynamic resistance (3) protein inlet (4) protein channel with high resistance (5) dilution by a factor of 1:10000 (6) mixing device (7) split point, where the diluted and mixed sample has a time delay of less than 10 ms after initial dilution (8) outflow with low resistance, in parallel to the measurement channel (9) broad measurement channel for slow flow rates (10) small serial channel with high resistance, in parallel to the outflow (11) outlet.

For microfluidic design and analysis of velocity profiles:

Navier-Stokes equation can be simplified to the Hagen-Poiseuille equation in case of laminar and steady-state flow:

$$\eta \Delta v - \nabla p = 0 \qquad (55)$$

For rectangular cross section of the microfluidic channels (*119*) the volume flow rate can be simplified in case of $h \ll b$ with errors of 13 % for $h=b$ and 0.2 % for $h=b/2$:

$$Q \approx \frac{h^3 b \Delta p}{12 \eta L}\left(1 - 0.63\frac{h}{b}\right) \qquad (56)$$

Combining all constants to R_{hyd} leads to the following relation, which allows calculating the volume flow rates in each microfluidic channel depending on the effective resistance $R_{hyd,eff}$ (analogous to an electrical resistance network):

$$\Delta p = R_{hyd,eff} Q \tag{57}$$

Widths and heights of the device channels were chosen so that fast mixing, sufficient large passage time through the confocal lasers and a broad observation time window were obtained. The dilution factor f was additionally controlled by the inlet pressures:

$$f = \frac{Q_{buffer}}{Q_{protein}} = \frac{\Delta p_{buffer} R_{hyd,protein}}{\Delta p_{protein} R_{hyd,buffer}} \tag{58}$$

For analysis of the kinetics of the Hsp90 homo dimer:

The following ansatz assumes single association and dissociation kinetics for the Hsp90 homo dimer:

$$M + M \underset{k_{dis}}{\overset{k_{ass}}{\rightleftarrows}} D \tag{59}$$

Because of the fast concentration decrease from a concentration above the dissociation constant $K=k_{dis}/k_{ass}$ to a concentration of more than 100 times below the dissociation constant, the association rate k_{ass} can be neglected in comparison the dissociation rate k_{dis}:

$$\frac{d[D]}{dt} = k_{ass}[M]^2 - k_{dis}[D] \approx -k_{dis}[D] \tag{60}$$

This leads to a simple exponential decay:

