PHYSIK-DEPARTMENT

TECHNISCHE UNIVERSITÄT MÜNCHEN

Modeling of Protein Assemblies Through Global Docking and Accelerated Molecular Dynamics Simulation

Danial Pourjafar-Dehkordi

FAKULTÄT FÜR PHYSIK

TECHNISCHE UNIVERSITÄT MÜNCHEN

Modeling of Protein Assemblies Through Global Docking and Accelerated Molecular Dynamics Simulation

Danial Pourjafar-Dehkordi

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. Franz Pfeiffer

Prüfer der Dissertation:

1. Prof. Dr. Martin Zacharias

2. Prof. Dr. Karen Alim

Die Dissertation wurde am 17.09.2021 bei der Technischen Universität München eingereicht und durch die Fakultät für Physik am 20.01.2022 angenommen.

Abstract

Proteins are the building blocks of the cell. Critical functions are driven and regulated by assemblies formed by proteins and other biomolecules (such as DNAs, RNAs). Understanding the structure of protein assemblies plays a pivotal role in discovering their function. Although the recent years have seen a great progress in experimental structure determination methods, there are still challenges and limitations, largely due to the highly dynamic and impressively diverse nature of protein assemblies.

Complementary to the experimental methods, molecular dynamics (MD) simulations provide an atomistic view of a wide range of biomolecular processes. The focus of this work is on the application of MD simulations in the structural analysis of protein assemblies. An overview of the current protein-protein docking tools and methodologies for predicting the complex structures is presented. Next, protein complexes of two members of the Rab small GTPases, Rab8a and Rab1b, that regulate cellular membrane trafficking are scrutinized using MD and free-energy simulations. The influence of post-translational modifications of Rab8a on its binding to the exchange factor Rabin8 is studied. Furthermore, using dihedral-angle biasing potential replica-exchange method, structural flexibility of the wild type and S111-phosphorylated Rab1b in complex with GTP/GDP is evaluated. Finally, a self-learning accelerated-sampling scheme that specifically identifies and compensates for low free-energy conformations is introduced and utilized to explore the large-scale conformational changes in bacterial Argonaute protein in the absence of substrate and upon binding to guide and target DNA strands. The effectiveness of the accelerated dynamics scheme is validated by the insight it provides from the transition mechanisms leading to the activation of the Argonaute protein.

Zusammenfassung

Proteine sind die Bausteine der Zelle. Kritische Funktionen werden durch Zusammenschlüsse von Proteinen und anderen Biomolekülen (z. B. DNAs, RNAs) gesteuert und reguliert. Das Verständnis der Struktur von Protein-Assemblies spielt eine entscheidende Rolle bei der Entdeckung ihrer Funktion. Obwohl in den letzten Jahren große Fortschritte bei den experimentellen Methoden zur Strukturbestimmung erzielt wurden, gibt es immer noch Herausforderungen und Einschränkungen, die größtenteils auf die hochdynamische und erstaunlich vielfältige Natur von Protein-Assemblies zurückzuführen sind.

Ergänzend zu den experimentellen Methoden bieten Molekulardynamik (MD)-Simulationen einen atomistischen Einblick eine große Vielfalt von biomolekularen Prozessen. Der Schwerpunkt dieser Arbeit liegt auf der Anwendung von MD-Simulationen bei der Strukturanalyse von Protein-Assemblies. Es wird ein Überblick über die aktuellen Protein-Protein-Docking-Tools und Methoden zur Vorhersage der komplexen Strukturen vorgestellt. Anschließend werden zwei Mitglieder der kleinen Rab-GTPasen, Rab8a und Rab1b, die den zellulären Membrantransport regulieren, mit Hilfe von MD- und Freie-Energie-Simulationen unter die Lupe genommen. Der Einfluss von posttranslationalen Modifikationen von Rab8a auf seine Bindung an den Austauschfaktor Rabin8 wird untersucht. Des weiteren wird die strukturelle Flexibilität des Wildtyps und des S111-phosphorylierten Rab1b im Komplex mit GTP/GDP unter Verwendung der Dihedral-Angle Biasing Potential Replica-Exchange-Methode untersucht. Schließlich wird ein selbstlernendes Accelerated-Sampling-Schema eingeführt, das spezifisch Konformationen mit niedriger freier Energie identifiziert und kompensiert, um die großräumigen Konformationsänderungen des bakteriellen Argonaute-Proteins in Abwesenheit von Substrat und bei Bindung an Leit- und Ziel-DNA-Stränge zu untersuchen. Die Effektivität des beschleunigten Dynamikschemas wird durch den Einblick in die Übergangsmechanismen, die zur Aktivierung des Argonaute-Proteins führen, validiert.

Contents

1. Introduction

Inside the cell is a strange place. The diversity of the biological macromolecules found there has a great deal of wonder for curious observers. They vary in shape, weight, and function, but all are necessary for the life cycle of cells.

The information of life is preserved in one simple coded dimension–the genetic information–and is transferred between nucleic acids and proteins. Deoxyribonucleic acids (DNAs) contain all the genetic information and are formed along a strand by a sequence of four different nucleotide bases. Each base is comprised of a phosphate group, sugar, and one of the four organic bases: adenine, cytosine, guanine, or thymine (denoted by A, C, G, T). A binds to T and G binds to C, making each strand of the DNA duplex complementary to the other. Ribonucleic acid (RNA) is usually found as single strand helix and has a slightly different base pairing since it uses the base uracil (U) instead of thymine. The genetic information stored in DNA is transcribed to RNA and then translated into a sequence of amino acids forming the protein. This is known as the central dogma of molecular biology. As it will be discussed later, the sequence of amino acid residues has a direct impact on the physical properties of the protein and therefore, the genes not only determine proteins' presence or absence, but also their way of operation.

The building blocks of proteins are the amino acids, all of which share a basic composition: an amino group, a central carbon atom (Cα) along with a sidechain specific to the amino acid, and a carboxyl group. The peptide linkage that connects all the amino acids is a single covalent bond between the carbon atom of the carboxyl group and the nitrogen atom from the amino group. The juxtaposition of amino acids gives rise to their secondary structure–either an α-helices or a β-sheets. The overall three-dimensional structure of a protein–also known as tertiary structure–takes a compact globular shape or an elongated conformation (Figure 1-1)

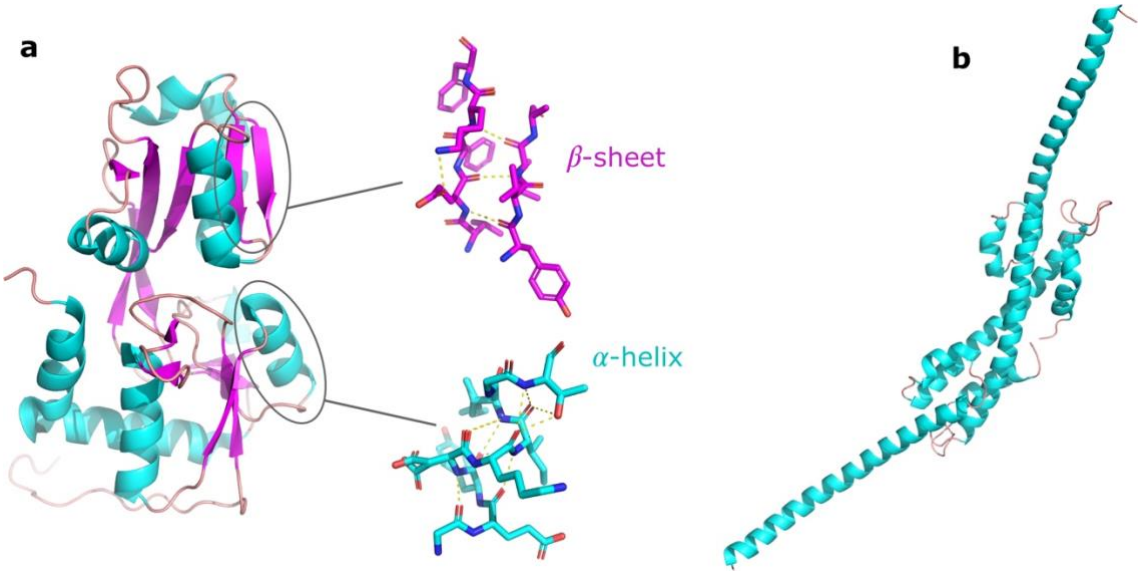

Figure 1-1. Protein's tertiary structure.
The 3D structure of a protein is composed of local folded segments that form α-helices (cyan) or β-sheets (magenta), both held in place by hydrogen bonds (yellow dashed line), as illustrated in cartoon representation of two sample proteins: **a**, substrate binding domain 2 of ABC transporter GlnPQ (Protein Data Bank (PDB): 4zef (1)) and **b**: dimeric structure of elongated fibrous segment of herpes simplex virus type 1 (PDB: 4tt0 (2)).

The common amino acids are considered to be only 20, each of which has certain features owing to the unique sidechain it contains. Glycine, for instance, has only a hydrogen atom as the sidechain, which allows it to adopt unusual dihedral angles and thereby, increasing the polypeptide chain's flexibility. Branched sidechains, seen in valine, isoleucine, and leucine, are stiffer and easier to be fixed in a certain position, which hardens the main chain. The amino acid properties are governed by their sidechain's electrostatic charge distribution. Polar sidechains form hydrogen bonds to water and other molecules. Neutral polar residues are found at the surface as well as inside proteins, while the nonpolar residues are buried inside the protein fold. As internal residues they usually form hydrogen bonds with each other or with the polypeptide backbone. Charged sidechains are commonly found at the molecular surface. Theory and experiments have shown that the protein folding is governed by a number of factors. Hydrogen bonds between the polar sidechains, van der Waals interactions of the tightly packed residues, hydrophobic interactions of the non-polar residues and the attractive and repulsive forces between electrostatically-charged residues are among the major contributors (3). These intra-molecular forces guide the protein towards adopting a tertiary structure in which it is active. This arrangement is known as the native conformation.

The proteins are not always readily in their active form. In fact, they are considered to be most of the times rather inert, waiting to be activated, or degraded and replaced. The switching of proteins between active and inactive states is widely encountered in nature and allows a dynamic regulation of processes that are essential for cell's survival, such as cell division,

signaling and transcriptional regulations (4). These mechanisms may be triggered by conformational changes, binding to effectors, or post-translational modification (PTM).

PTM is a reversible biochemical process in which the amino acid residues of a protein are covalently modified after its synthesis. Modifications can be of numerous types namely, phosphorylation, methylation, ubiquitination, etc. They increase both the structural and functional diversity of proteins and are known to facilitate protein folding and enhance the function (5). Moreover, PTMs can influence the stability of proteins and interfere in their interactions with other proteins. The majority of proteins operate not as isolated molecules, but as components of larger macromolecular assemblies that drive certain biological functions. Proteins are usually surrounded by many potential binding partners, creating a crowded environment. The vast majority of proteins are specific in the choice of binding partners–they must have complementary shape and surface physical properties. On the other hand, many proteins are found to be involved in more than one complex (6). Understanding the effects of a covalent modification on the protein itself, and on its association with other proteins is of great importance in deciphering their function and their biological pathways.

Using high-resolution structural characterization techniques one can determine the final structure of protein complexes. However, during the binding process proteins usually transition to a number of intermediate states that, for a thorough understanding of the process, are of equal significance. In the work presented here, we use atomistic molecular dynamics (MD) combined with advanced sampling techniques to better understand the entire process of conformational transitions upon complex formation. A summary of the most important computational protein-complex prediction methods will be presented, some of which have reached a promising efficiency in protein complex detection (Chapter 3). Using MD simulations and free-energy calculations, we will study the influence of PTMs on two GTPase proteins involved in membrane trafficking processes (Chapters 4 and 5). Moreover, we will determine the contribution of different sidechain arrangements on the binding process with other effectors. Finally, we will present an enhanced sampling scheme that uses a self-learning algorithm for exploring the conformational landscape of large protein domain motions (Chapter 6). The application of the technique in bacterial Argonaute protein will be presented, where the interplay between the protein and DNAs creates a beautiful illustration of the dynamic conformational changes observed in biomolecules. But first, let us go through the fundamental theories used in this work.

2. Theory

2.1. Molecular dynamics simulations

In the work presented here, classical mechanics is used to describe the motions of atoms and molecules. In classical mechanics, an atom is treated as one particle that obeys Newton's equation of motion,

$$F_i = \frac{\partial^2 r_i(t)}{\partial t^2} m_i, \qquad 2\text{-}1$$

where F_i is the force exerted on atom i with the mass m_i, located at the position r_i at time t. Forces are calculated from a potential-energy function V–called "force field",

$$F_i = \frac{\partial V(r_1, \ldots, r_N)}{\partial r_i}. \qquad 2\text{-}2$$

In the force field used here–from the Amber software package (7)–the potential-energy function consists of four terms,

$$V = \sum_{bonds} k_b(r - r_0)^2 + \sum_{angles} k_\theta(\theta - \theta_0)^2 + \sum_{dihedrals} V_n[1 + cos(n\emptyset - \gamma)] + \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\varepsilon R_{ij}}\right], \qquad 2\text{-}3$$

where the first two terms represent the covalent-bond contributions using harmonic potential functions with the equilibrium distance and angle, r_0 and θ_0, and spring constants k_b and k_θ. The dihedral angles between bound atoms are represented using a periodic cosine function that has n minima, the scaling factor V_n and a phase shift of γ. The last term describes the non-bonded interactions namely, the van der Waals forces between atoms i and j with the parameters A_{ij} and B_{ij}, and the electrostatic interactions between the atom pair with partial

charges q_i and q_j, calculated according to Coulomb's law. Parameters for the bond length, angles and dihedral contributions are typically derived from quantum-chemical calculations. The Newton's equation of motion is integrated using the Verlet algorithm (8), which is derived from a truncated Taylor expansion of the coordinate of particle at time t,

$$r_i(t + \Delta t) = r_i(t) + \frac{\partial r_i(t)}{\partial t}\Delta t + \frac{1}{2}\frac{\partial^2 r_i(t)}{\partial t^2}\Delta t^2.$$

Similarly,

$$r_i(t - \Delta t) = r_i(t) - \frac{\partial r_i(t)}{\partial t}\Delta t + \frac{1}{2}\frac{\partial^2 r_i(t)}{\partial t^2}\Delta t^2.$$

The Δt is the time step. Summing these two equations, we have

$$r_i(t + \Delta t) = 2r_i(t) - r_i(t - \Delta t) + \frac{\partial^2 r_i(t)}{\partial t^2}\Delta t^2.$$

2.1.1 Temperature and pressure

In the previous section, we defined the internal interactions of a system composed of *N* particles and showed how one can simulate the evolution of such system over time by solving the Newton's equation of motion iteratively. Nevertheless, a realistic modelling of the biomolecules' surroundings requires replication of the relevant temperatures and pressures. The instantaneous temperature of a body of *N* particles at time *t* is,

$$T(t) = \sum_{i=1}^{N} \frac{m_i v_i^2}{N_{df} k_B},$$

where N_{df} is the particles' degrees of freedom, m_i their mass and v_i their velocity. There are several formalisms for adjusting the temperature, namely the Berendsen and Nosé-Hoover thermostats (9,10). The fundamental idea, however, relies on rescaling the velocities using a new auxiliary term in the Hamiltonian, such that the overall temperature remains close to a reference temperature. To simulate at constant pressure, the barostatic formalism, such as Berendsen (9) and Parrinello-Rahman (11), consider the cell's (simulation box) size and shape variables that are dynamic and can change during the simulation.

2.1.2 Periodic boundary condition

In the physiological conditions that MD aims to model, the protein or DNA of interest is surrounded by copious water molecules and ions. On the other hand, the computational cost of modelling large number of solvent molecules is very high. The common workaround is to use periodic boundary conditions. To mimic large systems, images of a smaller–though sufficiently sized–simulation box that contains the solvated molecule of interest in physiological ion concentrations are juxtaposed in all directions. The images are identical and therefore, only the atoms within the central box need to be simulated. Atoms can enter or leave the box and leaving of an atom from one side entails its entering of the opposing side. The larger box, however, is imitating the environment of a much larger solvent box.

2.1.3 Cut-off radius

The large number of atoms that the simulation box contains increases the computational cost of pairwise van der Waals- and electrostatic-force calculations. To avoid this, we use a cut-off radius of 9 Å and neglect the short-range atomic interactions beyond that distance. The potentials at the cut-off radius are set to zero to avoid discontinuity in the system's energy function. With the application of a cut-off radius, we cannot simply neglect the long-range interactions; instead, we take advantage of the periodic boundary condition in a technique called Particle Mesh Ewald (PME).

2.1.4 Particle mesh Ewald Method

In PME, the Coulomb potential is the summation of two terms,

$$V_c = \sum_{i,j} V_{sr} + \sum_{\boldsymbol{k}} \widetilde{\Phi}_{lr}(\boldsymbol{k})|(\tilde{\rho}(\boldsymbol{k}))|^2.$$

The first term sums the short-range electrostatic interactions, and the second term calculates the long-range interactions based on the Fourier transforms of the potential ($\widetilde{\Phi}_{lr}$) and the charge density ($\tilde{\rho}$). The short-ranged interactions are calculated with high accuracy using the actual charge positions, while the long-range calculations are accelerated by interpolation of charge positions on a mesh which is provided by the periodic boundary conditions.

2.2. Enhanced sampling methods

If one considers all arrangements of dihedral angle and sidechain rotations, the number of possible states for a protein grows exponentially. These states, owing to the force field constants,

are separated by large energy barriers and the MD simulations are usually stuck in one with only small fluctuations of bond lengths and angles observed. Such states are known as local energy minimum. Since the early days of molecular simulations, enhancement of the sampling to go beyond energy minima and capture more relevant states within limited simulation times has been the focus of many researchers. These efforts have resulted in development of several fruitful techniques. In Chapter 5 we will introduce and utilize a replica-exchange technique that promotes sidechain and dihedral-angle rotations. In Chapter 6 we will report on another accelerated-sampling technique, which has some similarities to the Metadynamics method. Therefore, it is useful to briefly introduce the two methods. Further details on our implementations will be discussed in the chapters.

2.2.1 Metadynamics

A very well-established class of enhanced sampling methods–called Metadynamics–involves filling of the energy minima in a controlled and history-dependent manner using a bias potential–typically, a Gaussian function (12,13). Metadynamics technique runs in the following procedure: i) running a short simulation, ii) computing the histogram ($H_t(s)$) of the collective variable (CV) of interest, iii) updating the bias,

$$B_{t+1}(s) = B_t(s) + T\log\big(H_t(s)\big),$$

and iv) resuming the simulations. In the first phase, the bias is zero and the system will remain in the first local minimum. In the next runs, the bias is introduced and updated, and thus the system explores a different range of CV values. The approximation of the histogram in Metadynamics is a Gaussian of width σ and height w,

$$T\log(H_t(s)) \sim w\exp\left(-\frac{(s_t - s)^2}{2\sigma^2}\right).$$

This results in,

$$B_t(s) = w\sum_{t\prime < t}\exp\left(\frac{(s_{t\prime} - s)^2}{2\sigma^2}\right).$$

This sum causes increasingly higher deviations in the CVs. At some point, the energy minimum is entirely filled with the bias and the system moves to other states.

Metadynamics has been successfully applied to a variety of problems and its development it still ongoing (14,15). One of its advantages is that the CV histogram is provides a direct estimate

of the free energy. The choice of CV in metadynamics, on the other hand, is of critical importance and sometimes a major difficulty. While CVs are typically a lower-dimensional projection of the atomic coordinates, they must represent the energy-minima states as well. Moreover, the values of the CVs should be distinguishable in the minima and in the transition states. Finally, one should avoid having too many of them, as each CV adds a dimension in the space that needs to be filled with the bias potential, which makes the simulation costly.

2.2.2 Hamiltonian replica-exchange MD method

Another strategy to augment MD sampling is the replica-exchange technique. Here we briefly outline the formulation of the Hamiltonian replica-exchange method used as the advanced sampling technique throughout this study. The probability of configuration X in the mth replica obeys the Boltzmann distribution $P_m(X)$ at the temperature T_m,

$$P_m(X) = \frac{1}{Z_m}\exp(-\beta_m \mathcal{H}(X)),$$

where β_m is the inverse temperature, $1/k_B T_m$ with k_B being the Boltzmann constant, $\mathcal{H}(X)$ is the Hamiltonian and Z_m is the partition function. The overall probability of the system with M replicas, P_{all}, is the multiplication of the probability of each replica,

$$P_{all} = \prod_{i}^{M} P_i(X_i).$$

The probability that the configuration X of the mth replica is exchanged with configuration X' of the nth replica is the transition probability, written as,

$$W(X, \mathcal{H}_m; X', \mathcal{H}_n)$$

The probability of the reverse process is,

$$W(X', \mathcal{H}_m; X, \mathcal{H}_n)$$

The balance condition for the extended system to reach Boltzmann equilibrium requires

$$P_{all}[(\dots; X, \mathcal{H}_m; X', \mathcal{H}_n; \dots)]W(X, \mathcal{H}_m; X', \mathcal{H}_n) = P_{all}[(\dots; X', \mathcal{H}_m; X, \mathcal{H}_n; \dots)]W(X', \mathcal{H}_m; X', \mathcal{H}_n).$$

This leads to

$$\frac{W(X,\mathcal{H}_m;X',\mathcal{H}_n)}{W(X',\mathcal{H}_m;X,\mathcal{H}_n)} = \exp(-\Delta),$$

where,

$$\Delta \equiv \beta\{[\mathcal{H}_m(X') + \mathcal{H}_n(X)] - [\mathcal{H}_m(X) + \mathcal{H}_n(X')]\}.$$

Based on the Metropolis criteria for the transition probability,

$$\begin{cases} W(X,\mathcal{H}_m;X',\mathcal{H}_n) = 1 & for \quad \Delta \leq 0 \\ \exp(-\Delta) & for \quad \Delta > 0 \end{cases}.$$

That is, the exchange is accepted if the result in a lower overall energy level. If the overall energy is unfavorable, the exchange is accepted only according to the Boltzmann weighted probability. The exchange attempts are done between the neighboring replicas (16).

2.3. Free energy calculations based on MD simulations

Describing the thermodynamics of a system using free energy has been one of the most important objectives of biomolecular simulations. Having an estimation of a system's free energy can minimize the need for experimental measurements and has many useful applications in various fields, namely in drug design. In many cases, the aim is to calculate the free-energy difference between two states that are either different in conformation or are the bound/unbound states of two ligands to a single receptor. To this end, one needs to have an accurate-enough physical model of the biological system and be able to sample the regions in the phase-space that correspond to those states. This implies, however, two limitations; one is the accuracy of the current force fields, and the second is insufficient sampling. Improvement of the force fields is an everlasting goal of the researchers in this field. The latter also can be overcome by coupling free-energy calculations to enhanced sampling techniques, such as Metadynamics or Hamiltonian replica-exchange MD. There exist a number of techniques that allow calculation of the free-energy differences between two well-defined thermodynamic states, such as

thermodynamic integration or free energy perturbation (FEP) theory, both of which rely on the notion of a coupling parameter (17–21). Here, we will introduce the FEP theory.

Consider a biomolecular system with constant number of particles, *N*, at a given set of coordinates $\boldsymbol{q}$ and momenta $\boldsymbol{p}$. The partition function, Z, for such a system is,

$$Z = \frac{1}{h^{3N}N!}\iint \exp(-\beta\mathcal{H}(\boldsymbol{q},\boldsymbol{p}))d\boldsymbol{p}d\boldsymbol{q},$$

where h is Planck's constant and the fraction corresponds to the zero entropy of the ideal gas. $\mathcal{H}(\boldsymbol{q},\boldsymbol{p})$ is the Hamiltonian, representing the overall energy and $\iint d\boldsymbol{p}d\boldsymbol{q}$ stands for integration over all 6N coordinates of the phase space. The free energy of the system corresponding to the canonical ensemble, i.e., at temperature and volume, is,

$$G = -\frac{1}{\beta}\ln(Z).$$

Here, we wish to estimate the free-energy difference between two states A and B, for instance, corresponding to different conformations of a protein, is,

$$\Delta G_{\mathrm{BA}} = G_{\mathrm{B}} - G_{\mathrm{A}} = -\frac{1}{\beta}\ln\left(\frac{Z_{\mathrm{B}}}{Z_{\mathrm{A}}}\right).$$

The Hamiltonian that is used for the free-energy difference calculations is a function of the two conformations' Hamiltonians, $\mathcal{H}_{\mathrm{A}}$ and $\mathcal{H}_{\mathrm{B}}$. The new combined Hamiltonian, $\mathcal{H}_{\mathrm{comb}}$, is dependent on λ, a coupling parameter, such that for one value of λ, $\mathcal{H}_{\mathrm{comb}} = \mathcal{H}_{\mathrm{A}}$, and for another value, $\mathcal{H}_{\mathrm{comb}} = \mathcal{H}_{\mathrm{B}}$.

Given a sufficient sampling of a set of configurations, an estimation of the difference in free energy can be achieved by simply using the probability distribution along λ i.e., counting how often a given value of λ is encountered during the simulation. If λ_{A} and λ_{B} correspond to state A and B, the relative probability of λ_{A} and λ_{B} is derived from the partition functions of the states,

$$\frac{p(\lambda_{\mathrm{B}})}{p(\lambda_{\mathrm{A}})} = \frac{\iiint \exp(-\beta\mathcal{H}(\boldsymbol{q},\boldsymbol{p};\lambda))\,\delta(\lambda - \lambda_{\mathrm{B}})d\boldsymbol{p}d\boldsymbol{q}d\lambda}{\iiint \exp(-\beta\mathcal{H}(\boldsymbol{q},\boldsymbol{p};\lambda))\,\delta(\lambda - \lambda_{\mathrm{A}})d\boldsymbol{p}d\boldsymbol{q}d\lambda} = \frac{Z_{\mathrm{B}}}{Z_{\mathrm{A}}},$$

where δ is the delta function. Thus,

$$\Delta G_{\mathrm{BA}} = -\frac{1}{\beta}\ln\left(\frac{p(\lambda_{\mathrm{B}})}{p(\lambda_{\mathrm{A}})}\right).$$

The obtained λ probability distribution has to be reweighted to an unbiased distribution if it includes any configurational biasing of any sort. In the replica-exchange framework presented in the following chapters, however, the free-energy differences are calculated solely based on the sampling in the unbiased reference replica, where the exchanges are reweighted via the Metropolis criterion that leads to a Boltzmann distribution.

3. Protein-Protein Complex Structure Prediction: Methods and Tools[1]

Almost all processes in the cell dependent on formation of protein complexes. Proteins and other biological molecules form assemblies that have specific functions, such as transcription and translation of genetic information or transport of materials across cell membranes (22,23). Understanding a complex's structure is central to understanding its function. Recent years have witnessed a surge in the number of resolved compound structures owing to powerful molecular-biology instruments and equipment. X-ray crystallography, for instance, has been extremely fruitful in determining the structure of more than five thousand protein compounds at atomic – or sub-atomic–resolutions (24). Another example is Cryogenic electron microscopy (CryoEM), which has been revolutionary for the field of structural biology, especially for larger assemblies. In CryoEM, contrary to X-ray crystallography, it is not necessary to purify and crystallize large amounts of proteins. It only requires a sufficient number of snapshots from different viewpoints, which collectively form a high-resolution image (25). In addition, using Nuclear Magnetic Resonance (NMR) the structures of many complexes–although mostly dimeric– have been elucidated (26,27). NMR is limited to complexes of small size (~20 kD). Nevertheless, it can be used to assist structural modelling and affinity determination of larger complexes given that the partner proteins have known structures.

Despite the great progress in the experimental determination of protein-protein complexes, it is still a challenge to determine all putative protein-protein complexes of the cell (28). The number of proteins in the cell is in the order of thousands or tens of thousands (29,30). However, the number of relevant complexes and assemblies amount to hundreds of thousands (24). Moreover, weak, or transient protein-protein interactions are often too unstable to allow structural

[1] This chapter has been previously published in similar form in: Pourjafar-Dehkordi, Danial, and Martin Zacharias. "Prediction of protein–protein complex structures by docking." *PROTEIN INTERACTIONS: Computational Methods, Analysis and Applications*. 2020. 59-85.

determination at high atomic resolutions. There are experimental approaches to detect such assemblies, however a detailed insight into the complex is missing. A realistic and accurate structure prediction protocol, therefore, is of increasing importance (29,31,32).

The abundance of resolved protein-protein complex structures provides us with a template, using which we can construct a model for complex formation (33). A recent study suggests that the majority of natural protein-protein interactions can be modelled by a template-based approach (34). The hurdle to overcome is to properly map the target protein sequence to the template sequence. Template-based protein-protein complex modelling is a powerful tool that has been reviewed (35) and will not be part of the present chapter. It is noteworthy that even template-based models often require further refinements. This will be further discussed in the final section of the chapter. If, however, the target-template similarity is insufficient or protein partners are unknown, the template-based models fall short of their functionality. Here is where protein-protein docking methods can be extremely useful. Protein-protein docking methods are computational techniques that predict the structure of a complex starting from the structure of its isolated protein monomers (31,36).

Most of protein-protein docking algorithms use a concept known as *optimal complementarity* as the main target criteria for predicting interactions. According to this concept, the partner structures are considered as rigid bodies, whose interactions follow a "lock and key" principle, proposed by E. Fischer (37). The binding affinity and specificity are then determined by the associated changes in the free energy, caused by structural and physiochemical changes in the protein partners. These methods are called the rigid-body protein-protein docking methods (32). The principles of these methods will be described in the first section of the chapter.

Nevertheless, based on frequent experimental observations, partners appear to induce conformational changes upon binding, that are prerequisite for complex formation. In other words, biological macromolecules usually are not rigid and can undergo various types of motion at physiological temperatures. Based on these observations, the induced-fit concept of partner proteins has evolved (37,38). According to this concept, a pre-existing ensemble of interconvertible conformational states are at equilibrium. Within this ensemble there are states close to the bound and unbound forms and all states are accessible, even in the unbound form, albeit with a potentially lower statistical weight. Binding to partner molecules shifts the equilibrium towards the bound form. For a realistic prediction, the protein-protein docking algorithm needs to account for such conformational differences among states. Inclusion of conformational changes either directly or in the form of pre-calculated ensembles in docking approaches will be reviewed in the second section of this chapter.

Biochemical and biophysical experiments supplement docking algorithms by identifying amino-acid residues that interact with partners, or by setting upper and lower bounds for the inclusion of residues in the vicinity of the binding site (39–41). Additionally, the knowledge

gained from bioinformatics–for instance, the conservation of surface residues or co-evolution of residues on partners–augments to the precision of docking methods (28). These will be covered in more details in the third section.

To create an accurate structural model, complex structures obtained from an initial rigid- or semi-rigid-body docking algorithms need to be refined. Indeed, most of current protein-protein docking methods have an initial exhaustive systematic search, from which a subgroup of complexes are selected to go through a secondary refinement step (31,42). In the last section such flexible refinement steps and various techniques for scoring the predictions will be

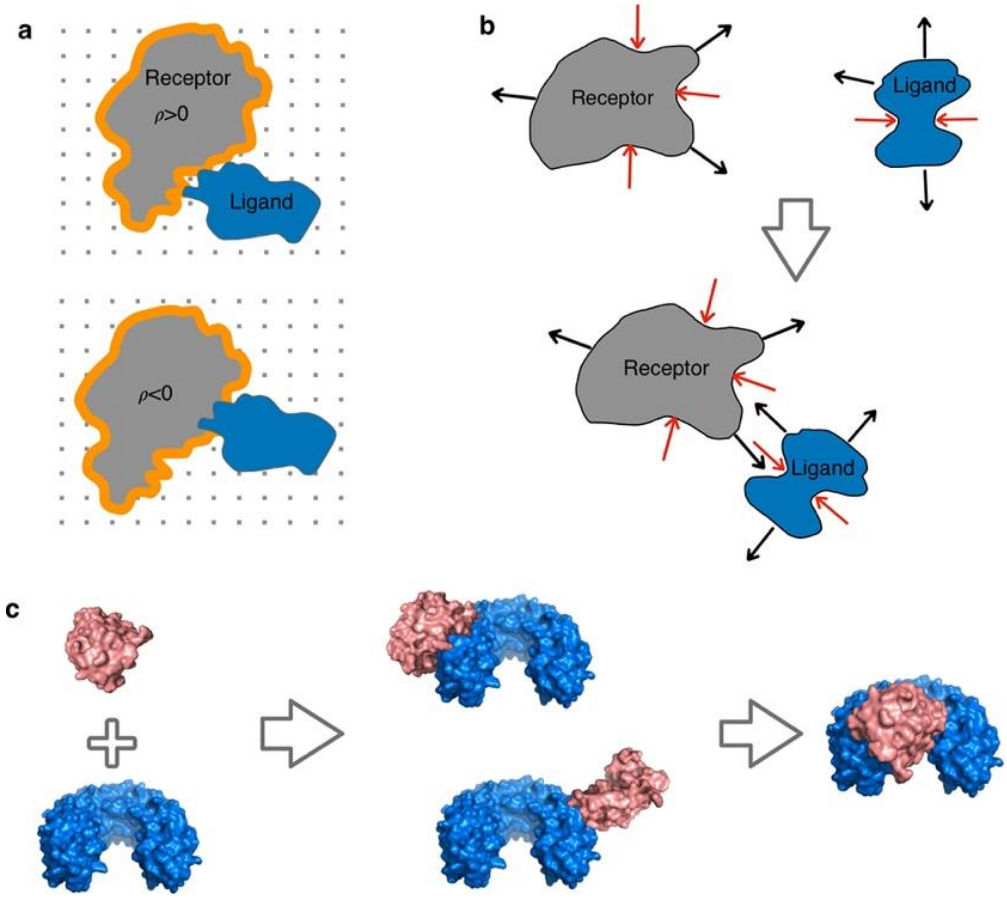

Figure 3-1. The most common types of docking methodologies.
a, docking based on solving a correlation task; the overlap of a ligand protein (blue) with the surface of a receptor protein (grey) is calculated by solving a correlation integral (43). The correlation is positive if the ligand overlaps with the surface region (orange) but becomes unfavourable with increasing overlap with the interior of the receptor. Using a grid to discretize the protein the correlation is rapidly solved using FFT and the best possible overlap for a given relative protein orientation is extracted. The resolution of the FFT grid determines the model's accuracy. **b**, geometric hashing to match surface descriptors on both protein partners (44). Concave (red arrows) and convex areas (black arrows) on the protein's surface are illustrated. Docked structures are assessed based on the level of matching between concave and convex regions. **c**, docking by systematic energy minimization or short MD simulations starting from thousands of different juxtapositions of the partner proteins with the aim to find an energetically stable complex (on the right of the panel).

discussed and possible directions of improvement will be outlined. Finally, in a conclusion section very recent developments to predict protein interaction geometries using brute-force Molecular Dynamics (MD) simulations including full flexibility of binding partners and explicit inclusion of solvent molecules will also be discussed.

3.1. Rigid-body docking approaches

In protein-protein docking using rigid partner structures the aim is to generate a variety of possible interaction geometries, rapidly and exhaustively, with significant complementarity but a minimal overlap of partner structures. The degrees of freedom for each partner with respect to the other are limited to three translational and three rotational. A variety of computational methods have been developed in recent years to efficiently generate numerous putative bound geometries. To implicitly account for conformational adjustment of binding partners some nonspecific sterical overlap between docking partners is typically tolerated. Among the most common ones are fast Fourier transform (FFT) correlation techniques (35,43,45,46) to efficiently locate overlaps between complementary protein surfaces and geometric hashing methods to rapidly match geometric surface descriptors of proteins (Figure 3-1) (42,44). In the FFT-docking approach, the two protein partners are represented by cubic grids; the grid points are assigned discrete values for inside, outside and on the surface of the protein. A geometric complementarity score is calculated for the two binding partners by computing the correlation of the two grids representing each protein. Instead of summing up all the pair products of the grid entries, one can make use of the Fourier correlation theorem. The corresponding correlation integral can easily be computed in Fourier space. The discrete Fourier transform for the receptor grid needs to be calculated only once. Due to the special shifting properties of Fourier transforms the different translations of the ligand grid with respect to the receptor grid can be computed by a simple multiplication in Fourier space. This process is repeated for various relative orientations of the two proteins. Several available computer programs for protein-protein docking use the Cartesian FFT algorithm (Table 3-1).

A disadvantage of standard Cartesian FFT-based correlation methods is that for each relative orientation of one protein with respect to the other, there needs to be a FFT performed. Typically, the orientations vary by 10-15 º (45). The discretisation of the relative orientation can be avoided by correlating spherical polar basis functions that represent, for example, the surface shape of protein molecules (47). More recent approaches allow for solving the whole multidimensional search process–including orientation and translation–in Fourier space, instead of solving only the translation correlation task for discrete relative orientations of the proteins in Cartesian space (48). Using FFT correlation technique it is possible to solve a rigid-

body docking problem for protein partners within minutes on standard PC. It has been successfully applied in the field of protein-protein docking and is part of most protein-protein docking webservers (Table 3-1).

Recently, new multidimensional correlation methods have been developed that allow the correlation of multi-term potentials. Each function needs to be expressed in terms of spherical basis functions. characterizing the surface properties of the protein partners (49). Utilizing such rapid scoring methods, the full partition function of the rigid-body docking problem can be calculated (48). One significant drawback of the FFT correlation method is that it is directly applicable only to protein dimer predictions and requires sequential application, when it comes to molecular assemblies consisting of many protein partners.

In addition to the FFT correlation method, protein-protein binding arrangements are identified using a technique called *geometric hashing* (Figure 3-1). It has been developed originally to match complementary subsections of several datasets. The protein surface is typically represented as a set of triangles that are stored in a hash table. By means of a hash key, matching triangles on the surface of the other partner is quickly identified. These triangles account for points on the molecular surface, having a certain geometrical (concave, convex) and/or physico-chemical (polar, hydrophobic) character. Complex geometries are evaluated based on the level of complementarity between the triangles of the binding partners. PATCHDOCK, for instance, is a program employing geometric hashing (50).

A third group of rigid-body approaches use multi-start energy minimization, Brownian dynamics or Monte Carlo simulations to predict the bound complex (31). A great advantage of these methods is that they can include a range of conformational flexibility already at the initial systematic search step. However, the computational demand is larger compared to FFT-based correlation methods or geometric hashing. Often, the search needs to be limited to predefined regions of the binding partners. Another advantage is their capability to simultaneously dock multiple protein partners. To reduce the computational costs, coarse-grained protein models were employed, which allows to energy-minimize thousands of start configurations (51,52). Available programs that belong to this class are ATTRACT, RosettaDock, SwarmDock and HADDOCK (53–59). Several docking approaches are also available as webservers (Table 3-1).

With the help of the community-wide Critical Assessment of Predicted Interactions (CAPRI) experiment the progress in protein-protein docking prediction methods has been extensively monitored over the last 15 years (60–62). CAPRI is a challenge, in which different groups test docking methods' blind prediction of protein-protein complex structures. For protein partners with small differences between unbound and bound conformation and some experimental hints on the interaction region accurate predictions of complex structures are possible (62,63). However, when protein partners undergo significant conformational changes upon association or for protein homology models the predictions are often incorrect or very limited accuracy.

Computational approaches to realistically predict protein-protein binding geometries need to account for such conformational changes.

Table 3-1. Protein-protein docking programs and associated websites or webservers.

PROGRAM (REF.)	SEARCH METHOD	PROTEIN REPRESENTATION
3D DOCK (64)	Correlation FFT	Discrete
ATTRACT (53)	Guided: multi-minimization	Coarse grain
CLUSPRO (46)	Correlation FFT	Discrete
DOT (65)	Correlation FFT	Discrete
GRAMM-X (66)	Correlation FFT	Discrete
HADDOCK (59)	Guided: data-driven, MD	Atomic
HEX (67)	Correlation polar FFT	Discrete
ICM-DISCO (68)	Guided: MC minimization	Atomic
MOLFIT (69)	Correlation FFT	Discrete
PATCHDOCK (50)	Geometric	Surface
ROSETTADOCK (55)	MC & minimization	Coarse grain to atomic
ZDOCK (45)	Correlation FFT	Discrete
MEGADOCK (70)	Correlation FFT	Discrete
SWARMDOCK (57)	MC on a Swarm	Discrete
F2DOCK (71)	Correlation FFT	Discrete
FRODOCK (72)	Guided: distance restraints	Discrete
PYDOCK (73)	Correlation FFT	Discrete
PROBE (74)	Geometric, MC minimization	Discrete

3.2. Description of protein flexibility during systematic docking search

Protein-protein docking tools tend to perform better when starting from bound structures–those that have been extracted from the known complex–, compared to when they start from structures determined in the absence of the partner. This is due to the conformational changes that components of protein complexes experience upon binding, which might be of local (sidechain or loop transitions) or global (domain motions) nature (Figure 3-2). Unfortunately, the structure of the constituents of a complex are not always resolved; in such cases, one needs

to turn to comparative modelling–using their homologs and a known template with sufficient sequence similarity. The homology-based models are typically of lower accuracy than the experimentally determined structures and can also contain sidechain or loop misplacements that interfere with rigid-body docking process. Indeed, rigid-body docking methods fail to predict the native complex if the bound and unbound structures differ significantly (62,63). Therefore, a refinement step, in which the conformational readjustments are accounted for, is desirable. Several approaches of flexible refinement, ensemble docking and explicit inclusion of flexibility during the entire docking process have been developed to include possible conformational changes during docking (31).