$$[D] \approx D_0 e^{-k_{dis} t} \tag{61}$$

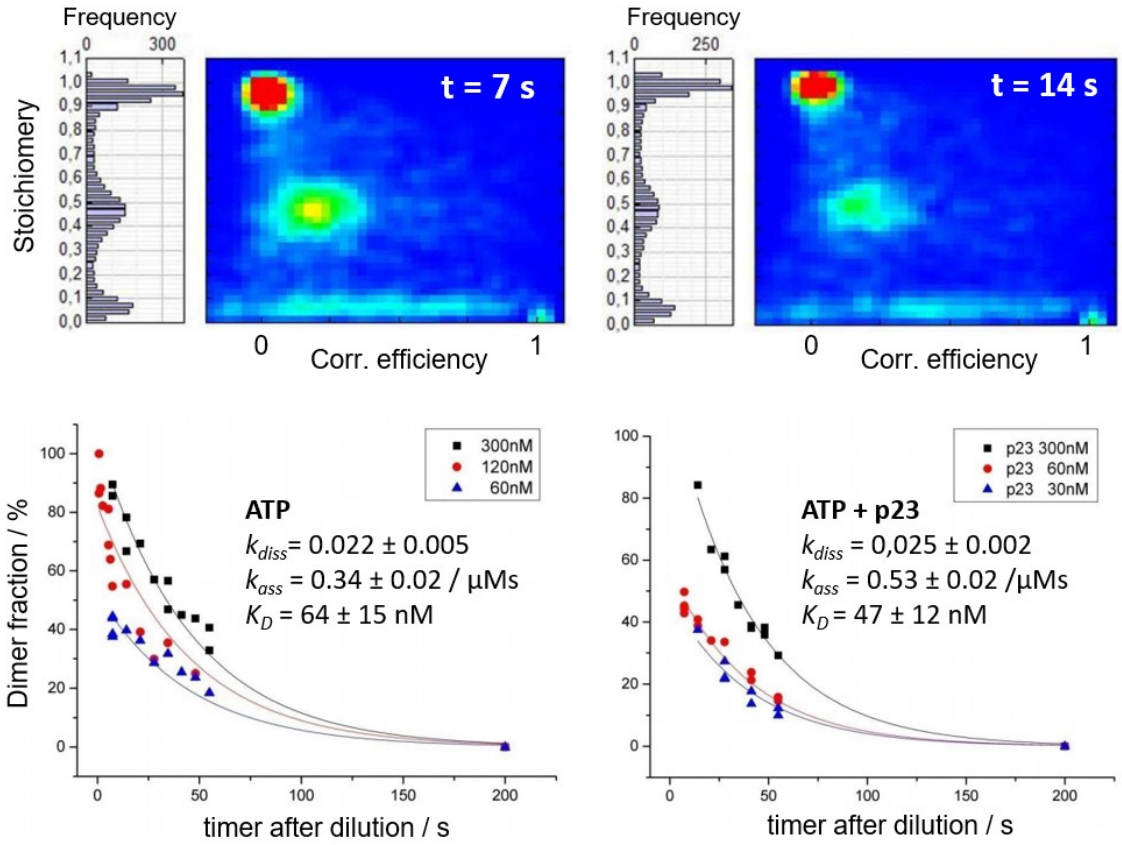


Figure 88: Dissociation of the hsp90 dimer in dependence of nucleotides and co-chaperones. The dissociation rate was calculated from the stoichiometric populations representing monomer and dimer fractions at certain time points in the microfluidic measurement chamber (top). The dimer fraction was fitted with single exponentials revealing dissociation rates for Hsp90 in presence of ATP (bottom left) or in presence of ATP+p23 (bottom right). Varying the starting concentrations in the aggregation chamber allowed obtaining the dimer fraction at time near zero and thus the dissociation constants and association rates.

The dissociation constant K_D for Hsp90 and ATP nicely fits to literature. The dissociation rate k_{dis} is faster than the ATPase rate indicating dissociation before hydrolysis. The association rate k_{ass} in presence of ATP+p23 is faster than the association rate in presence of ATP alone, indicating association at the N-terminal domains. If dissociation happens after or due to hydrolysis, the slower ATPase rate of about 1/60 s^{-1} should be concentration-dependent or the association rate should be limited by the monomer conformation. However, similar dissociation rates for ATP and APO corroborate dissociation before or after hydrolysis, but not due to hydrolysis. Thus, the Hsp90 homodimer might be describable with the morpheein model of allosteric regulation. Notably, the slow transition rates from the open to the closed conformation might be mediated by the zipper-construct for the immobilized molecules. Luckily, the fluctuation model of Hsp90's open state would be still consistent with this dissociation aspect and the FRET efficiency distributions measured with the microfluidics are equal to the ones measured with the zipper-construct.

Finally, the microfluidic device demonstrates a novel way for measuring the stoichiometries and affinities of protein complexes, as demonstrated for the Hsp90-p23 complex in Figure 89.