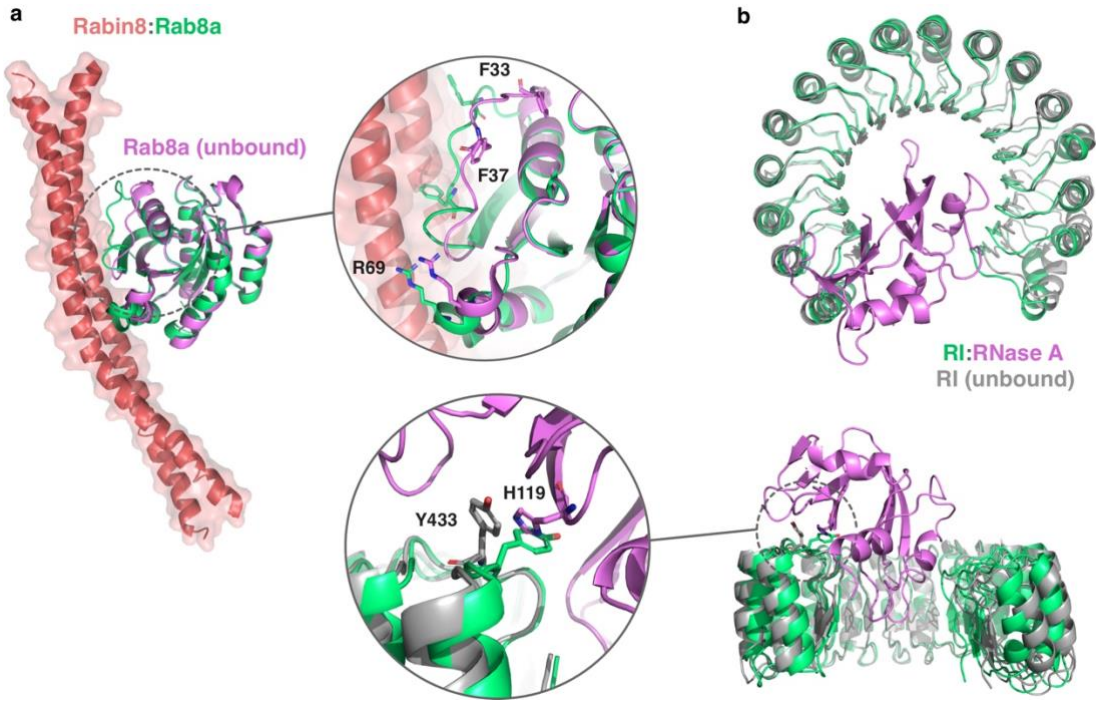

Figure 3-2. Conformational changes upon protein complex formation.
These changes are either local (e.g., sidechain or loop motions) or global. **a**, functional loop regions known as switch I & switch II in the small G-protein Rab8a adopt a closed arrangement in the unbound state (magenta, PDB: 4lhw), but adopt a more open conformation once bound to the Guanine nucleotide exchange factor, Rabin8 (red and green, PDB: 4lhy). **b**, global backbone changes (opening/closing motion) are observed by comparing the cartoon representation of the Ribonuclease Inhibitor (RI) in the unbound (grey, PDB: 2bnh) vs. in complex with Ribonuclease A (RNase A) (green and magenta, PDB: 1dfj). In addition to the global changes also side changes are observed (stick models in the inset of panel b).

3.3. Experimental and bioinformatics data incorporated in docking

Like any other approach, protein-protein docking methods come with limitations, such as false positives, or the diversity of final predictions, especially when there are large proteins involved. Nevertheless, they have improved by integrating low- or high-resolution experimental data to steer the docking engine towards the correct results, or to filter out false predictions at post-processing stages, although it might have added to the workflow's complexity (53,75). The term *integrative modelling* is typically used for methods that combine experimental and bioinformatics data from various sources to generate structural models of molecular assemblies (28,76,77). A variety of experimental data can be used to guide docking of proteins and we will discuss only the most common types of data.

Cross-linking has been used increasingly in the recent years to supervise and validate protein docking data. Mass spectrometry is a common experimental technique used to identify the location of the cross-links, which are included as upper-bound distance restraints during docking, or screening the solutions (78). Mechler et al. used distance restraints between specific residues–obtained from cross-linking experiments–to adjust and score the models proposed by HADDOCK (79). ATTRACT docking software also incorporates such data (80). Small-angle X-ray scattering in solution (SAXS) is another technique for characterizing structural and dynamics of biomolecules at low resolutions (28). SAXS data, calculated as a form of convolution, has served in several docking programs as a scoring function (81). Most of the current methods use SAXS data as a filter to refine and rank the final predictions (82–84), whereas other methods directly use them during the sampling stage (81,85).

Despite the great progress of CryoEM experiments in the recent years, the resolution is in some cases low (>10 Å). Nevertheless, this low-resolution electron-density envelope can be very useful for evaluating the shape of the macromolecular complex. For example, HADDOCK can contain CryoEM data in addition to other sources of experimental and bioinformatics restraints to generate putative models of complexes (86). The generated solutions are then scored based on HADDOCK scoring function. The ATTRACT approach has also the option to perform docking of an arbitrary number of protein molecules guided by low-resolution CryoEM density envelopes (87). Based on this implementation, it was demonstrated that inclusion of low-resolution (15-20 Å) CryoEM data is highly efficient to guide docking to near-native geometries for the majority of complexes in a large benchmark set (88).

3.4. Flexible refinement and final scoring of docked complexes

Pons et al. have investigated the limitations of rigid-body docking strategies in combination with a rescoring step using rigid-body FFT-correlation-based docking and scoring with the PyDock approach (89). This approach employs an electrostatic Coulomb contribution with a surface-area-based solvation term and a van der Waals term. The protocol performed very well for most proteins that undergo minor conformational changes upon complex formation (<1 Å root-mean-square deviation (RMSD) between unbound and bound structures), but unsatisfactory results for cases with significant binding-induced conformational changes or with homology-modelled proteins. Hence, improvement of the prediction accuracy of proposed binding modes is necessary for specific scoring indicating a coupling between realistic scoring and accurate prediction of the complex structure.

Almost all docking approaches employ a refinement and rescoring step applied to a subset of docking solutions–even those methods that include some degree of conformational flexibility during the initial search. The number of structures that undergo a refinement step are typically in the order of hundreds or thousands of the initial solutions and therefore, computational efficiency of the refinement step is critical. The FireDock approach (90) uses a combination of rigid-body moves and sidechain optimization to refine docking solution obtained by PatchDock (50). The refinement steps of the program HADDOCK consist of a series of energy minimizations and dynamics in dihedral variables followed by MD simulation in Cartesian coordinates that allows inclusion of explicit solvent (91). Most FFT-based approaches such as CLUS-PRO or ZDOCK employ energy minimization and sometimes also short MD simulations to remove sterical overlaps and improve the surface complementarity (92,93). The Rosetta molecular modelling suite is frequently used for optimizing the geometry of protein-protein complexes (56,94). The Rosetta program typically uses internal dihedral angles as conformational variables. It can be used for flexible refinement of just the sidechains at the interface, but also in combination with “backrub” motions to modify the backbone geometry (95). The approach is used both for refinement of solutions from FFT-correlation docking and for solutions obtained directly from a coarse-grained Rosetta protein-protein docking approach. Often the local refinement of docked complexes using energy-minimization or MD approaches does not move the partner structures significantly from the starting geometry–it just optimizes the interface interactions. In many cases especially when the unbound protein structures differ from the conformation in the bound complex none of the solutions come very close to the native binding geometry. Often, even the *acceptable* solutions according to the CAPRI criteria are not

reached (Table 3-1). In such cases, not only must the refinement step improve the interface arrangements, but it also must globally move the partners closer. The iATTRACT approach (96) is a docking refinement strategy that tries to combine minimization in the global translational and rotational variables with a full atomic flexibility of only the predicted interface region. In this way small sidechain movements can trigger large-scale whole-body movements of partner proteins (outlined and illustrated in Figure 3-3).

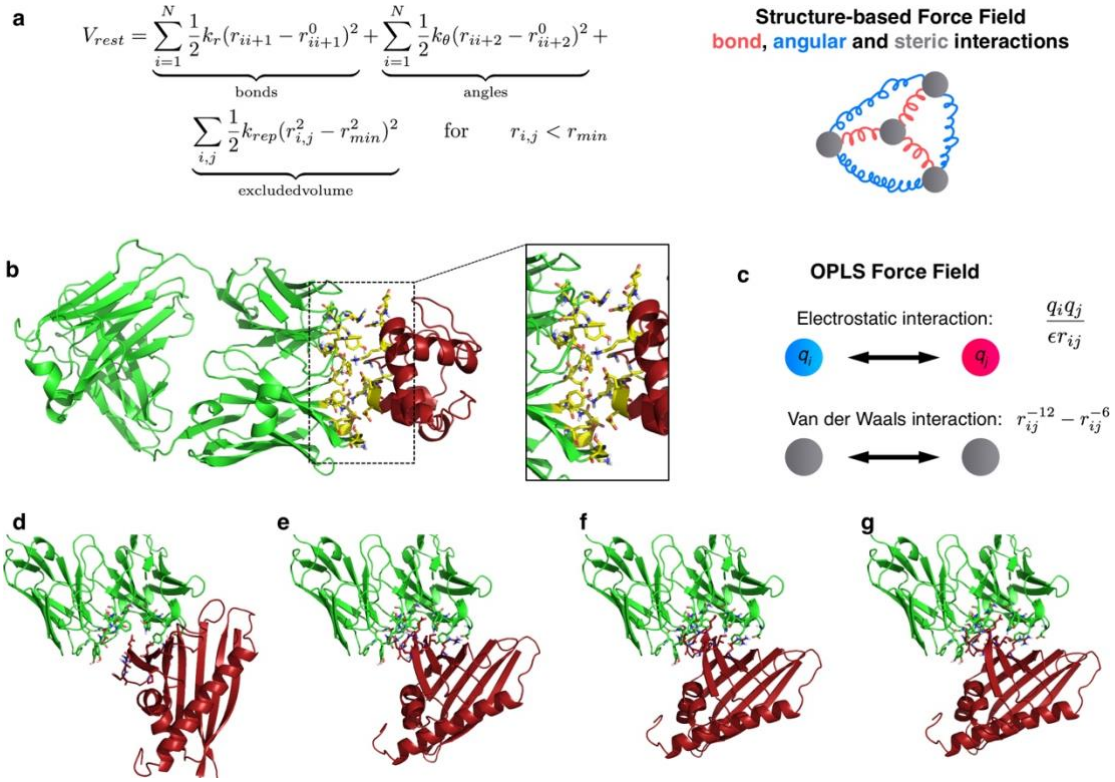

Figure 3-3. Refinement of protein-protein docking geometries by iATTRACT.
In iATTRACT full mobility of the interface residues (within 5-10 Å) is combined with the global mobility of the partners in terms of overall rotation and translation. The intra-molecular interactions of the interface atoms are described by an elastic network model (indicated in **a** and illustrated in **b**, receptor protein in green cartoon, ligand protein in red cartoon, interface as yellow stick model). The intermolecular interactions at the interface (yellow sticks) are described by a standard force field (OPLS (97), indicated in **c**). Small changes at the interface can trigger large-scale rearrangements in rotation and translation of a predicted start complex (**d**) as illustrated in a series of energy-minimization snapshots (**e-g**).

High-performance computers allow longer MD simulations on putative docked complexes for further refinement under realistic conditions–including full flexibility of partners and explicit solvent molecules (98–100). However, even in long MD simulations protein-protein complexes can be trapped for a long time in non-specific "sticky" complexes and do not move forward to reach a near-native state (100). Among several advanced sampling options, replica-exchange simulations (REMD) with an added biasing potential to the partners in the replicas (H-REMD

or BP-REMD approach) is a possibility to avoid such trapping in non-specific geometries and a broader range of conformations can be sampled. The replica exchanges allow also sampling of relevant states in a reference replica without biasing (101,102).

Many docking methodologies incorporate a two state-scoring procedure; an initial scoring followed by a final, more sophisticated evaluation (40). The initial scoring, for instance based on surface complementarity, eliminates unlikely docking solutions and restricts the number of relevant complexes to a few hundred or thousand, for an accurate evaluation. The final scoring employs a physics-based force-field function that includes van zacharder Waals interactions, electrostatic Coulomb interactions and solvation contributions. This is the implemented approach in HADDOCK (37) , Rosetta (34) , pyDock (54) and ATTRACT (82) approaches. In the final scoring, each term of the force fields are assigned a weight which are calibrated based on benchmark sets of known complexes and a large collection of decoy conformations.

Alternatively, knowledge-based docking solutions are also used (36,41,103–106), in which random distributions of residues on the protein surface are evaluated based on the observed probability of finding those residues at protein-protein interfaces. The advantage of this approach is that the effective free-energy function can be deduced by taking the logarithm of the probability ratio into account. Most of the designed statistical potentials are based on the pairwise contacts or distance between residues (41,106). However, the concept can be expanded to multi-body potentials. Recently, new methods for optimizing such potentials have been used to further improve statistical potentials based in part on machine learning applications (107,108).

In most cases, scoring of complexes is performed on single docked structures (e.g. cluster representatives) by just calculating the interaction energy between partner molecules, based on force fields or knowledge-based potentials. However, protein complexes are not static structures. To account for the flexible character of the complexes, molecular mechanics Poisson Boltzmann/generalized Born surface area (MMPBSA/MMGBSA) approaches can be employed (109). In these approaches, an ensemble of docked conformations in the vicinity of the starting complex is generated using MD simulations. This ensemble is analysed using a molecular-mechanics force field–similar to the single structure scoring–combined with either Poisson-Boltzmann or Generalized Born approach to implicitly account for solvation effects. This technique has been extremely successful for scoring protein-protein or protein-peptide complexes (110,111). It is, however, more demanding than scoring based on single structures.

It is important to note that almost all scoring methods are based on the interaction between partners. However, protein binding is driven not only by interactions between partners but also by the binding free energy. The binding free energy includes many parameters in addition to the interaction between partners. For example, it is influenced by the energy of deforming the unbound structure into the bound structure, by the entropic cost of reducing the rotational and

translational freedom of one partner relative to the other and by the changes in the conformational entropy of the partners–usually a restriction of conformational mobility. All these effects can be calculated in free energy simulations of protein-protein binding using restraint MD simulations (112–114). In such calculations one typically restraints the conformations of the partners to stay close to the starting structure in the bound (or predicted) complex and applies restraints to keep the relative orientation of the partners. Subsequently, dissociation of the partners is achieved along a separation distance coordinate by adding an appropriate biasing potential and the associated free energy along the dissociation path is obtained. Finally, simulations are performed at the bound and dissociated states to calculate the free energy of releasing the restraints. Such methods have been used to evaluate single complexes (115) or a set of complexes using a coarse-grained model (116). Moreover, the methodology was systematically tested on 20 test systems and including 50 decoy complex for each test case in combination with an implicit solvent description (117). It was concluded that the performance is indeed slightly better than scoring based on interaction energies, but further developments are necessary to improve accuracy and convergence of the methodology.

Table 3-2. CAPRI protein-protein docking criteria.

QUALITY	% NATIVE CONTACTS	LIGAND-RMSD	INTERFACE RMSD
HIGH	≥ 50	RMSD ≤ 1 Å	or RMSD ≤ 1 Å
MEDIUM	≥ 30	1 Å < RMSD ≤ 5 Å	or 1 Å < RMSD ≤ 2 Å
ACCEPTABLE	≥ 10	5 Å < RMSD ≤ 10 Å	or 2 Å < RMSD ≤ 4 Å
INCORRECT	< 10	–	–

3.5. Conclusions

Protein-protein docking approaches have seen steady progress in recent years that has been monitored regularly by the community-wide docking challenge CAPRI. Docking methods and docking servers are frequently used also by non-expert users working in the field of protein-protein interactions. In addition to prediction by protein-protein docking, complex structures for many of the natural protein-protein interactions can be generated based on template-based modelling methods followed by structural refinement. Even in cases with no appropriate template it is possible in many cases to include experimental information or data from bioinformatics to restrict or guide the search for protein-protein docking searches. Hence, the prediction of dimeric protein-protein interactions is a highly evolved field with many available approaches and many successful applications. However, the prediction of weak interactions that

may form the basis for an assembly of several proteins to form multi-protein complexes is still extremely challenging. These assemblies mediate many cellular functions and may also undergo rapid changes in the cell due to association and dissociation of sub-elements. Experimental data such as low-resolution density from CryoEM, Cryo-tomography or in vivo crosslinking data could be combined with docking methods to obtain structural models of larger molecular assemblies in the cell. These approaches may allow in the not-too-distant future the structural modelling of the assembly and disassembly of many multi-protein complexes even in a crowed cell-type environment.

4. Covalent Modification of Small GTPase Rab8a Impedes Binding to The Exchange Factor Rabin8[1]

GTPases are key players in cellular signaling processes. Post-translational modification of GTPases can modulate their function and signaling properties. Phosphorylation of Rab proteins, which belong to the Ras superfamily of small GTPases that regulate intracellular transport, has recently been implicated in the pathogenesis of Parkinson Disease (PD). For Rab8a, it was shown that the phosphorylation of residue serine 111 (pS111) is dependent on the protein kinase PINK1, and that mimicking the phosphorylation at S111 by a serine/glutamate substitution (S111E) impaired Rab8a activation by its cognate guanine nucleotide exchange factor (GEF) Rabin8. Here, we performed comparative molecular dynamics and free energy simulations on Rab8a and Rab8a:Rabin8 complexes to elucidate the molecular details on how pS111 and S111E may influence the interaction with Rabin8. The simulations indicate that S111E and pS111 establish an intramolecular interaction with the neighboring arginine residue 79 (R79). The interaction persists in the complex and perturbs a favorable intermolecular salt-bridge contact between R79 in Rab8a and aspartate 187 in Rabin8. Binding free energy analysis reveals that S111E and pS111, as well as the R79A mutation, drastically decrease the binding affinity for Rabin8. Combining the R79A mutation with S111E or pS111 nearly diminishes Rab8a−Rabin8 binding. In vitro experiments confirm our computational results showing a > 80% decrease in the nucleotide exchange rate of the respective Rab8a mutants in the presence of Rabin8 compared to that of the wild type. In addition to insights into how S111 phosphorylation of Rab8a influences GEF-mediated activation, the simulations demonstrate

[1] This chapter has been previously published in similar form in: Pourjafar-Dehkordi, Danial, et al. "Phosphorylation of Ser111 in Rab8a modulates Rabin8-dependent activation by perturbation of side chain interaction networks." *Biochemistry*. 2019, 58, 33, 3546–3554. Reprinted with permission from the American Chemical Society.

how sidechain modifications in general can allosterically influence the surface sidechain interaction network between binding partners.

4.1. Introduction

The Rab subfamily of small GTPases is involved in the spatial and temporal regulation of vesicular trafficking (118–121). The subfamily consists of ~60 Rab proteins in humans with specific intracellular localization to mediate signaling functions (122). The basis for mediating signaling processes is a molecular switch between an inactive guanosine diphosphate (GDP)-bound state and an active guanosine triphosphate (GTP)-bound state. To switch from the inactive to the active state, the binding of guanine exchange factors (GEFs) and activation by GEFs are required (118). GEFs are enzymes that stimulate the release of GDP and binding of free GTP to Rab proteins. Additional partner proteins, termed effectors, can recognize the active GTP form and promote cellular downstream processes. GTPase activating proteins (GAPs) eventually return Rabs back to their inactive state by stimulating the intrinsic GTPase activity (123,124). The Rab activity state is communicated to regulatory proteins and downstream interaction partners by two functionally important loop regions known as switch I and II. These regions are conformationally flexible in the inactive state, but they become structurally ordered in the active form. Due to their pivotal role in binding to interaction partners, any changes in the conformation of these regions are going to influence the binding profile of the GTPase.

The activity of Rab GTPases can be further modulated by post-translational modifications (PTMs) such as phosphorylation (125–127). For example, it has been found that Parkinson's disease (PD) kinase LRRK2 regulates a subset of Rab GTPases (128). Another example is the PTEN-induced kinase 1 (PINK1) that is a protein kinase which is indirectly involved in the phosphorylation of a conserved Ser111 residue in the Rab8a, Rab8b, and Rab13 GTPases *in vivo* (125). PINK1 is important for mitochondrial quality control, and mutations in PINK1 are associated with autosomal recessive Parkinson's disease (125). Recently, it was found that mimicking the phosphorylation of Ser111 by introducing a Ser111 → Glu ($S111E^{Rab8a}$) substitution significantly impairs Rab8a activation by its cognate GEF, Rabin8 (125).