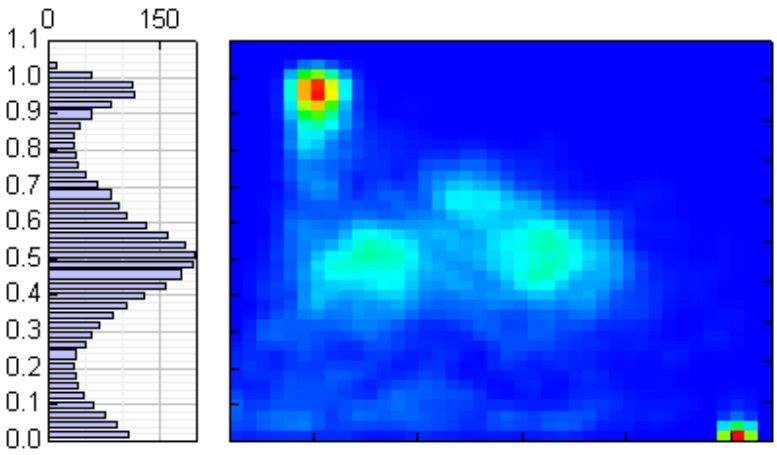


Figure 89: The stoichiometry and affinity of p23 molecules binding to the closed AMPPNP-bound Hsp90 complex was revealed by correlating stoichiometries and FRET efficiencies of single-molecule events, here 100 milliseconds after dilution with the presented microfluidic device. Hsp90 was labeled with Atto532 at position 61 and the co-chaperone p23 was labeled with Atto647N at position 311. 0.15 µM of labeled Hsp90 was exchanged with five-fold excess of wildtype Hsp90 and mixed with 2 µM of labeled p23, and then, the mixture was given to the microfluidic incubation chamber. The plot shows two equally populated states representing single p23 binding at one of the two possible binding positions of Hsp90 and one smaller populated state at a higher stoichiometry representing single p23 binding at one of the two possible binding positions of Hsp90, but with two donor dyes that are present in the symmetric Hsp90 dimer (agreeing with the five-fold wildtype excess). A population with stoichiometries between 0 and 0.5 would represent binding of two p23 molecules, but was not observed indicating a very low affinity for the simultaneous binding of two p23 molecules.

7 REFERENCES

1. M. S. Kim *et al.*, A draft map of the human proteome. *Nature* **509**, 575-581 (2014).
2. N. Savage, Proteomics: High-protein research. *Nature* **527**, S6-7 (2015).
3. U. Hensen *et al.*, Exploring protein dynamics space: the dynasome as the missing link between protein structure and function. *PLoS One* **7**, e33931 (2012).
4. D. T. Haynie, B. Xue, Superdomains in the protein structure hierarchy: The case of PTP-C2. *Protein Sci* **24**, 874-882 (2015).
5. J. M. Eckl, K. Richter, Functions of the Hsp90 chaperone system: lifting client proteins to new heights. *Int J Biochem Mol Biol* **4**, 157-165 (2013).
6. G. E. Karagoz, S. G. Rudiger, Hsp90 interaction with clients. *Trends Biochem Sci* **40**, 117-125 (2015).
7. D. L. Riggs *et al.*, Functional specificity of co-chaperone interactions with Hsp90 client proteins. *Crit Rev Biochem Mol Biol* **39**, 279-295 (2004).
8. M. Taipale *et al.*, A quantitative chaperone interaction network reveals the architecture of cellular protein homeostasis pathways. *Cell* **158**, 434-448 (2014).
9. K. A. Krukenberg, T. O. Street, L. A. Lavery, D. A. Agard, Conformational dynamics of the molecular chaperone Hsp90. *Q Rev Biophys* **44**, 229-255 (2011).
10. L. H. Pearl, C. Prodromou, Structure and mechanism of the Hsp90 molecular chaperone machinery. *Annu Rev Biochem* **75**, 271-294 (2006).
11. M. Taipale, D. F. Jarosz, S. Lindquist, HSP90 at the hub of protein homeostasis: emerging mechanistic insights. *Nat Rev Mol Cell Biol* **11**, 515-528 (2010).
12. L. Neckers, P. Workman, Hsp90 molecular chaperone inhibitors: are we there yet? *Clin Cancer Res* **18**, 64-76 (2012).
13. J. R. Porter, C. C. Fritz, K. M. Depew, Discovery and development of Hsp90 inhibitors: a promising pathway for cancer therapy. *Curr Opin Chem Biol* **14**, 412-420 (2010).
14. K. Moulick *et al.*, Affinity-based proteomics reveal cancer-specific networks coordinated by Hsp90. *Nat Chem Biol* **7**, 818-826 (2011).
15. N. Rohner *et al.*, Cryptic variation in morphological evolution: HSP90 as a capacitor for loss of eyes in cavefish. *Science* **342**, 1372-1375 (2013).