The crystal structure of GDP-bound Rab8a in complex with Rabin8 provided molecular insights into the mode of this Rab-GEF interaction (120). Interestingly, in this three-dimensional structure, the $S111^{Rab8a}$ is not directly interacting with Rabin8 (i.e., is not part of the interface). Hence, the decrease on the rate of Rab8a activation by Rabin8 due to $S111^{Rab8a}$ phosphorylation cannot be readily explained by a PTM-induced obstruction of the protein-protein interface. However, $S111^{Rab8a}$ is located opposite a negative surface patch of the Rabin8 (125). Therefore,

the repulsion of charges between the phosphorylation-mimicking S111E^{Rab8a} mutation or the

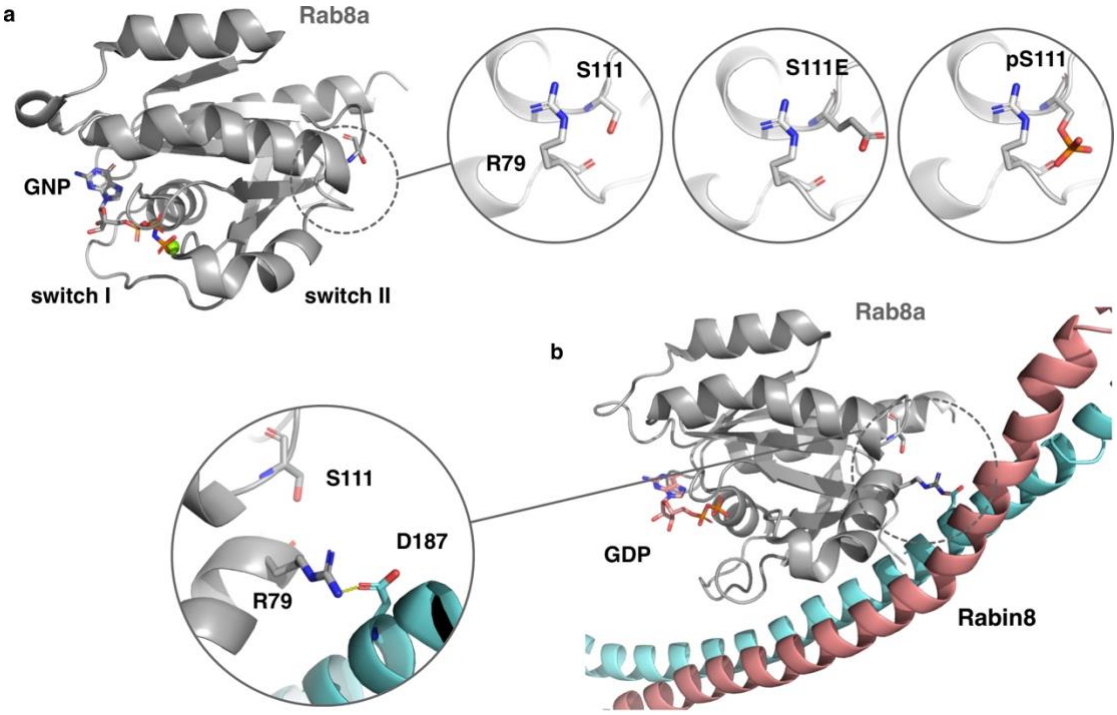

Figure 4-1. The start structures of the simulations.
a, cartoon representation of the crystal structure of Rab8a bound to GNP–the GTP analogue–which served as the starting structure for Rab8a simulations (PDB: 4lhw). GNP is shown in atom-colour coded sticks and Mg^{2+} in green sphere. Start structure of the Rab8a variants were generated by replacing S111^{Rab8a} with glutamic acid (S111E^{Rab8a}) or phosphoserine (pS111^{Rab8a}) *in silico*. **b**, cartoon representation of the crystal structure of the Rab8a:Rabin8 complex and location of bound GDP and of residues S111^{Rab8a}, R79^{Rab8a} and D187^{Rabin8} as sticks. The insets illustrate the contact of R79^{Rab8a} and D187^{Rabin8} and the negative patch (red) of Rabin8's electrostatic surface around residue D187^{Rabin8}, respectively.

Ser111^{Rab8a} phosphorylation and the negative surface patch of Rabin8 may provide a molecular cause for the less efficient complex formation.

In the present study we investigate the Rab8a:Rabin8 complex using molecular dynamics (MD) simulations to elucidate the molecular mechanism by which the Ser111^{Rab8a} phosphorylation impairs the interaction with Rabin8. We compare the Rabin8-binding of wild type Rab8a with the corresponding S111E^{Rab8a} mutant and pS111^{Rab8a} modified form. In the complex, D187^{Rabin8} interacts with the switch II residue R79^{Rab8a} to form a favorable salt-bridge interaction stabilizing the complex. However, in case of the Ser111 phosphorylation or S111E substitution, we identified intramolecular sidechain interactions in Rab8a between S111E^{Rab8a}/pS111^{Rab8a} and R79^{Rab8a}. This interaction weakens or even disrupts the interaction with D187^{Rabin8} in the complex with Rabin8. The simulations demonstrate that R79^{Rab8a} plays a key role in mediating polar interactions between Rab8a and Rabin8 that can be perturbed by introducing the S111E^{Rab8a} mutation or upon S111^{Rab8a} phosphorylation. The simulation results could be

confirmed by experimental in vitro measurements showing that the Rabin8-mediated nucleotide exchange rate of Rab8a variants ($S111E^{Rab8a}$, $R79A^{Rab8a}$) is decreased by >80% compared to wild type Rab8a.

The study gives insights into the molecular mechanism of signaling modulation by phosphorylation of Rab8a. It furthermore is a model system on how modifications of polar and charged residues adjacent to (but not localized in) a protein-protein interface can allosterically modulate binding strength.

4.2. Results and Discussions

In a previous study it has been found that the phosphorylation-mimicking $S111E^{Rab8a}$ substitution (and possibly also $S111^{Rab8a}$ phosphorylation) in Rab8a impairs the activation by its cognate GEF, Rabin8 (125). To investigate the molecular origin of this effect, we performed a series of MD simulations of isolated Rab8a variants and in complex with Rabin8, starting from the known structures (Figure 4-1). The substitutions were performed *in silico* to yield Rab8a $S111E^{Rab8a}$ and $pS111^{Rab8a}$ variants, both in isolated Rab8a and in the complex with Rabin8 (see Methods). As a first step, we performed simulations on the isolated Rab8a variants

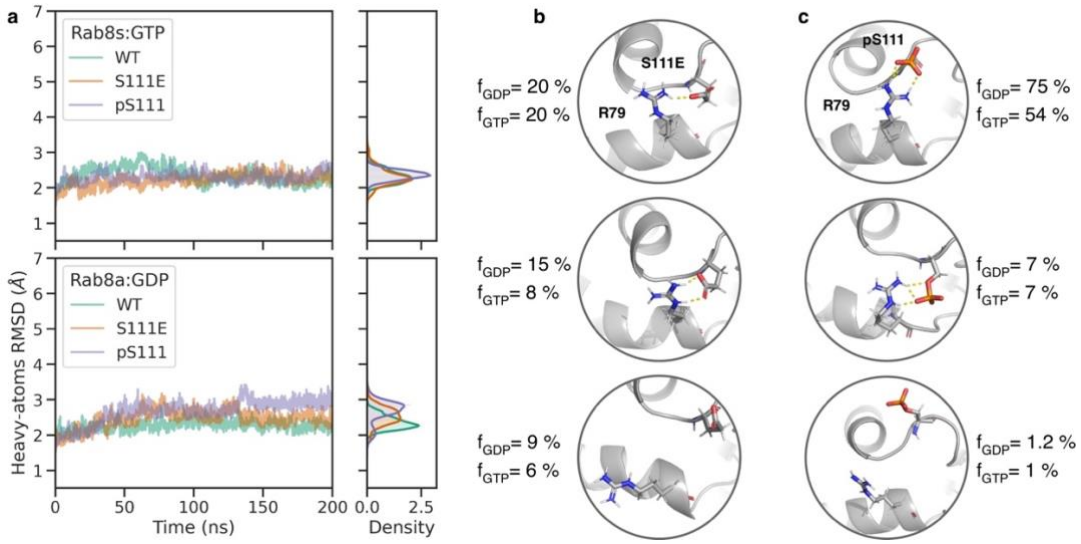

Figure 4-2. Results from the simulation of the isolated Rab8a in complex with GTP and GDP. **a**, root-mean- square deviation (RMSD) of protein's heavy atoms with respect to the initial structure wild type (WT), S111E- and pS111-Rab8a variants bound to GTP (upper left panel) and GDP (lower left panel). The RMSD probability density distributions are also indicated. **b**, representative conformations of the three most populated clusters obtained from cluster analysis of the trajectories along with their population for GDP- bound S111E-Rab8a (**b**) and pS111-Rab8a (**c**) cases. The numbers next to snapshots report the population of the clusters and their equivalent clusters in the GTP-bound state.

(bound to GTP or GDP) starting from the experimental structure in complex with a bound GTP analog (PDB: 4lhw) (120). On a simulation time scale of 200 ns, the sampled states remained close to the start structure with a similar overall root-mean-square-deviation (RMSD) relative to the start structure of all variants (Figure 4-2a). On this time scale, fluctuations of the switch I and II regions (for the GDP-bound cases) but no unfolding processes were observed. However, both the phosphorylation mimetic $S111E^{Rab8a}$ as well as the $pS111^{Rab8a}$ variant formed transient hydrogen bonding states with the sidechain of $R79^{Rab8a}$ in the neighboring switch II region (Figure 4-2). Especially $pS111^{Rab8a}$ formed a robust salt-bridge contact to $R79^{Raba8}$ with two H-bonds for 75 % of the simulation time. Only in less than 10 % ($S111E^{Rab8a}$) and 2 % ($pS111^{Raba8}$) of the simulation time, there was no contact with $R79^{Rab8a}$. The sampling of these intramolecular H-binding contacts was observed in both the simulations with GTP- and GDP-bound Rab8a. In contrast, there were no stable contacts between $S111^{Rab8a}$ and $R79^{Rab8a}$ observed in the MD simulations of wild type Rab8a.

4.2.1 Simulation of Rab8a in complex with Rabin8

As a next step, MD simulations of the Rab8a variants in complex with Rabin8 were performed starting from the known crystal structure (GDP-bound, PDB: 4lhy, residue substitution at position S111Rab8a by *in silico* mutation). For all the variants, as indicated by the RMSD plots and the trajectories snapshots in Figure 4-3, the protein interface remained unchanged, showing that the mutations did not alter the overall structure. However, in the case of wild type Rab8a in complex with Rabin8, the $R79^{Rab8a}$ can form a hydrogen-bonded salt-bridge contact to D187 of Rabin8 ($D187^{Rabin8}$). This hydrogen-bonded state was also observed as the dominant local conformational cluster during the simulations of the wild type Rab8a in complex with Rabin8 suggesting that the contact contributes favorably to the stability of the Rab8a-Rabin8 complex (Figure 4-4). It is characterized by a short distance between sidechain groups of $R79^{Rab8a}$ and $D187^{Rabin8}$ and a larger distance of $R79^{Rab8a}$ and $S111^{Rab8a}$. Indeed, several conformational clusters for the arrangement of the $D187^{Rabin8}$, $R79^{Rab8a}$ and $S111^{Rab8a}$ sidechains could be distinguished for the wild-type case during the simulations, mostly with direct intermolecular contacts between $R79^{Rab8a}$ and $D187^{Rabin8}$ (Figure 4-4).

In both cases of $pS111^{Rab8a}$ modification and $S111E^{Rab8a}$ substitution, the simulations indicate that the intramolecular contact to $R79^{Rab8a}$ persists in the simulations in complex with Rabin8

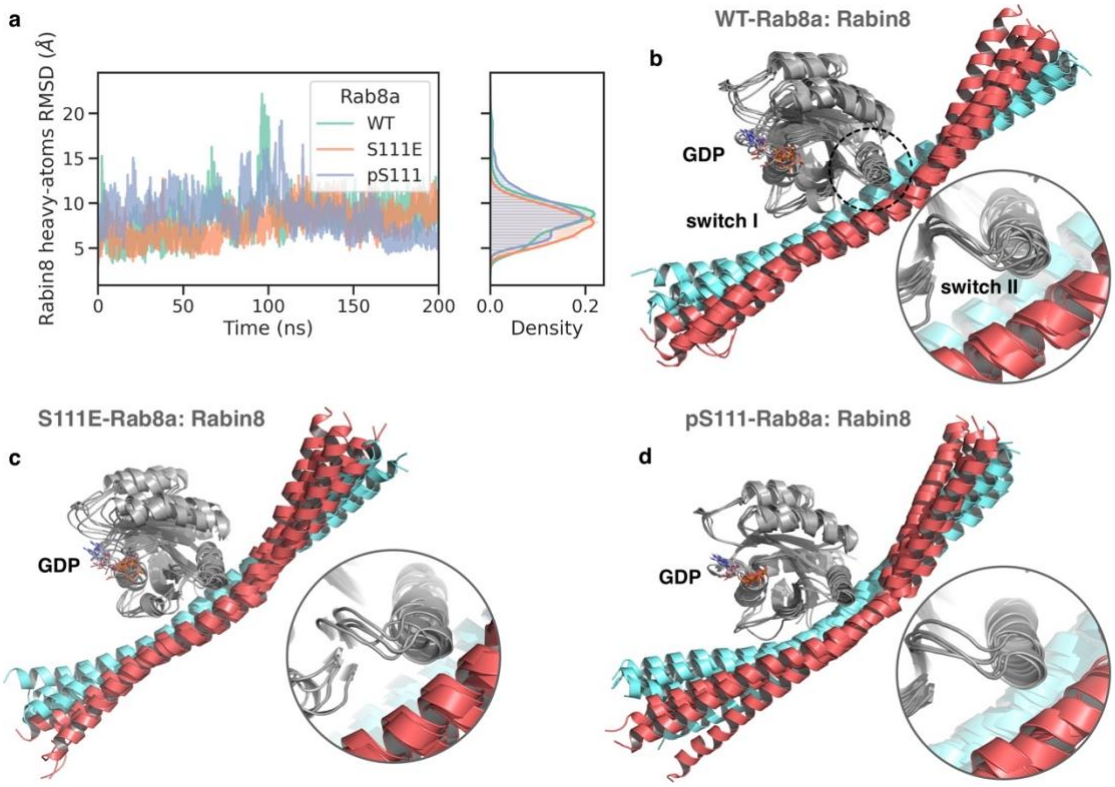

Figure 4-3. MD simulation of GDP-bound Rab8a variants in complex with Rabin8.
a, heavy atoms RMSD of Rabin8 with respect to the initial structure after superimposing the complex on the Rab8a partner protein vs. simulation time. The RMSD probability density is also illustrated. Superposition of five snapshots (cartoon representation) obtained from the simulations of WT (**b**), S111E (**c**) and pS111 (**d**) variants of Rab8a in complex with Rabin8 (snapshots taken every 50 ns).

(Figure 4-4). The average distance of residue S111^{Rab8a} to R79^{Rab8a} is considerably shorter than that of the wild type (WT complex, 7.8 Å; S111E-Rab8a complex, 5.1 Å; pS111-Rab8a complex, 4.1 Å). In the most populated states, the R79^{Rab8a}-D187^{Rabin8} contact is disrupted or shows only a weak H-bonding geometry and some sampled states allow for simultaneous contacts of all three residues, such that the average R79^{Rab8a}-D187^{Rabin8} distance is similar in all simulations of the complexes (WT complex, 5.4 Å; S111E-Rab8a complex, 5.2 Å; pS111-Rab8a complex, 5.3 Å). The cluster analysis of the arrangement of the three residues in the simulations of the S111E^{Rab8a} and pS111^{Rab8a} variants indicated sampling of fewer distinct clusters compared to the wild type (Figure 4-4). Interestingly, due to the formation of the intramolecular contact between residue S111^{Rab8a} and R79^{Rab8a}, the average distance of residue S111^{Rab8a} relative to D187^{Rabin8} is also reduced in the simulations of the pS111^{Rab8a} and S111E^{Rab8a} variants (WT complex, 9.2 Å; S111E-Rab8a complex, 8.0 Å; pS111-Rab8a complex, 7.1 Å). It brings the negative charges of these residues closer together, and thus is expected to weaken the binding. Finally, control simulations of the double substitution S111E-R79A-Rab8a or pS111-R79A-Rab8a in complex with Rabin8 indicate no stable arrangement with close

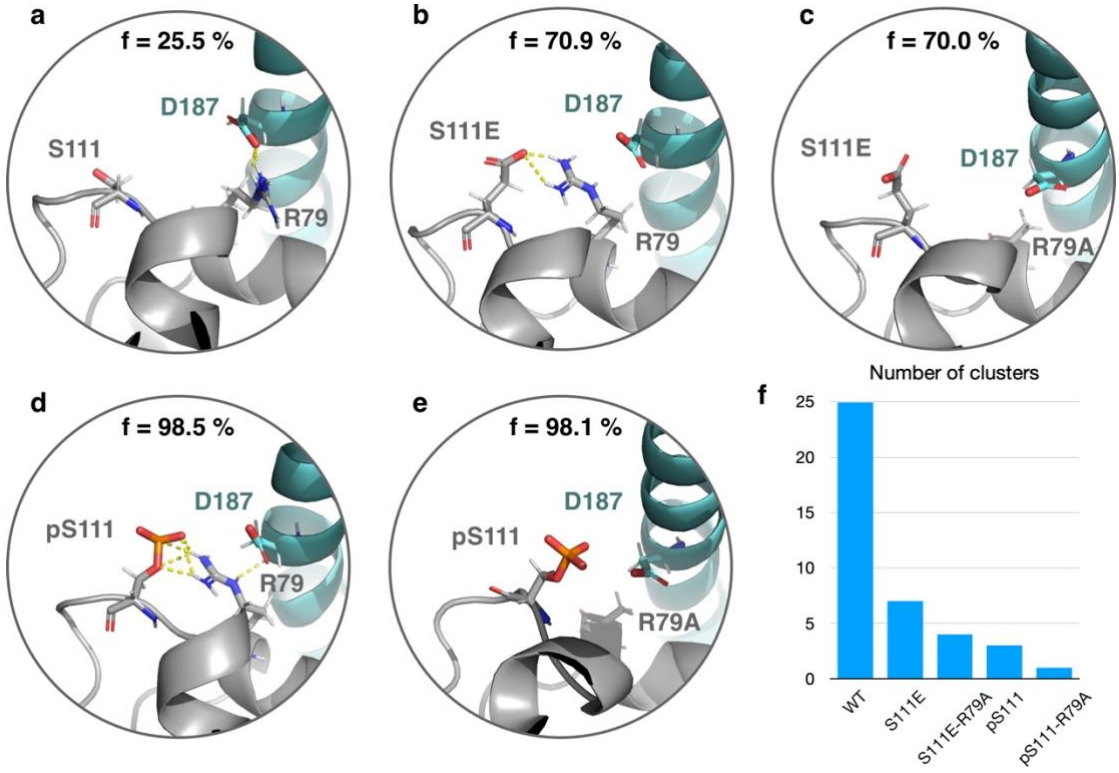

Figure 4-4. Dominant conformational states observed in simulations of Rab8a:Rabin8 complexes.

Dominant conformational states observed in simulations of Rab8a:Rabin8 complexes. **a**, representative snapshot of the most populated cluster obtained from 200 ns MD simulations of wild-type Rab8a in complex with Rabin8. **b**, same as **a**, but for S111E-Rab8a in complex with Rabin8. **c**, the most populated cluster representative for the MD simulations of the S111E-R79A-Rab8a:Rabin8 complex. **d**, representative snapshot of the pS111-Rab8a:Rabin8 complex. **e**, same as in **d** but for the pS111-R79A-Rab8a:Rabin8 complex. Hydrogen bonds are illustrated with yellow dashes. The numbers on the lower right in each panel indicate the population of the clusters as percentage of the frames.

contacts of the sidechains $R79^{Rab8a}$, $S111^{Rab8a}$ and $D187^{Rabin8}$ (snapshots of most dominant states are illustrated in Figure 4-4).

4.2.2 Influence of S111-phosphorylation on binding affinity to Rabin8

To quantify the effect of the $S111E^{Rab8a}$ and $pS111^{Rab8a}$ substitution in Rab8a on the binding to Rabin8, we evaluated the trajectories using the molecular mechanics-Poisson-Boltzmann surface area (MMPBSA) method (129,130). It should be emphasized that these calculations do not include all contributions to binding (they neglect for instance conformational entropy contributions) but allow a semi-quantitative comparison of relative binding affinity of the variants. The assumption is that the reduction in conformational entropy upon complex formation is similar for all variants. The MMPBSA calculations indicate only a small binding

affinity difference in favor of Rab8a:GDP vs. Rab8a:GTP to Rabin8, which is in line with the experimental evidence that there is no strong preference for Rabin8 GEF interactions with the GDP vs. GTP bound forms of Rab8a (120). Both, the substitution $S111E^{Rab8a}$ as well as the $pS111^{Rab8a}$ modification resulted in a calculated reduction of the binding strength (by about ~9-10 kcal/mol). Apparently, this gives an estimate of the energetic contribution of the favorable $R79^{Rab8a}$-$D187^{Rabin8}$ contact that is frequently sampled in the case of the WT-Rab8a:Rabin8 complex, but either disrupted or weakened by the presence of the nearby negatively charged residues $S111E^{Rab8a}$ or $pS111^{Rab8a}$ in the Rab8a variants. It is important to emphasize that the calculated magnitude of this binding energy reduction is likely an overestimation because the entropic contributions to restrict the conformational freedom of the sidechain motion upon binding are not included. Since our MD simulations indicate that the $R79^{Rab8a}$ residue may play a key role in mediating the effect of the $S111^{Rab8a}$ modification, which does not directly contact the Rabin8 partner, we used the "alanine scan" option in the MMPBSA approach to investigate the effect of substituting the $R79^{Rab8a}$ by alanine ($R79A^{Rab8a}$). Already in case of the wild type in complex with Rabin8, the $R79A^{Rab8a}$ is predicted to significantly reduce the Rab8a:Rabin8 binding affinity (Table 4-1). In case of the $S111E^{Rab8a}$ and $pS111^{Rab8a}$ variants of Rab8a, a further reduction in affinity was observed. With the loss of contact between $R79^{Rab8a}$ and $D187^{Rabin8}$ by replacing the arginine with alanine ($R79A^{Rab8a}$), an attractive force between the two proteins is eliminated, and by addition of the phospho-mimetic S111E substitution (S111E-R79A-Rab8a), a repulsive electrostatic force between the two negatively charged residues ($S111E^{Rab8a}$ and $D187^{Rabin8}$) is predicted to weaken the complex affinity even more. This is reflected by the sum of Coulomb and polar-solvation contributions that represent the electrostatic contribution to binding. This contribution is more positive (by 5-9 kcal/mol) for all the $R79A^{Rab8a}$ variants compared to the cases with no $R79^{Rab8a}$ mutation (Table 4-1).