16. G. E. Karagoz *et al.*, Hsp90-Tau complex reveals molecular basis for specificity in chaperone action. *Cell* **156**, 963-974 (2014).

17. T. O. Street *et al.*, Elucidating the mechanism of substrate recognition by the bacterial Hsp90 molecular chaperone. *J Mol Biol* **426**, 2393-2404 (2014).

18. F. Hagn *et al.*, Structural analysis of the interaction between Hsp90 and the tumor suppressor protein p53. *Nat Struct Mol Biol* **18**, 1086-1093 (2011).

19. S. J. Park, B. N. Borin, M. A. Martinez-Yamout, H. J. Dyson, The client protein p53 adopts a molten globule-like state in the presence of Hsp90. *Nat Struct Mol Biol* **18**, 537-541 (2011).

20. P. C. Echeverria, A. Bernthaler, P. Dupuis, B. Mayer, D. Picard, An interaction network predicted from public data as a discovery tool: application to the Hsp90 molecular chaperone machine. *PLoS One* **6**, e26044 (2011).

21. H. Zhang *et al.*, A dynamic view of ATP-coupled functioning cycle of Hsp90 N-terminal domain. *Sci Rep* **5**, 9542 (2015).

22. M. Jahn *et al.*, The charged linker of the molecular chaperone Hsp90 modulates domain contacts and biological function. *Proc Natl Acad Sci U S A* **111**, 17881-17886 (2014).

23. M. Mickler, M. Hessling, C. Ratzke, J. Buchner, T. Hugel, The large conformational changes of Hsp90 are only weakly coupled to ATP hydrolysis. *Nat Struct Mol Biol* **16**, 281-286 (2009).

24. T. Vreven, H. Hwang, B. G. Pierce, Z. Weng, Prediction of protein-protein binding free energies. *Protein Sci* **21**, 396-404 (2012).

25. C. Graf, C. T. Lee, L. Eva Meier-Andrejszki, M. T. Nguyen, M. P. Mayer, Differences in conformational dynamics within the Hsp90 chaperone family reveal mechanistic insights. *Front Mol Biosci* **1**, 4 (2014).

26. K. K. Frederick, M. S. Marlow, K. G. Valentine, A. J. Wand, Conformational entropy in molecular recognition by proteins. *Nature* **448**, 325-329 (2007).

27. D. R. Southworth, D. A. Agard, Client-loading conformation of the Hsp90 molecular chaperone revealed in the cryo-EM structure of the human Hsp90:Hop complex. *Mol Cell* **42**, 771-781 (2011).

28. J. C. Gebhardt, Z. Okten, M. Rief, The lever arm effects a mechanical asymmetry of the myosin-V-actin bond. *Biophys J* **98**, 277-281 (2010).

29. A. Mogilner, M. Mangel, R. J. Baskin, Motion of molecular motor ratcheted by internal fluctuations and protein friction. *Physics Letters A* **237**, 297-306 (1998).

30. N. K. Lee *et al.*, Accurate FRET measurements within single diffusing biomolecules using alternating-laser excitation. *Biophys J* **88**, 2939-2953 (2005).

31. B. Schuler, E. A. Lipman, P. J. Steinbach, M. Kumke, W. A. Eaton, Polyproline and the "spectroscopic ruler" revisited with single-molecule fluorescence. *Proc Natl Acad Sci U S A* **102**, 2754-2759 (2005).

32. I. H. Stein, V. Schuller, P. Bohm, P. Tinnefeld, T. Liedl, Single-molecule FRET ruler based on rigid DNA origami blocks. *Chemphyschem : a European journal of chemical physics and physical chemistry* **12**, 689-695 (2011).

33. T. Ha *et al.*, Probing the interaction between two single molecules: fluorescence resonance energy transfer between a single donor and a single acceptor. *Proceedings of the National Academy of Sciences* **93**, 6264-6268 (1996).

34. M. J. Comstock *et al.*, Protein structure. Direct observation of structure-function relationship in a nucleic acid-processing enzyme. *Science* **348**, 352-354 (2015).