In addition to the average effect of the substitutions we also split the trajectories into sets of frames that belong to different conformational clusters formed by the residues $R79^{Rab8a}$, $S111^{Rab8a}$ and $D187^{Rabin8}$ (Figure 4-5). The MMPBSA analysis of these sets of frames indicates that each cluster of sidechain arrangements contributes differently to the binding affinity. Some clusters make a favorable and others a less favorable or even unfavorable contribution to binding. For example, in case of the wild type simulations at around 100 ns mostly conformations with a $R79^{Rab8a}$-$D187^{Rabin8}$ distance > 4 Å are sampled resulting in an overall slightly positive interaction energy ($\Delta E > 0$) compared to states with a smaller $R79^{Rab8a}$-$D187^{Rabin8}$ distance. The $S111E^{Rab8a}$ substitution or phosphorylation results in a shift of the clusters to more states with an unfavorable effect on binding.

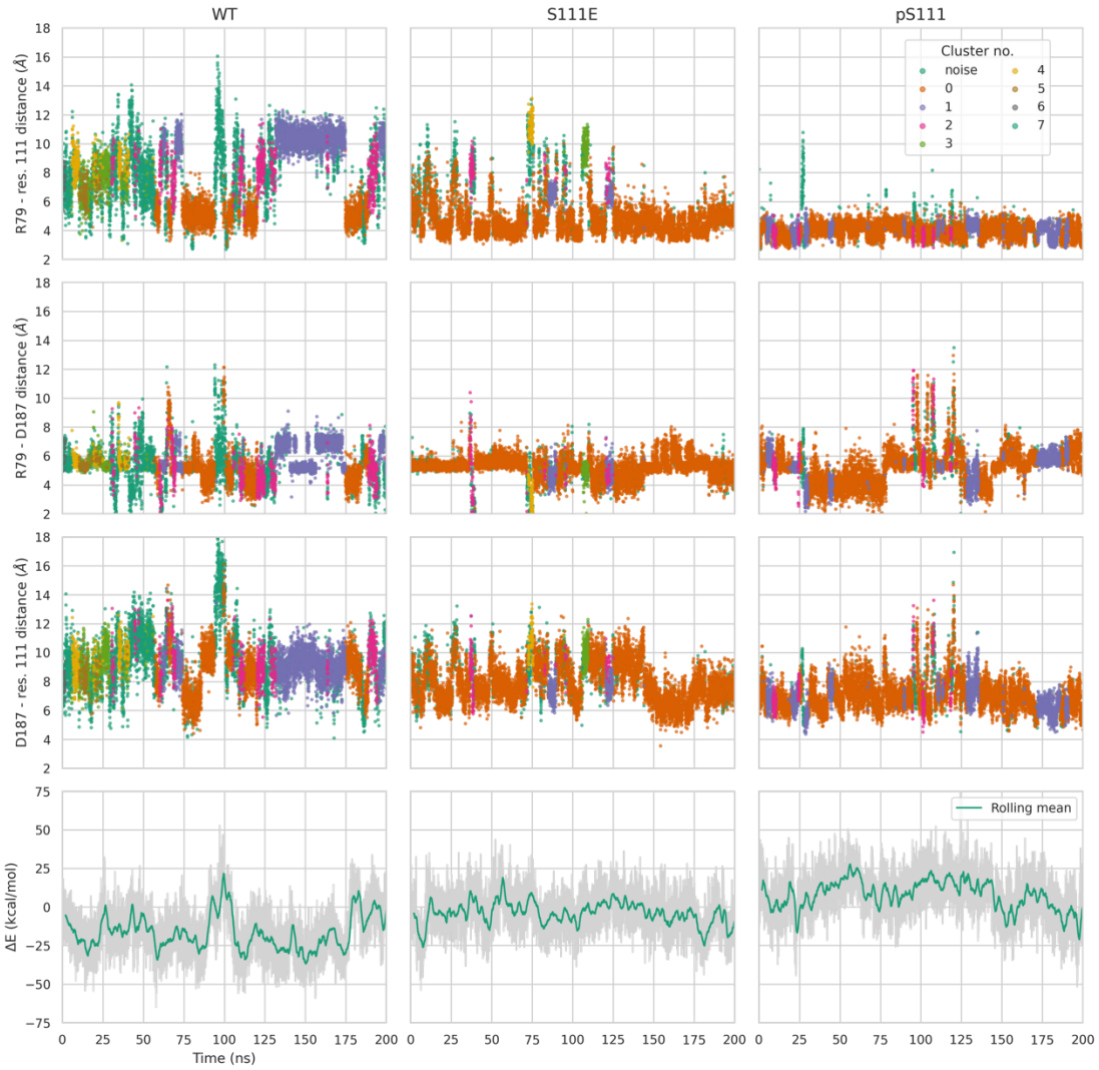

Figure 4-5. Sampling of conformational substates and interatomic distances involving residues R79^{Rab8a}, S111E^{Rab8a}, pS111^{Rab8a} and D187^{Rabin8}.

In the top panels the assignments to most populated clusters are indicated by different point colours. The clusters are numbered based on their population in a descending order. The intramolecular distances between R79^{Rab8a}-D187^{Rabin8} and residue S111^{Rab8a} are calculated between the arginine amino group (NH_2) or OG of aspartic acid and the sidechain OG of S111^{Rab8a}, CG of S111E^{Rab8a} and OG of pS111^{Rab8a}, respectively. The intermolecular distances between R79^{Rab8a} and D187^{Rabin8} are the hydrogen-bond distances of the arginine amino group and OG of D187^{Rabin8}. Rab8a–Rabin8 interaction energies are calculated using the MMPBSA approach. The green lines in energy plots indicate the mean over a rolling 2 ns window.

The interaction energy (indicated as ΔE) is more positive compared to the wild type simulations especially in time intervals that correspond to conformations with increased R79^{Rab8a}-D187^{Rabin8} distances (the lower panel in Figure 4-5). The arrangement forms a model system how modifications of residues not being part of an interface can mediate or control binding affinity

by changing the network of contacts of the nearby residues (in the present case, of $R79^{Rab8a}$) with residues on the partner protein.

The simulations suggest differences in the binding affinity of the Rab8a variants to Rabin8, which should result in decreased nucleotide exchange rates for those variants compared to wild type Rab8a. Except for the $pS111^{Rab8a}$ variant, it was possible to generate, express and purify all Rab8a variants. We determined the kinetics of Rabin8-stimulated nucleotide exchange reactions for WT-Rab8a, S111E-Rab8a, R79A-Rab8a and S111E-R79A-Rab8a. The Rabin8-simulated GDP → GTP exchange rates were calculated for each Rab8a variant based on the time-dependent change of intrinsic tryptophan fluorescence, as reported in ref. (131). We found that the phospho-mimetic $S111E^{Rab8a}$ and the $R79A^{Rab8a}$ mutants resulted in > 80 % reduction of the nucleotide exchange rate compared to wild type Rab8a. In the case of the S111E-$R79A^{Rab8a}$ double mutant, the nucleotide exchange activity was almost completely abolished (~ 95 % decrease). Since the mutations are not on the Rabin8 side, the decreased activity must be due to the weakened binding of Rabin8 to Rab8a variants. Hence, the experimental results are in line with the computational prediction that the presence of $S111E^{Rab8a}$ or $pS111^{Rab8a}$ prevents $R79^{Rab8a}$ to engage in interactions with Rabin8, and therefore decreases the binding affinity. Note that the simulations clearly indicate qualitatively similar effects of $S111E^{Rab8a}$ and $pS111^{Rab8a}$ that in the latter case need to be considered as a prediction because experiments were performed for only the phospho-mimetic $S111E^{Rab8a}$ and the $R79A^{Rab8a}$ mutants.

4.3. Conclusions

Post-translational modifications may play a role in modulating signaling properties of GTPases (125,132). Indeed, for Rab8a it has been previously shown that mimicking the phosphorylation at serine 111 ($S111E^{Rab8a}$) reduces the GEF-mediated activation by Rabin8 (125). Interestingly, the structure determination of the Rab8a:Rabin8 complex revealed that $S111^{Rab8a}$ is not directly located at the interface to Rabin8, and that even upon phosphorylation no direct contact to Rabin8 is sterically possible (120,125). In the study presented here, we investigated the molecular details on how $S111E^{Rab8a}$ and $pS111^{Rab8a}$ affect the interaction with Rabin8 using a series of MD-simulations and free energy calculations. Both the *in silico* $S111E^{Rab8a}$ substitution as well as the $pS111^{Rab8a}$ modification tend to form intramolecular salt bridge-like contacts to the nearby $R79^{Rab8a}$ residue within the switch II region. This was observed in the isolated Rab8a variant and in the complex with Rabin8 as dominant conformational states that in case of the complex perturbs and even disrupts the $R79^{Rab8a}$-$D187^{Rabin8}$ salt-bridge contact in contrast to the wild-type complex. The importance of $R79^{Rab8a}$ for stabilizing the binding of Rab8a to Rabin8

could be demonstrated by analyzing the R79A^{Rab8a} substitution as well as double mutations (R79A^{Rab8a}-S111E^{Rab8a}, R79A^{Rab8a}-pS111^{Rab8a}). Although the predicted effects on the change in binding free energy due to the substitutions are likely to be larger than the experimental binding free energy changes, the order of the effect of each variant correlates very well with experimental data on the binding to Rabin8 and the GEF efficiency of Rabin8. *In vitro* experiments investigating the Rabin8-mediated nucleotide exchange reactions confirmed the computational results. We show that the exchange rates of all Rab8a (S111E, R79A) variants are decreased by >80% compared to wild type Rab8a, suggesting that the residue R79^{Rab8a} in Rab8a plays an essential role in mediating the binding of Rabin8, and that S111E^{Rab8a} interferes with this function.

It is well known that phosphorylation of protein residues can alter conformational equilibria, and in turn influence binding to protein partners. Examples are the phosphorylation of Arg/Ser rich proteins (RS-proteins) resulting in conformational changes and disruption of binding to RNA molecules (133), or the phosphorylation of Tyr residues in Ras-GTPases that alter the switch I and II conformation directly affecting the interaction with effectors (134,135). The present study demonstrates a potential mechanism on how a chemical modification of a residue, which is not part of the binding interface, can modulate signaling events due to altering the neighboring sidechain interaction network that is part of the interface with the partner protein. Such mechanism can be of particular importance when it comes to the fine-tuning of cell signaling events, where a complete disruption of the binding, and thus signaling, would be detrimental. The perturbation or alteration of interacting sidechain networks can potentially be the basis of allosteric effects mediated not by chemical modification of sidechains, but by binding of an allosteric effector adjacent to the interface with another binding partner. Indeed, for HLA-DR (MHC class II) molecules the conformational change of a Trp sidechain induced by binding of the co-chaperone HLA-DM (at a site not overlapping with the peptide binding groove) has been found to control peptide binding and exchange (136). Similar to the present case the allosteric effect is then mediated by perturbation of a sidechain interaction network that mediates the interaction with the binding partner.

4.4. Methods

4.4.1 Simulation protocol

All isolated Rab8a simulations started from the crystal structure bound to phosphoaminophosphonic acid guanylate ester (GNP) (PDB: 4lhw) (120). The Rabin8-bound simulations started from the crystal structure of Rab8a:Rabin8 complex (PDB: 4lhy) (120). The nitrogen atom between the β- and ɣ-phosphate in GNP was exchanged with an oxygen atom to

model the Rab8a:GTP structure. The terminal phosphate was removed to form an initial model of the Rab8a:GDP complex. In the model of the phosphorylated complex, the Ser111 ($S111^{Rab8a}$) was replaced with phosphoserine ($Ser111{:}pS111^{Rab8a}$) to form the phosphorylated Rab8 structures. The Amber ff14sb force field was used for the proteins, additional force field parameters for GDP, GTP and phosphoserine were taken from the Amber parameter database at their fully unprotonated states (137–139). A study by Mann et al. (140) supports the unprotonated state of GTP as the most populated state at neutral pH. For Rab8a:GDP simulations, a Mg^{2+} ion with two bound water molecules was placed next to GDP-phosphate. Using sodium and chloride ions, the salt concentration was adjusted to 0.1 M and the systems were solvated with the TIP3P water model (141,142). The solvated systems were equilibrated by a first energy minimization (5000 steps), followed by 25 ps of heating and 50 ps of density equilibration, followed by a simulation in NPT ensemble at 300 K. During these equilibration phases, all protein nucleotide heavy atoms as well as magnesium ions were restraint with a harmonic potential at force constant of 5.0 kcal $mol^{-1}Å^{-2}$. Data gathering production simulations were performed without any restraints. The pmemd version of the Amber 16 software package (143) in combination with hydrogen mass repartitioning (144) was used which allows a simulation time step of 4 fs. Long range interactions were included using the particle mesh Ewald (PME) method combined with periodic boundary conditions and a 9 Å cut-off for real space non-bonded interactions. Trajectories were processed and analyzed using CPPTRAJ program (143). The DBSCAN algorithm was used for clustering of the trajectories with a distance cutoff of 1.0 Å of heavy atoms root mean square deviation (RMSD) and a frame interval of 200ps. Figures were generated using PyMol software package (145).

4.4.2 Binding affinity calculations

The interaction energies between Rab8a and Rabin8 were calculated using MMPBSA tool of the AMBER software suite (129). The binding free energy of an aqueous complex of two bound proteins can be approximated as

$$\Delta G_{\text{binding}} \approx \Delta E_{\text{MM}} + \Delta G_{\text{solvation}} - T\Delta S$$

$$\Delta E_{\text{MM}} = \Delta E_{\text{vdW}} + \Delta E_{\text{Coulomb}} + \Delta E_{\text{polar solvation}}$$

$$\Delta G_{\text{solvation}} = \Delta E_{\text{cavity}} + \Delta E_{\text{dispersion solvent}}$$

where ΔE_{MM}, $\Delta G_{\text{binding}}$ or $\Delta G_{\text{solvation}}$ and $-T\Delta S$ represent the gas-phase molecular mechanical energy change, the solvation free energy change and the conformational entropy

change upon binding, respectively (146). In our calculations, the entropy term was neglected. Five production simulations of 2 ns in duration started from the complex of GDP-bound Rab8a and Rabin8 (PDB: 4lhy) at 300K, generating snapshots every 10 ps. A similar set of simulations were carried out starting from the same structure but with $S111E^{Rab8a}$ mutated. were carried out starting from the same structure but with $S111E^{Rab8a}$ mutated. Using the "alanine scan" feature of MMPBSA, the contribution of $R79^{Rab8a}$ in each case was evaluated. Calculations were carried out on a sum of 1000 frames for each complex at an ion concentration of 0.1M.

Table 4-1. Calculated binding free energies (kilocalories per mole) of Rabin8 in complex with Rab8a variants.

Mean contributions to binding	Rabin8:Rab8a$_{GDP}$	Rabin8:R79A-Rab8a$_{GDP}$	Rabin8:S111E-Rab8a$_{GDP}$	Rabin8:S111E-R79A-Rab8a$_{GDP}$	Rabin8:pS111-Rab8a$_{GDP}$	Rabin8:pS111-R79A-Rab8a$_{GDP}$	Rabin8:Rab8a$_{GTP}$
$\Delta E_{\text{VAN DER WAALS}}$	-112.9	-111.0	-110.5	-109.2	-121.6	-118.9	-116.9
$\Delta E_{\text{COULOMB}}$	-648.7	-480.1	-460.8	-288.9	-407.2	-245.7	-535.9
$\Delta E_{\text{POLAR\_SOLVATION}}$	666.3	506.1	488.0	324.8	441.6	285.1	560.6
ΔE_{CAVITY}	-87.5	-83.8	-85.9	-82.7	-95.6	-92.7	-89.7
$\Delta E_{\text{DISPERSION\_SOLVENT}}$	171.3	165.9	169.2	164.6	181.2	177.0	173.3
$\Delta G_{\text{BINDING}}$	-11.6 ± 0.3	-2.9 ± 0.3	-0.01 ± 0.3	8.5 ± 0.3	-1.8 ± 0.4	4.8 ± 0.4	-8.7 ± 0.4
$\Delta\Delta G_{\text{BINDING}}$	-	8.6 ± 0.4	11.5 ± 0.5	20.0 ± 0.5	9.8 ± 0.5	16.4 ± 0.47	2.8 ± 0.5

5. Conformational Switching in Small GTPases Upon Binding To GTP/GDP[1]

Rab GTPases constitute the largest branch of the Ras protein superfamily that regulate intracellular membrane trafficking. Its signaling activity is mediated by the transition between an active GTP-bound state and an inactive GDP-bound state. In the inactive state the switch I and II segments adopt largely disordered flexible conformations whereas in the active state these regions form defined conformations. The switch I and II segments are central for recognition of Rab GTPases by interacting partners. Phosphorylation of the Rab1b-GTPase at residue Ser111 (pS111) results in modulation of signaling activity due to alterations of the protein interaction interface but possibly also due to modulation of the conformational flexibility. We have studied the flexibility of native and pS111-Rab1b in complex with GTP or GDP using extensive molecular dynamics (MD) simulations and an advanced sampling DIhedral Angle-biasing potential Replica-Exchange Molecular dynamics (DIA-REMD) method. The DIA-REMD technique promotes backbone and sidechain dihedral transitions along a series of replica simulations in selected segments of the protein segments and through exchanges also improves sampling in an unbiased reference simulation. Application to the Rab1b system results in significantly enhanced sampling of different switch I/II conformational states in the GDP-bound Rab1b. The pS111 modification is found to reduce the conformational flexibility even in the presence of GDP, which may influence signaling activities. The stabilizing effect can be attributed to the formation of additional surface salt bridges between arginine residues and pS111 that are not present in the native structure. The DIA-REMD technique could be a valuable approach for studying also other signaling proteins that contain flexible segments.

[1] This chapter has been previously published in similar form in: Pourjafar-Dehkordi, Danial, and Martin Zacharias. "Influence of a Ser111-Phosphorylation on Rab1b GTPase conformational dynamics studied by advanced sampling simulations." *Proteins.* 2021; 1-9. Reprinted with permission from John Wiley & Sons.

5.1. Introduction

Rab proteins belong to the class of small GPTases and are key actors in a variety of intracellular trafficking events in eukaryotic cells. Central elements of Rab GTPases are molecular switches that can alternate between a guanosine triphosphate (GTP)-bound active state and a guanosine diphosphate (GDP)-bound inactive state. Activation is mediated by guanine nucleotide exchange factors (GEFs), that catalyze the release of GDP and exchange of GDP by GTP. Hydrolysis of the GTP into GDP and switching back to the inactive state can be facilitated through binding to GTPase activating proteins (GAPs) and activation of the intrinsic GPTase activity (123,147). The conformational changes associated with activation or inactivation of GTPases occur in characteristic switch regions I and II. These flexible segments can undergo transitions between structured (in the active state with bound GTP) and largely unfolded inactive states in the presence of bound GDP. Most signaling interactions of Rab GTPases are associated with the switch I and II segments. Specific post-translational modification (PTM) of amino acid residues can further modulate the activity and regulation of small GTPases (120,127,148). Such modifications include adenylylation, phosphorylation and phosphocholination of residues and reside either in the spatial vicinity or are directly located in the two conformational switch regions (127,148–150).

Growing evidence suggests that there are links between regulation of Rabs and Parkinson disease (PD)- related proteins (125,128,151,152). PTEN-induced kinase 1 (PINK1), for instance, is a serine/threonine kinase that functions as a mitochondrial damage sensor, whose mutations cause autosomal recessive PD (153). It has been shown that once activated, PINK1 regulates a number of Rab GTPases, namely Rab1b, 8a, 8b, and 13, through an indirect phosphorylation of a highly conserved residue, Serine 111 (125). This modification has been proven to diminish Rab8a's binding to its cognate GEF (152) by interfering with the network of surface sidechain interactions (131). The determination of the X-ray structure of Rab8a with phosphorylated Serine 111 (pS111) indicated only little difference compared to the native non-phosphorylated GTPase structures. These include in the case of the pS111 variant small adjustments in the backbone structure of switch II with bound GTP and a better resolved residual switch II structure in the presence of GDP (152). Also, for the closely related Rab1b, Nuclear Magnetic Resonance (NMR) spectroscopy indicated no major structural alterations in solution due to the pS111 modification. However, these studies could not resolve the behavior of the switch I and and switch II regions of the Rab1b GTPase. Thermal melting experiments revealed an overall structural stabilization of the pS111-Rab1b both with bound GDP and in the presence of the GTP-analog GppNHp (152). Hence, the influence of the pS111 modification on the conformational flexibility of Rab1b is still not understood. The

increased stability of the pS111 variant indicates that the pS111 modification potentially affects the ensemble of conformational states of the switch regions not detected by the structural studies and that it can play a role for the interaction with other signaling proteins. To complement the biochemical and structural studies we employ comparative MD simulations to investigate the effects of this PTM on the conformational dynamics of Rab1b both in complex with GDP and GTP.

MD simulations have played already a substantial role in gaining a deeper understanding of GTPases. For example, the catalytic cleavage of GTP at high resolution was studied by a combination of X-ray, Fourier-transform infrared spectroscopy (FTIR) and combined quantum mechanics and molecular mechanics (QM/MM) in atomic detail by Gerwert and coworkers (154). Several computational studies have addressed the transition from active to inactive states and have illustrated the interplay of interactions between the protein's backbone atoms, the nucleotide's phosphate atoms and the Mg^{2+} ion along the transition process (155,156). Furthermore, accelerated MD simulations indicate that the mechanism by which the small GTPases bind to the nucleotide is based on conformational selection due to sampling of multiple conformations regardless of the nucleotide bound. Previous free energy simulations on the adenylylation of a Rab GTPase predicted a stabilization of the GTPase active form even in the presence of GDP which was later also confirmed experimentally (157).

In the present study we investigate the influence of the pS111 modification on the Rab1b dynamics using extensive comparative MD simulations of the phosphorylated and non-phosphorylated S111 variants with and without bound GDP or GTP. During conventional MD (cMD) simulations only moderate conformational fluctuations are observed in the switch regions even in the presence of bound GDP. To improve sampling of relevant states, we also employ an advanced sampling technique DIA-REMD, introduced previously, that promotes conformational transitions of specific protein regions during an MD simulation study (158,159). In the DIA-REMD technique, the low-energy backbone and sidechain dihedral angles are penalized with a biasing potential that promotes transitions to conformations separated by energy barriers. Different levels of the biasing potential are applied along a series of parallel running replicas. At preset intervals exchanges between the replicas are attempted and accepted by a Metropolis criterion. Our study demonstrates that the procedure indeed significantly enhances the sampling of relevant switch I and II conformations also in the reference replica without a biasing potential (resulting in correct canonical sampling in the original force field). In addition, the simulations indicate an overall stabilizing effect of the pS111 modification on the Rab1b active conformational state in the presence of GTP but also GDP.

The stabilization is mainly mediated by an interaction of the phosphate group with the R79 residue located in the switch II segment (and another Arg residue) and provides an explanation

for the experimentally observed increased stability of the pS111 variant of Rab1b. The study also demonstrates that the application of the DIA-REMD technique could be helpful to study other flexible protein or peptide segments that may play a role in signal transduction processes.

5.2. Results

5.2.1 Comparative molecular dynamics simulation of GTP/GDP-bound Rab1b.

It is well established that the structural flexibility of GTPases can be strongly modulated by bound GTP or GDP nucleotides especially in the switch I and II regions. We performed unrestrained explicit solvent MD simulations on the Rab1b GTPase with bound GTP and GDP starting from the same structure (PDB ID: 3nkv, Figure 5-1) that represents the active GTP-bound form. The two simulations were also performed with the S111-phosphorylated variant of Rab1b (pS111-Rab1b). On the timescale of 2 μs in all cases, the overall structures of Rab1b and the arrangements of its switch regions remained overall close to the start structure (Figure 5-1). No rearrangement or dissociation of the bound nucleotides was observed. Even the switch I and II regions resulted in only modest fluctuation and variation from the start structure. Slightly larger deviations of the switch I and II segments were observed with bound GDP vs.

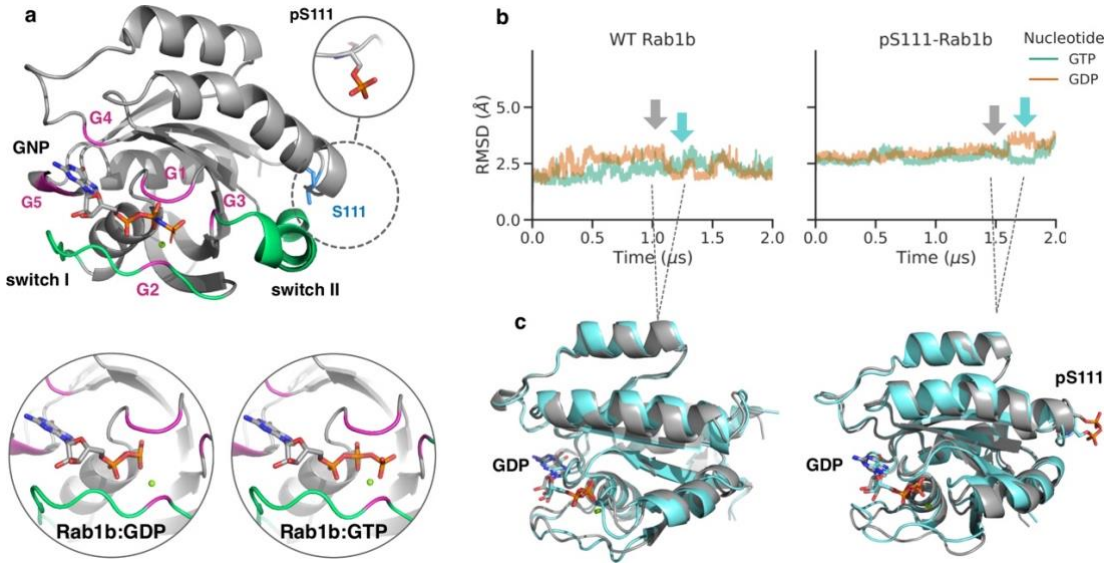

Figure 5-1. Start structure of the Rab1b variants free MD simulations.
a, crystal structure of the *active* Rab1b in complex with a GTP analog (PDB: 3nkv), which served as the start structure. The switch regions are highlighted in green, the conserved G-motifs in magenta and serine 111 in blue. The GNP atoms are represented as sticks and the Mg^{2+} ion as a green sphere. **b**, root-mean-square deviation (RMSD, of all non-hydrogen atoms) during the 2 μs continuous MD simulations. Snapshots taken form the final parts of the simulations of the GDP-bound Rab1b variants (wild type left, pS111-Rab1b right) taken at different RMSD values indicate small conformational changes but no sign of unfolding in the switch regions.

GTP but no unfolding of the segments. The deviations in the switch regions seen in the simulations are significantly smaller compared to experimentally observed conformational differences of the switch regions in inactive (GDP-bound) vs. active (GTP-bound) Rab1b (150,152). It indicates that due to the presence of energy barriers, the timescale of 2 μs of the present MD simulations might be insufficient to sample the expected structural transitions observed experimentally for the active and inactive forms.

5.2.2 DIA-REMD simulations to enhance sampling of transitions in switching regions

Next, we performed DIA-REMD simulations of GTP- and GDP-bound Rab1b. In this technique a number of replica simulations with increasing levels of a specific biasing potential are employed to promote transitions of dihedral angles. The biasing potential encourages conformational transitions of the switch segments especially in the higher replicas. Conformations can then exchange to the reference replica upon frequent exchanges between replica runs. All analysis of the trajectories in the following sections were performed only on the data obtained from the reference replica.

The simulation of the wild type Rab1b in the presence of bound GTP indicates minor deviation of the switch I and II regions (< 4 Å) during the entire 200 ns of DIA-REMD, similar to the results obtained with MD simulations (Figure 5-2). In contrast, in the GDP-bound Rab1b variant, after 50-100 ns significant conformational transitions were observed in both switch I and II regions that reach 6-15 Å, beyond what was observed in the MD simulations. During the first 50 ns, the biasing potential-adjustment algorithm updates the potential levels to optimize the exchange acceptance (more details in Methods). Therefore, only the last 2/3 of the trajectories were used for further analysis. The overall RMSD distribution remained unchanged throughout the last 2/3 of the simulation. Cluster analysis of the trajectories based on switch I and II conformations in Figure 5-4 reveals the most populated states adopted during the simulations. While the free MD frames, regardless of the nucleotide bound, were clustered in only one batch (for the selected clustering threshold of 2 Å root mean square deviation (RMSD)), the DIA-REMD trajectories yielded entirely different results depending on the type of nucleotide. For the GTP-bound complexes, there is only one dominant state found, which is characterized by a well-ordered arrangement of the switches, similar to the start structure. However, the GDP-bound Rab1b adopted a total of 41 states, 18 of which represent at least 1 % of the simulation time. The superimposition of the representative frames in Figure 5-4 clearly indicates the disordered switch regions as well as the localization of the deformations in these two segments.

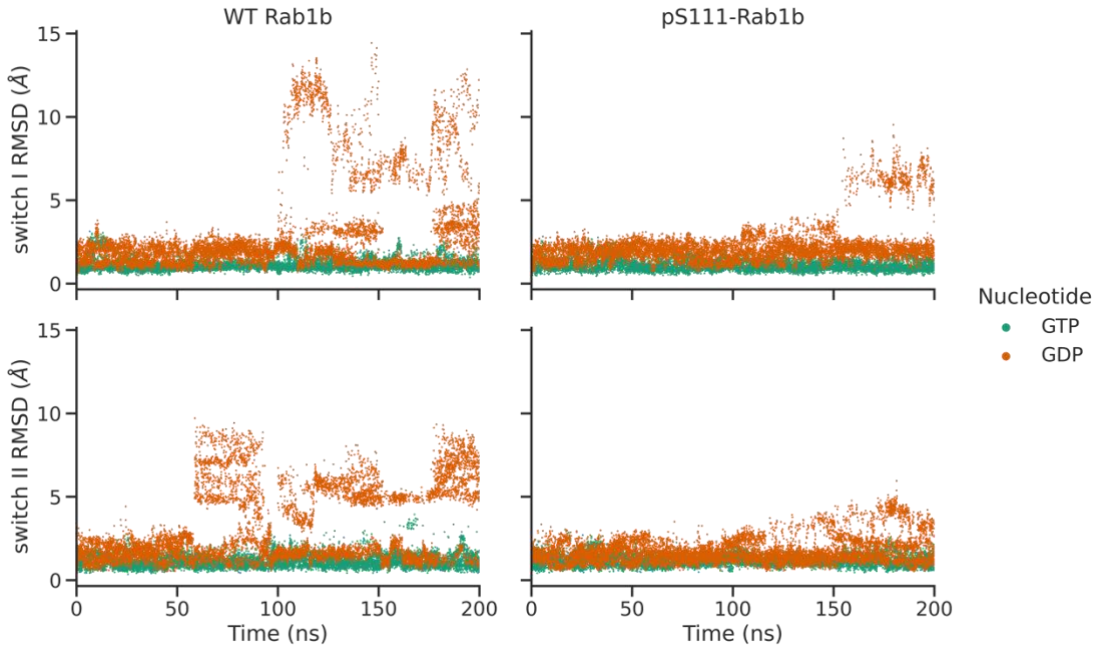

Figure 5-2. DIA-REMD results from wild type (left) and S111-phosphorylated (right) variants. RMSD (non-hydrogen atoms) of the switch I (upper panel) and switch II (lower panel) segments with respect to the native Rab1b conformation in the active state during the DIA-REMD simulations (reference replica). Amino acid residues involved in the RMSD calculation of switch I and switch II are residues 31-43 and 68-79, respectively. Data points are taken in 20 ps intervals.

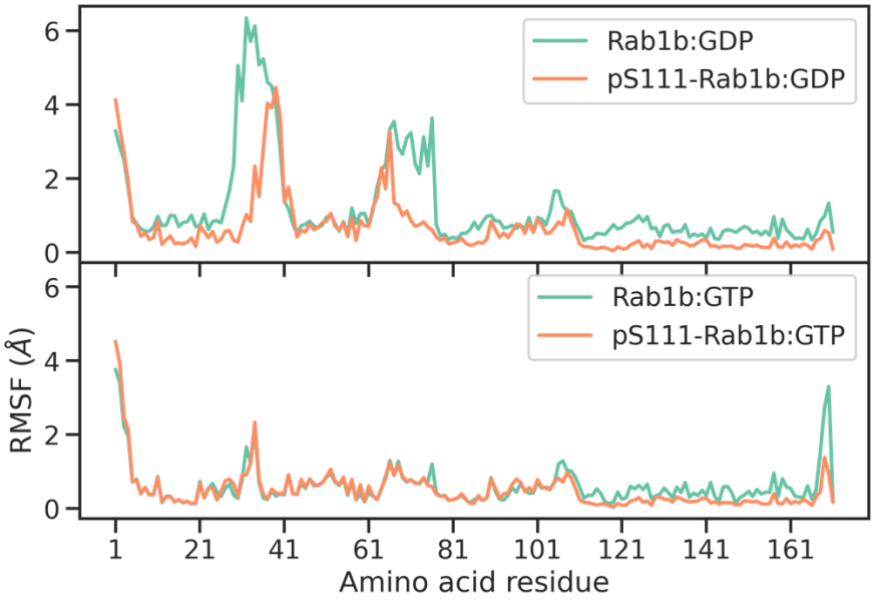

Figure 5-3. Per-residue RMSF of the Rab1b variants during DIA-REMD simulations. The S111-phosphorylation of the GDP-bound Rab1b resulted in stabilization of the protein in switch regions, while it did not cause any significant change in the GTP-bound variant fluctuations.

5.2.3 Structural changes in S111-phosphorylated Rab1b.

In the next step, we studied the structural changes caused by the phosphorylation of Rab1b on S111 employing exactly the same DIA-REMD simulations. Similar to the wild type Rab1b the RMSD values of the pS111-Rab1b:GTP variant remained close to the start structure (Figure 5-2). The DIA-REMD simulation of the pS111-Rab1b:GDP complex indicated enhanced deviations and fluctuations of the switch I and II regions compared to the GTP-bound form. However, the RMSDs of the sampled switch I and switch II in the pS111-Rab1b:GDP remained below 9 Å and 6 Å respectively. Moreover, per-residue root-mean-square fluctuations (RMSF) plots indicated a reduced flexibility in these two regions compared to the wild type (see Figure 5-3). Hence, the simulations indicate that S111-phosphorylation has a stabilizing influence on the switch I and II conformations compared to the wild type Rab1b. A cluster analysis based on the sampled switch I and II conformations indicates fewer clusters (only five) and smaller conformational deviation from the start structure observed within those clusters (Figure 5-4). Interestingly, the cluster analysis of the pS111-Rab1b:GDP complex revealed new hydrogen bonds involving the pS111 mainly formed between pS111 and the neighboring arginines (see Figure 5-5). In particular R79 is located in the switch II segment. In the (by far) most populated conformational cluster the phosphate group of pS111 forms hydrogen bonds both to R79 and R172. The additional H-bond attractions and interactions of the helix dipole with pS111 may stabilize the switch II region and in turn also the switch I segment, resulting in a decreased population of unfolded switch I and II even in the presence of bound GDP.

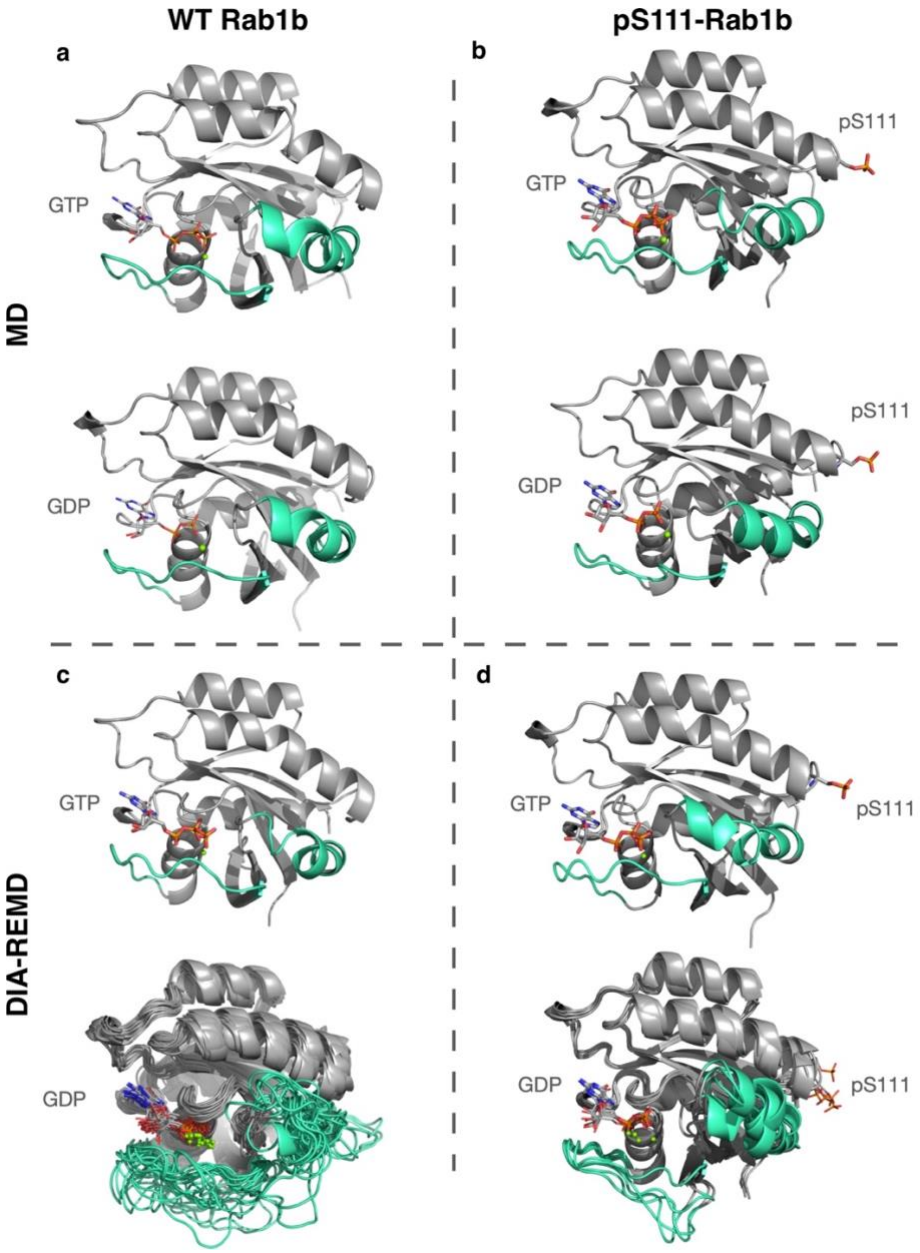

Figure 5-4. Cluster analysis of the simulation trajectories.
The representatives of all clusters obtained from unrestrained MD (**a**, wild type Rab1b, **b**, pS111-Rab1b) and DIA-REMD (**c**, Rab1b, **d**, pS111- Rab1b) for all Rab1b variants. Clustering was performed based on the pairwise RMSD of the switch I (residues 31-43) and switch II (residues 68-79) regions (cyan). The rest of the proteins are in grey cartoon, nucleotides are represented in sticks and the Mg^{2+} ion as green sphere.

Figure 5-6 shows the probability distribution of the switch I & II RMSDs based on a kernel density estimator. The plots were obtained from the DIA-REMD simulations of wild type and S111- phosphorylated Rab1b in the presence of GTP and GDP. In the wild type variants and in the presence of GDP, the RMSD distribution covers a broad range, indicating a significant deviation from the conformation found in the active GTP-bound state. In contrast, for the GTP-

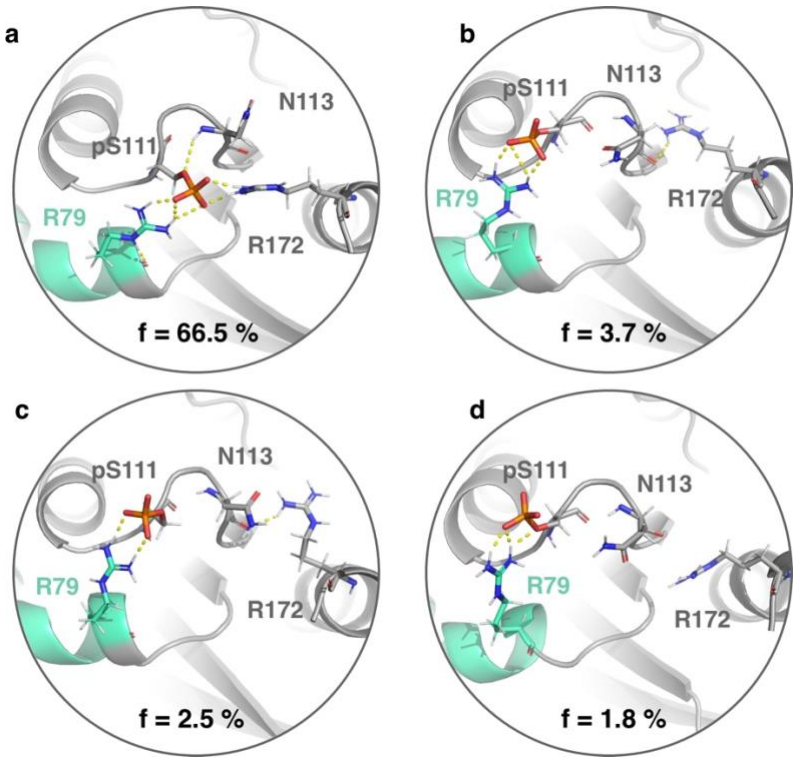

Figure 5-5. Representative conformations of the most populated clusters obtained from the simulations of the pS111-Rab1b:GDP complex.
In each case (a-d) the percentage of the frames represented by the cluster is also indicated. The yellow dashed lines represent hydrogen bonds.