35. C. D. Kinz-Thompson, A. K. Sharma, J. Frank, R. L. Gonzalez, Jr., D. Chowdhury, Quantitative Connection between Ensemble Thermodynamics and Single-Molecule Kinetics: A Case Study Using Cryogenic Electron Microscopy and Single-Molecule Fluorescence Resonance Energy Transfer Investigations of the Ribosome. *J Phys Chem B* **119**, 10888-10901 (2015).

36. C. Ratzke, B. Hellenkamp, T. Hugel, Four-colour FRET reveals directionality in the Hsp90 multicomponent machinery. *Nat Commun* **5**, 4192 (2014).

37. L. Skjaerven, A. Martinez, N. Reuter, Principal component and normal mode analysis of proteins; a quantitative comparison using the GroEL subunit. *Proteins* **79**, 232-243 (2011).

38. F. Marinelli, J. D. Faraldo-Gomez, Ensemble-Biased Metadynamics: A Molecular Simulation Method to Sample Experimental Distributions. *Biophys J* **108**, 2779-2782 (2015).

39. J. W. Pitera, J. D. Chodera, On the Use of Experimental Observations to Bias Simulated Ensembles. *J Chem Theory Comput* **8**, 3445-3451 (2012).

40. A. T. Brunger, P. Strop, M. Vrljic, S. Chu, K. R. Weninger, Three-dimensional molecular modeling with single molecule FRET. *J Struct Biol* **173**, 497-505 (2011).

41. S. Kalinin *et al.*, A toolkit and benchmark study for FRET-restrained high-precision structural modeling. *Nat Methods* **9**, 1218-1225 (2012).

42. A. Muschielok *et al.*, A nano-positioning system for macromolecular structural analysis. *Nat Methods* **5**, 965-971 (2008).

43. V. M. O. Kühn, Charge and Energy Transfer Dynamics in Molecular System. *Wiley-VCH* **3 edition**, (2011).

44. P. J. Walla, Modern Biophysical Chemistry: Detection and Analysis of Biomolecules. *Wiley-VCH* **2 edition**, (2015).

45. T. Förster, Zwischenmolekulare Energiewanderung und Fluoreszenz. *Annalen der Physik* **437**, 55-75 (1948).

46. E. J. Bowen, R. Livingston, An Experimental Study of the Transfer of Energy of Excitation between Unlike Molecules in Liquid Solutions1. *Journal of the American Chemical Society* **76**, 6300-6304 (1954).

47. A. L. Speelman *et al.*, Using molecular dynamics and quantum mechanics calculations to model fluorescence observables. *The journal of physical chemistry. A* **115**, 3997-4008 (2011).

48. D. Beljonne, C. Curutchet, G. D. Scholes, R. J. Silbey, Beyond Forster resonance energy transfer in biological and nanoscale systems. *J Phys Chem B* **113**, 6583-6599 (2009).

49. M. Yang, G. R. Fleming, Influence of phonons on exciton transfer dynamics: comparison of the Redfield, Förster, and modified Redfield equations. *Chemical Physics* **275**, 355-372 (2002).

50. A. Munoz-Losa, C. Curutchet, B. P. Krueger, L. R. Hartsell, B. Mennucci, Fretting about FRET: failure of the ideal dipole approximation. *Biophys J* **96**, 4779-4788 (2009).

51. C. B. Anfinsen, Principles that govern the folding of protein chains. *Science* **181**, 223-230 (1973).

52. M. Gruebele, Protein folding: the free energy surface. *Curr Opin Struct Biol* **12**, 161-168 (2002).

53. C. N. Pace, B. A. Shirley, M. McNutt, K. Gajiwala, Forces contributing to the conformational stability of proteins. *FASEB J* **10**, 75-83 (1996).

54. L. Pauling, R. B. Corey, H. R. Branson, The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain. *Proc Natl Acad Sci U S A* **37**, 205-211 (1951).