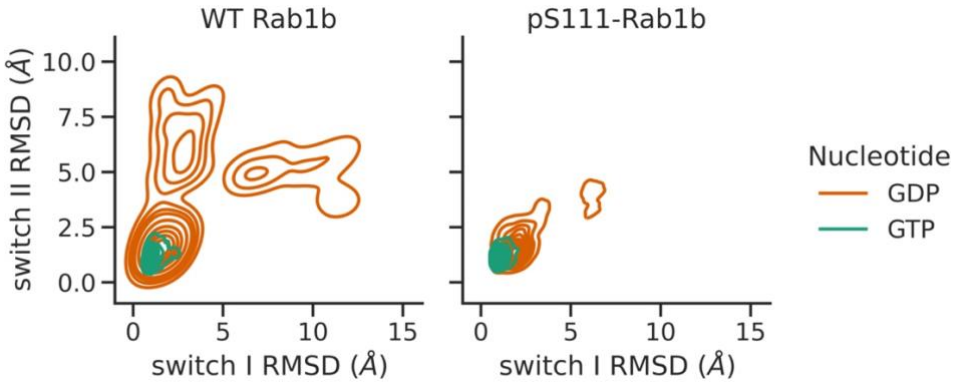

Figure 5-6. Probability distribution of Rab1b switch regions RMSD.
Switch I and switch II amino acid residues deviations from the active Rab1b conformation for wild type (left) and S111-phosphorylated (right) variants. Probability density levels are given in 1/20 fractions of the maximum probability.

bound Rab1b the distribution of RMSDs throughout the simulation was limited to a region near the GTP-bound start structure. A similar distribution was obtained in the GTP-bound complex when S111 was phosphorylated, while in the GDP-bound pS111-Rab1b variant, we observed a contraction in comparison to the wild type. Overall, our data suggests that the phosphorylation of S111 may stabilize the switch regions and reduces the flexibility of the GDP-bound Rab1b.

5.3. Discussion

GTPases such as Rab1b alternate between active (GTP-bound) and inactive (GDP-bound) states which involves structural changes mainly in the switch I and II regions. Rab1b is a member of the Rab branch of GTPases and can undergo post-translational modifications modulating its function. Phosphorylation of S111 is frequently found in several Rab members (e.g. Rab1b and the closely related Rab8a) and in the case of Rab8a, experimental structures of both active and inactive forms with and without S111 phosphorylation are known (152). Our simulation study demonstrates that even during relatively long MD simulations of 2 μs, no transitions to the inactive Rab1b conformation are observed. It indicates the presence of energy barriers for the necessary dihedral transitions that cannot be overcome on the time scale of these simulations. They are attributed to the interactions between the Mg^{2+} ion and conserved residues on switch I and II, namely T35 and G60 in the Ras subfamily (155). However, the application of the DIA-REMD method to specifically enhance the sampling of dihedral transitions in the switch regions allowed the sampling of switch I and II conformational transitions. The large conformational changes were observed in case of a bound GDP, but not in the GTP-bound cases.

Other enhanced sampling techniques have been used previously to sample conformational states of GTPases. It includes enhanced sampling along specific reaction coordinates to drive the switch folding and unfolding transition using Umbrella sampling (157) or Metadynamics (160). Alternatively, accelerated MD simulations using a biased/deformed force field have also been used (161). The disadvantage in the former approach is the necessity of defining a reaction co- ordinate prior the simulation and in the latter case it is the likely oversampling of irrelevant conformations, due to the altered force field. In contrast, in the present approach no definition of a reaction coordinate is necessary and only relevant conformations fully compatible with the original reference force field are considered–no re-weighting of sampled states is necessary. The technique could also be useful in case of other signaling proteins that include multiple conformational states of flexible switching elements.

The DIA-REMD simulations still indicate a free energy minimum near the Rab1b conformation representing the active state in the presence of bound GDP, but also significant sampling of partially unfolded switch I and II segments. The experimental structures of GDP bound Rab GTPases indicate largely disordered switch I and II segments as the dominant conformational state. Hence, insufficient sampling may still be an issue for the DIA-REMD simulations. However, it has been demonstrated that much broader sampling is achieved on the 200 ns time scale compared to 2 μs during the MD simulations (note, however, that the total demand of the DIA-REMD simulations is 8 × 200 ns= 1.6 μs). The much broader sampling of the DIA-REMD simulations at approximately the same total simulation time indicates that it can be very helpful

to overcome energy barriers and to identify flexible protein segments. In addition to sampling issues, current force fields may over-stabilize folded structures relative to disor- dered conformational states (162,163). Indeed, it has been demonstrated that current force fields yield varying results for disordered segments of proteins and this could also explain the significant sampling of the active state in the presence of bound GDP. In the future, the DIA-REMD approach might be helpful to systematically test force field improvements for modeling disordered segments in proteins.

Our results suggest that the phosphorylation of S111 leads to a stabilization of the closed form in both switch regions of the GDP-bound Rab1b complex, while it has no significant effect on the GTP-bound variant. The RMSD plots from the switch regions in Figure 5-6 show a limited range of sampling of open conformations in the presence of phosphorylated serine 111 in Rab1b bound to GDP compared to the wild-type variant. However, the X-ray crystal structures of the closely related inactive and active Rab8a states do not change significantly upon S111 phosphorylation. It is important to emphasize that our simulations do not exclude the possibility that the pS111-Rab1b switch I and II adopt disordered conformations also in solution but indicate an ensemble of structures that is influenced by the pS111 modification. S111 resides in the vicinity of the switch II and once phosphorylated, it forms an intra-molecular bond with the sidechain of R79 on the switch II and the R172 sidechain. This arrangement was found in a large fraction of the entire simulation time (Figure 5-5) and can be interpreted as a likely reason for the reduced flexibility and stabilization of the structure in the presence of pS111. It is important to note, that such hydrogen bonding between pS111 and R79 was also found in the X-ray structure of the closely related pS111-Rab8a GTPase (152). Furthermore, the pS111-Rab1b exhibits an enhanced thermal stability compared to the wild-type, which can be explained qualitatively by the additional pS111-mediated hydrogen bonding (152). In addition to a possible stabilization of the switch region the appearance of a hydrogen bonded salt bridge between pS111 and R79 can also directly interfere with Rab-effector binding (152,164).

5.4. Methods

5.4.1 Simulation protocol

All Rab1b simulations started from the crystal structure of Rab1b (PDB: 3nkv); the adenylylation modification at Y77 was removed (150). The nitrogen atom between the β- and γ-phosphates in GppNHp was exchanged with an oxygen atom to model the Rab1b:GTP structure. The γ-phosphate was removed to generate a start model of the Rab1b:GDP complex. T72 and S111 residues were replaced with unprotonated phosphothreonine and phosphoserine to form the phosphorylated Rab1b structures. Amber ff14SB force field was used for the

proteins (137). Additional parameters for GDP, GTP, phosphothreonine and phosphoserine were taken from Amber parameter database at their fully unprotonated states (138,139). An experimental study by Mann et al. confirmed the unprotonated state of γ-GTP as the dominant protonation state (140). For Rab1b:GDP simulations, a Mg^{2+} ion with two bound water molecules was placed next to GDP β-phosphate. Using sodium and chloride ions the salt concentrations were adjusted to 0.1 M and the complexes were solvated with TIP3P water model (141). The solvated complexes were equilibrated by energy minimization (2500 steps of steepest descent) followed by heating to 300 K (100 ps) and 1 ns of density equilibration at constant pressure (1 bar) and a temperature of 300 K. During heating and equilibration the protein heavy backbone and nucleotide atoms as well as magnesium ions were restraint with a harmonic potential at force constant of 5.0 kcal $mol^{-1}Å^{-2}$. All data gathering continuous MD simulations and replica-exchange (REMD) simulations were performed without any restraints. The pmemd.cuda module of the Amber 16 software package (143) with a time-step of 2 fs, periodic boundary conditions and the particle mesh Ewald (PME) method to account for long range interactions was employed. Trajectories were processed and analyzed using cpptraj program (165) of Amber16. Hierarchical clustering using complete-linkage with a minimum distance of 2 Å RMSD was performed on all frames of the continuous simulations and on the sampling in the reference replica in case of the DIA-REMD technique. The figures were generated using PyMol software package (145).

5.4.2 DIA-REMD technique

The DIA-REMD technique includes a biased potential that promotes transitions from low-energy backbone and sidechain dihedrals states. The starting setup is composed of eight parallel MD simulations–referred to as "replicas"–of the solvated Rab GTPase systems. Favorable dihedral angle combinations of the protein backbone (such as α-helices, β-sheets and left-handed α-helical regimes in the Ramachandran plot) as well as the first sidechain dihedral angles of a preset peptide segment are penalized by adding a penalty potential along the replica simulations. An advantage of the approach is the possibility to limit the biasing to certain peptide segments in the protein structure (only the flexible switch I, residues 31-43, and switch II, residues 68-79, were included in the dihedral angle biasing). The first reference replica, however, runs under the control of the unmodified force field. Every 1000 steps an exchange was attempted between the neighboring replicas and it was allowed or rejected based on the Metropolis criterion. For the biasing potential on the backbone dihedral angles two-dimensional potentials that depend on the ϕ and ψ backbone dihedral angles were employed. The potentials have a maximum at the favorable states (ϕ_c, ψ_c)) in the Ramachandran plot and fall off smoothly (parameters given in Table 5-1) . Let x_i be the shortest distance between the position

of residue *i* on the ϕ-ψ diagram (Ramachandran plot) and an energetically stable conformation, $R(\phi_C, \psi_C)$;

$$x_i = \sqrt{(\phi_i - \phi_c)^2 + (\psi_i - \psi_c)^2} \ .$$

The penalty force is a circular plateau with radius r_1 centered at *R* with the maximum potential (E_{max}), which continuously decreases down to zero at r_2. Moreover, a one-dimensional potential was applied to promote sidechain rotations. Hence, the total potential was calculated by summing the 2D backbone and the 1D sidechain dihedral-angles penalty force.

$$\begin{cases} if \ \ x_i < r_1 & V(x_i) = E_{max} \\ if \ \ r_1 < x_i < r_2 & V(x_i) = \dfrac{E_{max}}{(r_2 - r_1)^4}((x_i - r_1)^2 - (r_2 - r_1)^2)^2 \\ if \ \ r_2 < x_i & V(x_i) = 0 \end{cases}$$

Test simulations showed that with eight replicas sufficient "mixing" between the windows was achieved. In order to assure high rates of replica exchange, the heights of the potentials were adjusted on the fly based on an evaluation of the acceptance ratio during the last 100 exchange attempts; if any of the eight average rates fell below 20 %, the difference in the energy potential between replicas were reduced by 10 %. If all windows show 60 % of average successful exchange, then the difference was increased by 10 %. Typically, after less than 20 ns the biasing levels reached stable levels. Adjustment was therefore stopped after 50 ns and only data gathering beyond 50 ns was used for analysis.

Table 5-1. Energetically favourable regions on the Ramachandran plot that were penalized during the replica-exchange simulations.

	R (ϕ_C, ψ_C)	**r_1**	**r_2**
α-helix	(-57, -47)	22.5	40
β-sheet	(80, 150)	30	40
Left-handed α-helix	(45, 45)	17.5	10

6. Structural Insight From a Self-learning Accelerated-Sampling Algorithm Into Domain Flexibility and Activation Mechanism of Argonaute

6.1. Introduction

The Argonautes constitute a family of proteins that are involved in both transcriptional and post-transcriptional gene regulatory mechanisms. They are present in all forms of life, sharing a well conserved tertiary structure, despite their limited sequence similarity. Prokaryotic Argonautes participate in gene regulation by binding to single-stranded DNAs that guide them towards target DNAs or RNAs that are either cleaved or repressed (166). Moreover, bacterial Argonaute protects its host against invasive genomic elements through directly targeting foreign DNA molecules (167,168).

The overall structure of most Argonautes features four characteristic domains. The N-terminal domain (N) functions as a wedge to unwind guide-target duplexes (169,170). The P-element-induced wimpy testis (PIWI)-Argonaute-Zwille (PAZ) domain hosts the binding pocket for the 3' region of the guide strand. The phosphorylated 5' end of the guide is stabilized at the interface between the middle (MID) and PIWI domains via interaction with a Mg^{2+} cation. The four domains are distributed in two lobes with PAZ and N in one lobe and MID and PIWI in the other lobe. The lobes are connected by two linker domains, L1 and L2.

In a series of structural and biochemical studies Patel and coworkers have determined the structure of the *Thermus Thermophilus* Argonaute (TtAgo) in several guide-bound (binary) and guide/target-bound (ternary) complexes–varying in duplex length and level of duplex complementarity (167,169,171,172). These structural snapshots portrayed a glimpse of the protein in various stages of the silencing process. The structures of Argonaute proteins from other organisms have also been studied (166,173), however, not as extensively as TtAgo. The

structural studies reveal that both 5' and 3' ends of the 21-mer guide DNA strand are anchored in corresponding binding pockets in the MID and PAZ domains respectively. The 3’ end binding to the PAZ domain is characteristic of the *inactive* domain arrangement, which is stabilized in the presence of complementary target strands with a length below 15 nucleotides. Upon target binding followed by base pairing beyond position 16 of the guide strand, a switch to the cleavage-compatible conformation–known as the *active* state occurs, in which the 3’ end is dissociated from PAZ. The active state is distinguished by a tetrad of DEDD amino-acids residues in the PIWI domain that attacks the cleavage site between positions 10 and 11 of the target strand (174). Moreover, with the detachment of the 3' end of the guide, a pivot-like movement of the PAZ domain towards the MID domain is observed (169,171,172,174). Single-molecule fluorescence resonance energy transfer (smFRET) studies of the active-complex formation demonstrated the dynamic rearrangements of PAZ and the 3'-end of the nucleic acid bound to it (175,176). Based on the structural, biochemical and biophysical studies a mechanistic stepwise scheme has evolved with an initial binding of the guide strand to TtAgo, anchoring of the 5’ and 3’ ends followed by binding of a complementary RNA or DNA strand that–if long enough–leads to dissociation of the guide's 3’ end from the PAZ domain, enzyme activation and finally, cleavage of the target. The dissociation of the cleaved target and rebinding of the 3’ end to PAZ leads again to the initial inactive guide-bound state. The critical role of the PAZ domain anchoring of the 3’ end of the guide strand to inactivate the enzyme is further supported by experiments employing short guide DNA or RNA molecules (169,172). Even guide RNAs as short as 9 nucleotides which do not allow the anchoring of the 3’ end at the PAZ domain result in efficient cleavage of target RNA/DNA strands (172). Furthermore, kinetic observations of guide and target RNAs binding to human Argonaute-2 (hAgo2), a homolog of TtAgo, revealed that the 3' end's release from PAZ is the rate-limiting step during the activation process (177). Despite previous structural and biochemical studies, a detailed atomistic view of the extension of the base pairing between guide and target strands leading to the dissociation of the 3’ end and subsequent domain rearrangements and enzyme activation has remained elusive.

TtAgo adopts at least two conformational states (i.e., inactive and active) that involve domain rearrangements, but it is likely that there are additional intermediate states. The available crystal structures of TtAgo in complex with different guide and/or target strands give an excellent overview on stable domain arrangements but do not give insight into the accessible global states for apo TtAgo, or binary and ternary complexes. In principle, Molecular Dynamics (MD) simulations can provide a dynamic high-resolution view on possible structural arrangements of TtAgo. However, crossing the energy barriers separating metastable conformational states, especially when it involves an interplay between a multidomain protein and nucleic acids, may require simulations times well beyond currently accessible MD timescales. Most previous MD

simulation studies of Ago proteins were focused on the nucleic acid recognition and binding process and the corresponding structural changes in hAgo2 (178) and TtAgo (179). In another study, combination of MD with bias-exchange metadynamics and protein-DNA docking methods revealed the induced-fit mechanism of the guide-strand loading in TtAgo (180). In hAgo2, however, a two-step mechanism RNA loading was suggested, based on a study that employed Markov State Models and protein-RNA docking (181).

In the present study we employ a replica-exchange based enhanced-sampling MD method to specifically accelerate domain motions in the TtAgo system in different DNA-bound states and in the absence of any substrates (apo state). In our approach a biasing potential is constructed by a mixture of Gaussians–similar to a Metadynamics simulation–along center-of-mass distance variables. The form and level of biasing in each replica is rapidly adjusted during an equilibration phase and integrated in the force field of the Hamiltonian replica-exchange MD (H-REMD) simulation. It allows exploration of regions in the global conformational landscape that are not sampled in regular MD simulations. In addition to characterizing the domain mobility in the apo state, in the binary and in the ternary complexes, we also apply the methodology to reveal the structural transitions that steer the protein from an inactive to an active conformation. In agreement with the experimental observations, we detect the dissociation of the guide strand's 3' end from the PAZ domain when extending the target/guide duplex beyond a critical length of 14 base pairs, due to sterical strain. The dissociated 3' end ultimately settles in a cleft between N and PIWI domains but remains conformationally highly mobile. The associated conformational substates and intermediates are also investigated.

6.1.1 Brief description of the self-learning algorithm

The bias potential is constructed by a mixture of Gaussians, each of which correspond to a low free energy region of the chosen collective variable (CV). The local free energy minima are dynamically identified by analyzing the trajectory using a clustering algorithm. Cluster analysis of the CVs are performed at regular intervals of 1 ns. The probability distribution of the visited configurations is reconstructed using a function that is comprised of three Gaussian functions. The algorithm provides the centers and weights of the Gaussians. Moreover, the maximum width of each Gaussian hill is determined by the distance between cluster's outermost data-point and its center. Ultimately, the Gaussians shape the adaptive bias potential energies that promote the system to wander in the unexplored regions of configuration space. Every CV is treated separately. This allows for a rapid flattening of the free-energy surface along each CV, which accelerates the escape from the basin. In the initial stages, where the system is energetically trapped in one basin, the Gaussian centers might be adjacent. As the system starts visiting other regions, the centers diverge accordingly. Our choice of CVs in the replica-

exchange simulations is aimed for individual domain deformations. We assumed three CVs, defined by COM distances between i) PAZ and N, ii) MID and PIWI and iii) the upper and lower lobes (see Figure 6-3). More details in Methods.

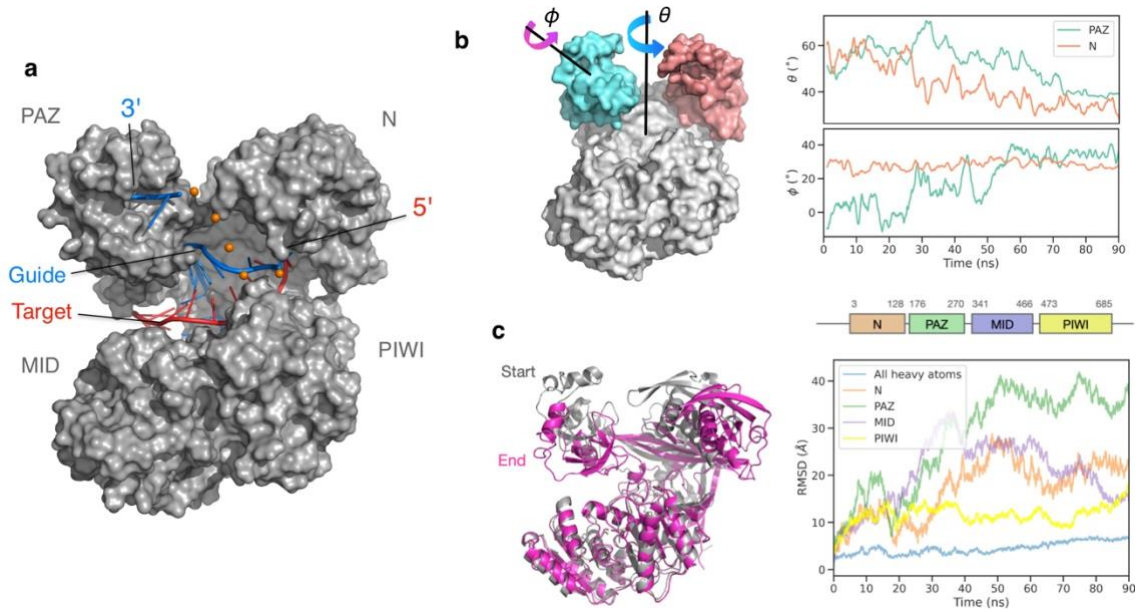

Figure 6-1. Overview of the start structure of Argonaute simulations.
a, crystal structure of prokaryotic Argonaute from Thermus Thermophilus bacterium (TtAgo, PDB: 4n41) bound to 21-mer guide (blue) and 15-mer target (red) DNA strands. The orange spheres show the presumed location of the missing bases. **b**, *in silico* removal of the DNAs causes a rapid rearrangement of the domains. The angles Θ and Φ describe each domain's rotation around the protein and around itself, respectively. **c**, The domains' RMSDs with respect to the initial structure. The last frame of the 90 ns-long free simulation served as the start structure in the following simulations of TtAgo in apo form.