55. I. V. Gopich, A. Szabo, Influence of diffusion on the kinetics of multisite phosphorylation. *Protein Sci* **25**, 244-254 (2016).

56. C. Du, S. C. Kou, Correlation Analysis of Enzymatic Reaction of a Single Protein Molecule. *Ann Appl Stat* **6**, 950-976 (2012).

57. L. J. Lapidus, W. A. Eaton, J. Hofrichter, Measuring the rate of intramolecular contact formation in polypeptides. *Proc Natl Acad Sci U S A* **97**, 7220-7225 (2000).

58. E. Pollak, P. Talkner, Reaction rate theory: what it was, where is it today, and where is it going? *Chaos (Woodbury, N.Y.)* **15**, 26116 (2005).

59. K. Dill, S. Bromberg, Molecular Driving Forces: Statistical Thermodynamics in Biology, Chemistry, Physics, and Nanoscience *Taylor & Francis Inc; Auflage: 2 Rev ed. (13. Dezember 2010)*, (2010).

60. V. Kudryavtsev *et al.*, Combining MFD and PIE for accurate single-pair Forster resonance energy transfer measurements. *Chemphyschem : a European journal of chemical physics and physical chemistry* **13**, 1060-1078 (2012).

61. B. K. Muller, E. Zaychikov, C. Brauchle, D. C. Lamb, Pulsed interleaved excitation. *Biophysical journal* **89**, 3508-3522 (2005).

62. B. K. Volker Buschmann, Felix Koberling, PicoQuant GmbH, Quantitative FCS: Determination of the Confocal Volume by FCS and Bead Scanning with the MicroTime 200.

63. K. Bacia, P. Schwille, Practical guidelines for dual-color fluorescence cross-correlation spectroscopy. *Nat Protoc* **2**, 2842-2856 (2007).

64. M. Hoeller, Advanced Fluorescence Fluctuation Spectroscopy with Pulsed Interleaved Excitation. *Dissertation, LMU München*, (2011).

65. W. Becker, The bh TCSPC Handbook, 6th edition. (2015).

66. R. M. Clegg, in *Laboratory Techniques in Biochemistry and Molecular Biology*. (Elsevier, 2009), vol. Volume 33, pp. 1-57.

67. A. Gust *et al.*, A starting point for fluorescence-based single-molecule measurements in biomolecular research. *Molecules* **19**, 15824-15865 (2014).

68. B. Schuler, Single-molecule FRET of protein structure and dynamics - a primer. *Journal of nanobiotechnology* **11 Suppl 1**, S2 (2013).

69. T. Cordes, A. Maiser, C. Steinhauer, L. Schermelleh, P. Tinnefeld, Mechanisms and advancement of antifading agents for fluorescence microscopy and single-molecule spectroscopy. *Phys Chem Chem Phys* **13**, 6699-6709 (2011).

70. J. Widengren, R. Rigler, U. Mets, Triplet-state monitoring by fluorescence correlation spectroscopy. *J Fluoresc* **4**, 255-258 (1994).

71. D. S. Banks, C. Fradin, Anomalous Diffusion of Proteins Due to Molecular Crowding. *Biophysical Journal* **89**, 2960-2971 (2005).

72. E. Nir *et al.*, Shot-noise limited single-molecule FRET histograms: comparison between theory and experiments. *J Phys Chem B* **110**, 22103-22124 (2006).

73. X. Michalet *et al.*, The power and prospects of fluorescence microscopies and spectroscopies. *Annu Rev Biophys Biomol Struct* **32**, 161-182 (2003).

74. W. Kuegel, Development and application of advanced single molecule fluorescence methods using PIE-MFD. *Ludwig-Maximilians-Universitaet Muenchen*, (2012).

75. J. P. Torella, S. J. Holden, Y. Santoso, J. Hohlbein, A. N. Kapanidis, Identifying molecular dynamics in single-molecule FRET experiments with burst variance analysis. *Biophys J* **100**, 1568-1577 (2011).

76. S. Kalinin, A. Valeri, M. Antonik, S. Felekyan, C. A. Seidel, Detection of structural dynamics by FRET: a photon distribution and fluorescence lifetime analysis of systems with multiple states. *J Phys Chem B* **114**, 7983-7995 (2010).

77. I. V. Gopich, A. Szabo, Theory of the energy transfer efficiency and fluorescence lifetime distribution in single-molecule FRET. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 7747-7752 (2012).

78. T. Torres, M. Levitus, Measuring conformational dynamics: a new FCS-FRET approach. *J Phys Chem B* **111**, 7392-7400 (2007).

79. D. Case *et al.*, Amber 14. (2014).

80. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics* **79**, 926-935 (1983).

81. J. A. Maier *et al.*, ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *Journal of Chemical Theory and Computation*, (2015).

82. T. Darden, D. York, L. Pedersen, Particle mesh Ewald: An N· log (N) method for Ewald sums in large systems. *The Journal of chemical physics* **98**, 10089-10092 (1993).

83. A. Shrake, J. Rupley, Environment and exposure to solvent of protein atoms. Lysozyme and insulin. *J Mol Biol* **79**, 351-371 (1973).

84. T. O. Street, L. A. Lavery, D. A. Agard, Substrate binding drives large-scale conformational changes in the Hsp90 molecular chaperone. *Molecular cell* **42**, 96-105 (2011).

85. M. Hessling, K. Richter, J. Buchner, Dissection of the ATP-induced conformational cycle of the molecular chaperone Hsp90. *Nat Struct Mol Biol* **16**, 287-293 (2009).

86. V. Ivanov, M. Li, K. Mizuuchi, Impact of emission anisotropy on fluorescence spectroscopy and FRET distance measurements. *Biophysical journal* **97**, 922-929 (2009).

87. J. A. Morales, E. Navarro, Minkowskian description of polarized light and polarizers. *Physical review. E, Statistical, nonlinear, and soft matter physics* **67**, 026605 (2003).

88. S. Kalinin, S. Felekyan, A. Valeri, C. A. Seidel, Characterizing multiple molecular states in single-molecule multiparameter fluorescence detection by probability distribution analysis. *The Journal of Physical Chemistry B* **112**, 8361-8374 (2008).

89. M. Antonik, S. Felekyan, A. Gaiduk, C. A. Seidel, Separating structural heterogeneities from stochastic variations in fluorescence resonance energy transfer distributions via photon distribution analysis. *The Journal of Physical Chemistry B* **110**, 6970-6978 (2006).

90. I. Gopich, A. Szabo, Theory of photon statistics in single-molecule Forster resonance energy transfer. *The Journal of chemical physics* **122**, 14707 (2005).

91. G. Arfken, The method of steepest descents. *Mathematical methods for physicists* **3**, 428-436 (1985).

92. M. M. Ali *et al.*, Crystal structure of an Hsp90-nucleotide-p23/Sba1 closed chaperone complex. *Nature* **440**, 1013-1017 (2006).

93. K. Sidera, E. Patsavoudi, HSP90 inhibitors: current development and potential in cancer therapy. *Recent Pat Anticancer Drug Discov* **9**, 1-20 (2014).

94. O. Genest *et al.*, Uncovering a region of heat shock protein 90 important for client binding in E. coli and chaperone function in yeast. *Mol Cell* **49**, 464-473 (2013).

95. D. E. Dollins, J. J. Warren, R. M. Immormino, D. T. Gewirth, Structures of GRP94-nucleotide complexes reveal mechanistic differences between the hsp90 chaperones. *Molecular cell* **28**, 41-56 (2007).

96. T. O. Street *et al.*, Cross-monomer substrate contacts reposition the Hsp90 N-terminal domain and prime the chaperone activity. *J Mol Biol* **415**, 3-15 (2012).

97. L. A. Lavery *et al.*, Structural asymmetry in the closed state of mitochondrial Hsp90 (TRAP1) supports a two-step ATP hydrolysis mechanism. *Mol Cell* **53**, 330-343 (2014).

98. A. K. Shiau, S. F. Harris, D. R. Southworth, D. A. Agard, Structural Analysis of E. coli hsp90 reveals dramatic nucleotide-dependent conformational rearrangements. *Cell* **127**, 329-340 (2006).

99. C. Ratzke, F. Berkemeier, T. Hugel, Heat shock protein 90's mechanochemical cycle is dominated by thermal fluctuations. *Proc Natl Acad Sci U S A* **109**, 161-166 (2012).

100. B. Musafia, V. Buchner, D. Arad, Complex salt bridges in proteins: statistical analysis of structure and function. *J Mol Biol* **254**, 761-770 (1995).

101. D. Xu, C.-J. Tsai, R. Nussinov, Hydrogen bonds and salt bridges across protein-protein interfaces. *Protein engineering* **10**, 999-1012 (1997).

102. K. A. Henzler-Wildman *et al.*, A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature* **450**, 913-916 (2007).

103. S. C. Kou, X. S. Xie, Generalized Langevin equation with fractional Gaussian noise: subdiffusion within a single protein molecule. *Phys Rev Lett* **93**, 180603 (2004).

104. D. Arti, R. Adhikari, Non-Markovian fluctuations in Markovian models of protein dynamics. *Journal of Statistical Mechanics: Theory and Experiment* **2011**, P04017 (2011).

105. W. Doster, The protein-solvent glass transition. *Biochim Biophys Acta* **1804**, 3-14 (2010).

106. T. Kawasaki, D. Coslovich, A. Ikeda, L. Berthier, Diverging viscosity and soft granular rheology in non-Brownian suspensions. *Physical review. E, Statistical, nonlinear, and soft matter physics* **91**, 012203 (2015).

107. Y. Wang, G. Zocchi, Viscoelastic transition and yield strain of the folded protein. *PLoS One* **6**, e28097 (2011).

108. K. Richter *et al.*, Conserved conformational changes in the ATPase cycle of human Hsp90. *J Biol Chem* **283**, 17757-17765 (2008).

109. F. H. Stillinger, P. G. Debenedetti, in *Annual Review of Condensed Matter Physics, Vol 4,* J. S. Langer, Ed. (2013), vol. 4, pp. 263-285.

110. T. Mittag, L. E. Kay, J. D. Forman-Kay, Protein dynamics and conformational disorder in molecular recognition. *J Mol Recognit* **23**, 105-116 (2010).

111. C. Schulenburg, D. Hilvert, Protein conformational disorder and enzyme catalysis. *Top Curr Chem* **337**, 41-67 (2013).

112. M. Chakraborty *et al.*, Protein Oligomerization Monitored by Fluorescence Fluctuation Spectroscopy: Self-Assembly of Rubisco Activase. *Biophysical Journal* **103**, 949-958 (2012).

113. M. A. Digman, P. W. Wiseman, C. Choi, A. R. Horwitz, E. Gratton, Stoichiometry of molecular complexes at adhesions in living cells. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 2170-2175 (2009).

114. G. Malengo *et al.*, Fluorescence Correlation Spectroscopy and Photon Counting Histogram on membrane proteins: Functional dynamics of the GPI-anchored Urokinase Plasminogen Activator Receptor. *Journal of biomedical optics* **13**, 031215-031215 (2008).

115. A. Orte, R. W. Clarke, D. Klenerman, Single-molecule fluorescence coincidence spectroscopy and its application to resonance energy transfer. *Chemphyschem : a European journal of chemical physics and physical chemistry* **12**, 491-499 (2011).

116. Y. Gambin *et al.*, Visualizing a one-way protein encounter complex by ultrafast single-molecule mixing. *Nat Meth* **8**, 239-241 (2011).

117. B. Wunderlich *et al.*, Microfluidic mixer designed for performing single-molecule kinetics with confocal detection on timescales from milliseconds to minutes. *Nat Protoc* **8**, 1459-1474 (2013).

118. M. H. Horrocks *et al.*, Single-Molecule Measurements of Transient Biomolecular Complexes through Microfluidic Dilution. *Analytical Chemistry* **85**, 6855-6859 (2013).

119. H. Bruus, Theoretical Microfluidics. *OUP Oxford*, (2008).