6.2. Results

6.2.1 Structural flexibility of the Argonaute protein during continuous MD simulations

TtAgo is one of the best studied Argonaute proteins and, like other Argonautes, consists of PAZ, N, MID and PIWI domains that undergo domain rearrangements during the enzyme functional cycle (Figure 6-1). A goal of our study is to characterize the accessible domain geometries in the apo TtAgo, the binary (TtAgo bound to 21-mer guide DNA) and the ternary (TtAgo bound to 21-mer guide and 14-mer target DNA) complex. To examine the global domain dynamics, we first examined Argonaute's structural flexibility using extensive regular unbiased MD simulations. Since the crystal structure of TtAgo in apo form is not available, the start structure was generated by removing the DNA from the *inactive* structure of TtAgo bound to 21-mer

guide and 15-mer target DNA strands (PDB: 4n41) (174). Within ~90 ns of MD simulation time, both PAZ and N domains moved towards MID and PIWI, thereby closing the substrate binding channel–also known as the central cleft. This arrangement is considered as a *closed* conformation and is in accordance with Argonaute's "rubber band" model. In this model, the substrate loading to the Ago protein is supported by the Hsp70/Hsp90 chaperone machinery that uses ATP to convert Argonaute from a closed to a more open structure that can accommodate the bulky strands (182). Such opening induces structural strain in the protein, analogous to a stretched rubber band. Release of this tension drives the strand separation without consuming any ATP. During unwinding, the guide strand that has the 5' end stably anchored in the pocket between MID and PIWI domains will remain in the protein, whereas the other unanchored strand–known as the passenger strand–will be discarded (182,183). The protein and the guide strand form a functional silencing complex. The motion of PAZ is both translational and rotational–it translates relative to the other domains (or rotates as a whole body relative to the center of TtAgo) and also rotates around itself–, while the N domain rotates only around the protein's axis (Figure 6-1**b**).

Extending the MD simulation up to 2 μs yielded an even further downward shift of the PAZ domain (Figure 6-2). This pivotal movement is reminiscent of the ones observed in the active TtAgo crystal structures, in which the PAZ has moved downwards and the 3' end of the guide is detached from it (PDB: 3hjf & 4nca) (169,174). It indicates that a motion of PAZ towards the active-like geometries even in the absence of guide and target DNAs is possible. We also performed unrestrained MD simulation of the binary (21-mer guide DNA) and ternary (21-mer guide and 14-mer target DNA) complexes starting from the inactive state (PDB: 4n41, after 1 ns equilibration). Contrary to the apo form, in the binary and ternary complexes the initial domain arrangements were largely preserved during the 2 μs simulations, indicated by low root-mean-square-deviation (RMSD) values. In particular the PAZ domain remained in an inactive arrangement bound to the 3'-end of the guide strand (Figure 6-2). The overall stability observed in these two variants is also attributed to the hydrogen-bond network formed between the guide strand and the protein's backbone atoms (179).

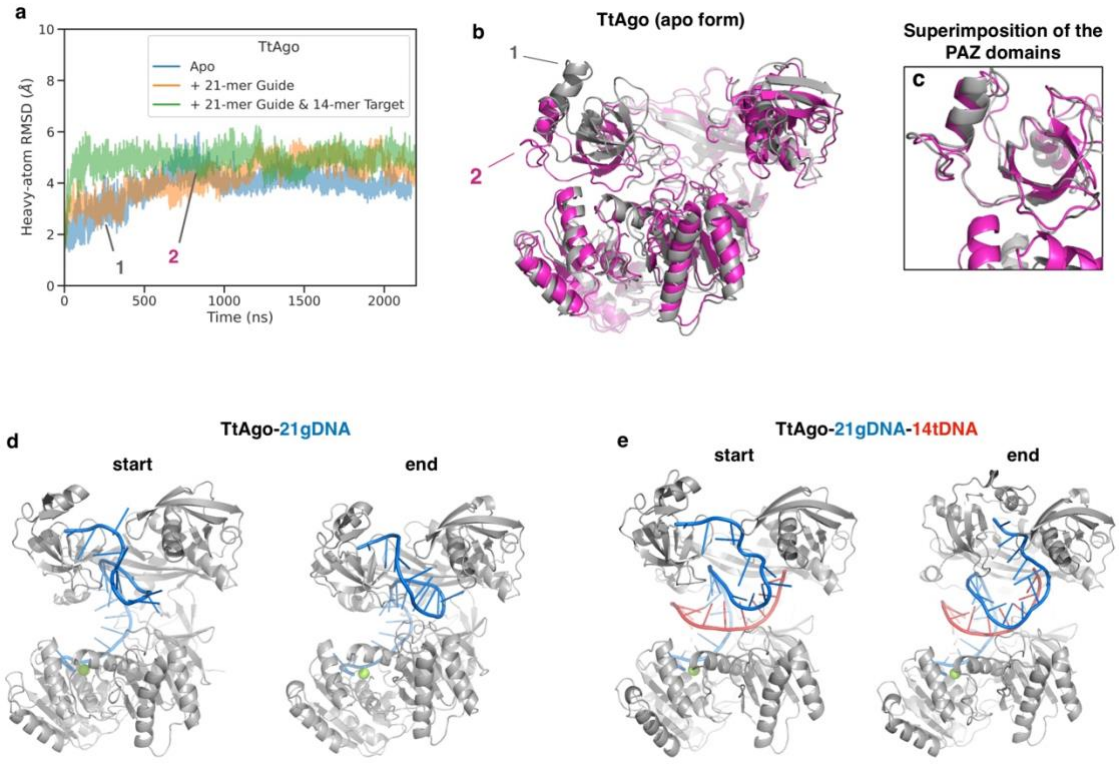

Figure 6-2. Unrestrained MD simulation results.
a, heavy-atom RMSD versus time for three TtAgo structures; apo, binary (21-mer guide DNA) and ternary (guide & 14-mer target DNA) structures. **b**, superimposition of two frames from the apo form simulation. **c**, superimposition of the frames in b based on the PAZ domain amino-acid residues. The conformational changes within the domain are negligible. The start and end structures of the binary and ternary complex simulations are shown in **d** and **e** respectively.

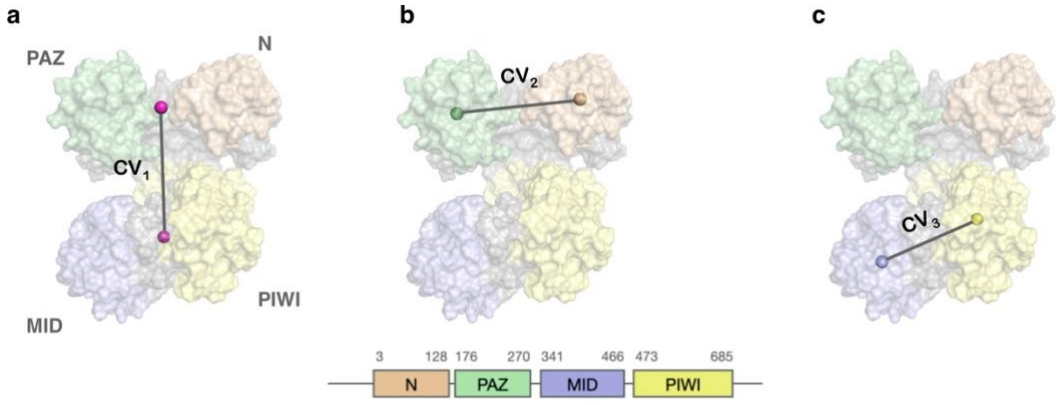

Figure 6-3. Three center-of-mass distances served as collective variables.
Each CV was independently biased by a potential to promote domain motions. The CVs were defined as the distance between centers-of mass of the two lobes (**a**), PAZ & N (**b**), and MID & PIWI (**c**).

6.2.2 Replica-exchange simulation of TtAgo in apo, binary and ternary complexes

During regular MD simulations only limited domain motions of TtAgo were observed. Even in the apo form, after it rapidly reached a closed conformation, the protein remained stuck in that domain arrangement. It might be an artefact due to a limited sampling of relevant conformational states on the time scale of the MD simulations. To sample putative domain arrangements more exhaustively, we employed H-REMD technique coupled with adaptive biasing potentials along pre-selected global collective variables (CVs).

The added biasing potentials act separately on each CV, and therefore it is advantageous that the coupling between the domain movements they promote be limited. Otherwise, in the higher replicas they may lead to sampling of conformations that are of low probability or relevance for the reference replica, which runs with the original force field. To this end, we selected center-of-mass (COM) distances between domains as global variables in a hierarchal manner (Figure 6-3). In TtAgo, the PIWI and MID domains are in close contact but do not directly interact with the PAZ and N domains. Hence, in the first CV we considered PIWI and MID as one unit, and PAZ and N as another unit, and the CV was defined as the distance between the

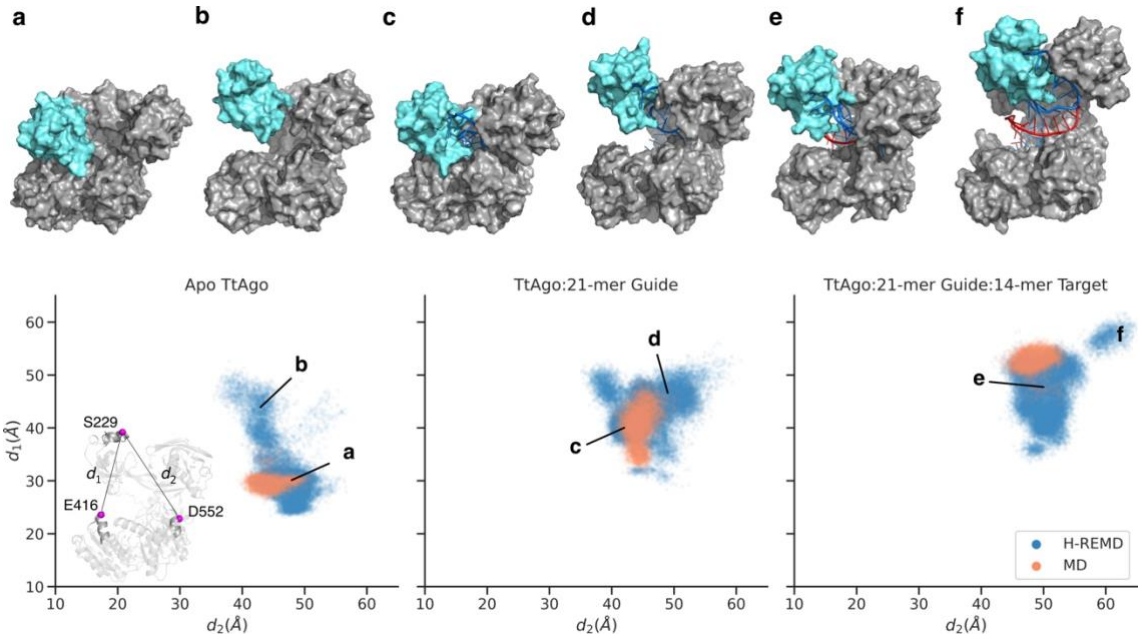

Figure 6-4. Free-energy profiles of the PAZ domain motion in TtAgo variants.
Comparison of the two-dimensional free-energy profiles of the apo protein along with the binary (middle) and ternary (right) complexes, as a function of α-carbon atoms distances d_1 and d_2 for the PAZ domain, which describe motion with respect to MID and PIWI. The red contours represent the data from 2 μs free MD simulations. Each contour represents an increase of 0.53 kcal/mol in free energy. Snapshots in **a-f** show the corresponding conformations on the free-energy profiles. The PAZ domain in shown in cyan colour, the guide and target strands in blue and red respectively.

COM of each of these two units (CV_1). The next two variables were defined as the COM distance between the N and PAZ domains (CV_2) and the COM distance between MID and PIWI domains (CV_3). For the bias force we employed one-dimensional potentials in the replicas that acted independently along the CVs (see Methods for details). During an equilibration phase of the H-REMD simulations the bias potentials were adjusted to accelerate the sampling along the collective variables and the exchanges among replicas.

The H-REMD simulation of the apo TtAgo protein started from the last frame of the 90 ns unrestrained MD simulation, in which the protein had adopted a closed conformation. The binary (21-mer guide DNA) and ternary (21-mer guide and 14-mer target DNA) variants started from the same structure as in the unrestrained MD runs. Evaluation of the sampled states and the free-energy calculations were done solely based on the sampling in the unbiased reference replica. A comparison of the conformational landscape of the PAZ domain in the three variants is illustrated in Figure 6-4. The plot reflects PAZ's movements relative to MID and PIWI domains, which–based on physical intuition–are measured by two α-carbon (Cα) distances, d_1 and d_2. The distance between S229 and E416 (d_1) represents PAZ's motion relative to MID, while the distance between S229 and E552 (d_2) represents PAZ's motion relative to PIWI. In all the three variants, a much broader range of the PAZ domain motion was sampled during the H-REMD simulation with an adaptive biasing potential compared to the unrestrained MD simulation. Snapshots obtained from the H-REMD simulations indicated that the apo TtAgo, in addition to the closed state, adopts conformations that are distinguished by large PAZ-MID distances ($d_1 \approx 50$ Å). Interestingly, in the open states, the PAZ domain's position relative to MID and PIWI is similar to the low-energy states of the guide-bound TtAgo complex. The similarity in PAZ arrangements between the apo protein and the binary complex reflects the capability of the TtAgo protein to rearrange the domains and widen the central cleft in order to accommodate the bulky DNA strands. Such conformations in the apo protein, however, are energetically not favorable, due to the absence of the guide's hydrogen-bond network.

The clustering results in Figure 6-5 indicate that in the guide- and guide/target-bound TtAgo, both 3’ and phosphorylated 5' end of the guide remained stably anchored in their binding pockets. Importantly, the coordination of the magnesium ion with the first phosphates of the 5' end–whose phosphorylation is critical for the cleavage activity–and the sidechain of V685 of the PIWI domain was also unaltered (166,172). Additionally, the seed region of the guide strand (position 2-8) exhibits the lowest flexibility and remained constantly solvent exposed. This arrangement reduces the energy barrier involved in base pairing with target strands (184).

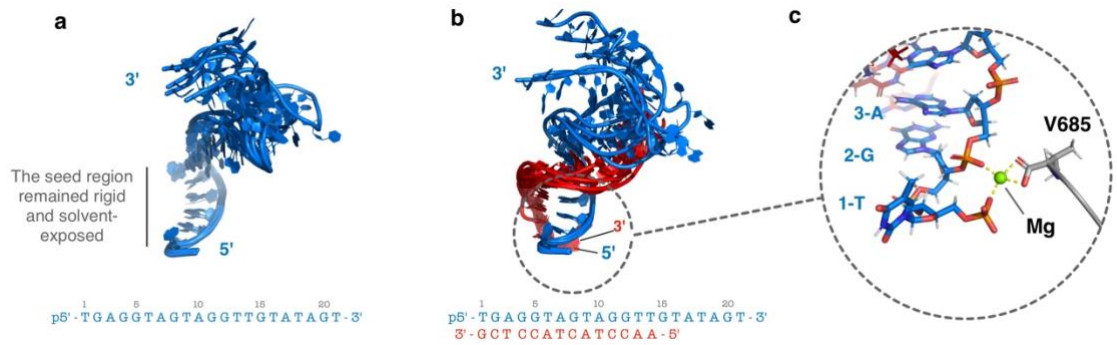

Figure 6-5. Superimposition of the cluster analysis results shows that the seed region of the DNA strand has the lowest flexibility.
Cluster representatives obtained from the simulation trajectories in the binary (**a**) and ternary (**b**) complex shows a non-uniform level of rigidity along the DNA strands. The seed region remained solvent-exposed in all clusters of the guide-bound complex. The guide and target strands are displayed in blue and red cartoon respectively. **c**, the interactions between the Mg ion (green sphere), V685 and the phosphates of the first and the third base remained stable during the simulations.

6.2.3 Extending the target length to position 16 triggers guide's 3'-end release from PAZ.

Next, we performed H-REMD simulations of TtAgo bound to 5'-phosphorylated 21-mer guide DNA and fully complementary 15- and 16-mer target DNAs. The 15th and 16th bases of the target strand were added to the previous ternary complex (PDB: 4n41) to form the start structure, while maintaining the Watson-Crick hydrogen bonds with the opposing base on the guide strand. Comparing the simulation results of the 15-mer and 16-mer complexes with the previous 14-mer complex led to interesting observations. Firstly, with the extension of the DNA duplex, the N-PIWI gap was found to be wider. The gap was measured as the distance between the α-carbon atoms of M82 from N and D552 from PIWI and was increased from 26 Å in the 14-mer target complex to 32 Å in the 15-mer and 46 Å in the 16-mer target complex (Figure 6-7). Secondly, the cluster analysis of the 14-mer target complex indicated that in 64% of the simulation time, the duplex bases at positions 12-14 had lost base pairing and were unstacked from their neighbors–a phenomena called fraying (Figure 6-6). On the contrary, we observed no faying events in the 15-mer and 16-mer target complex. The occurrence of fraying in the 14-mer target complex is further evident from high values of root mean-square atomic fluctuations (RMSF) in the bases 10 to 14 (Figure 6-7e). Overall, there seems to exist an interplay between the N domain and the DNA duplex. With 14-mer target, the duplex is weak and frayed due to the N domain's sterical interactions. With 15- and 16-mer targets, the duplex gains stability, and pushes back the N domain, which widens the gap between N and PIWI.

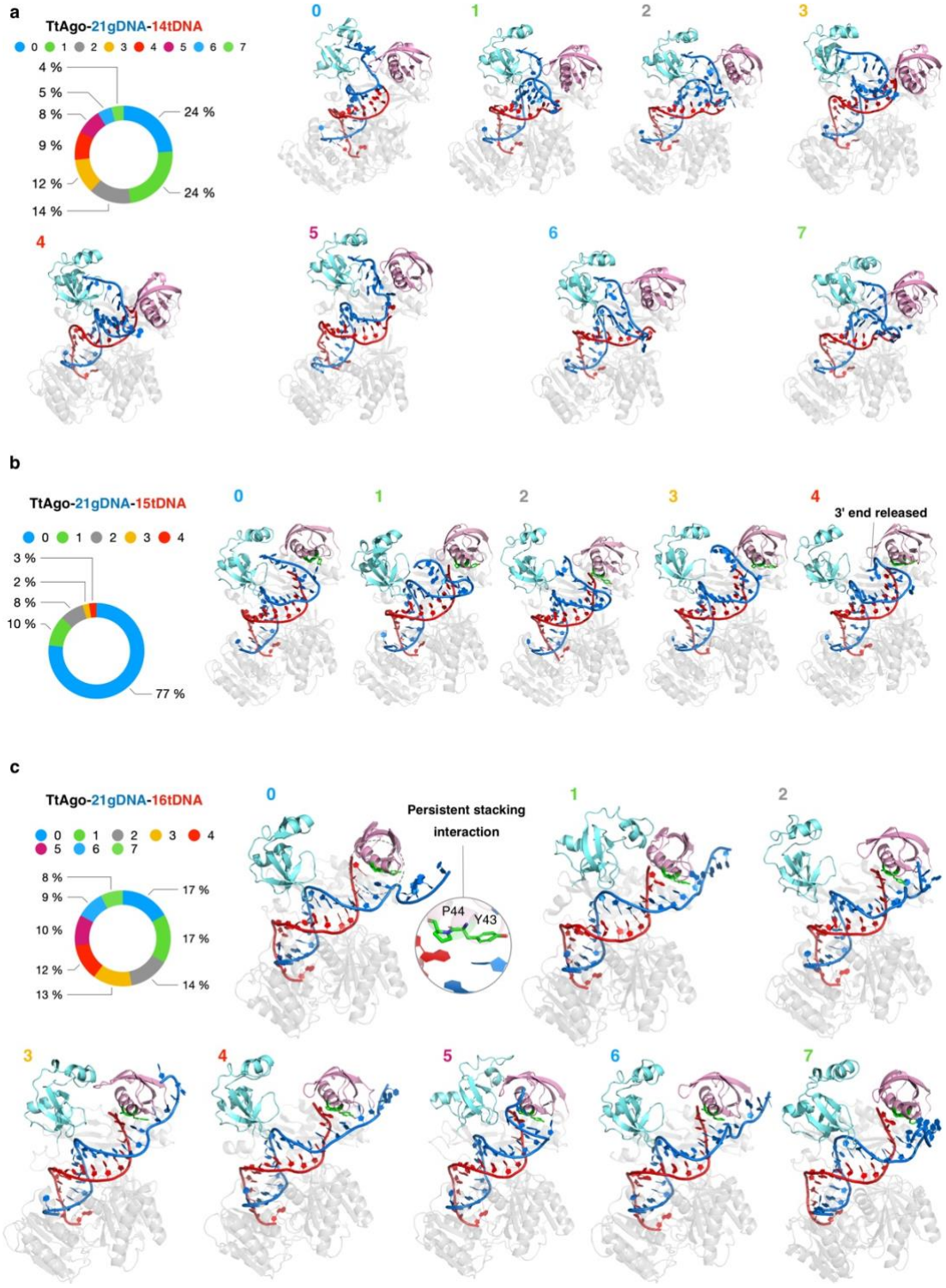

Figure 6-6. Clustering of the TtAgo ternary complex simulations.
Cluster populations (pie charts) along with the cluster representatives of three guide/target-bound TtAgo proteins with varying target length: 14 (**a**), 15 (**b**) and 16 (**c**) nucleotide bases. Cluster analysis in the 14-mer target complex was performed based on the RMSD of the PAZ and N domains. In the 15-mer and 16-mer target complexes the clustering was based on the 3'-end - PAZ distance (OP2 phosphate atom of the 3'end and hydroxyl group of Y226).

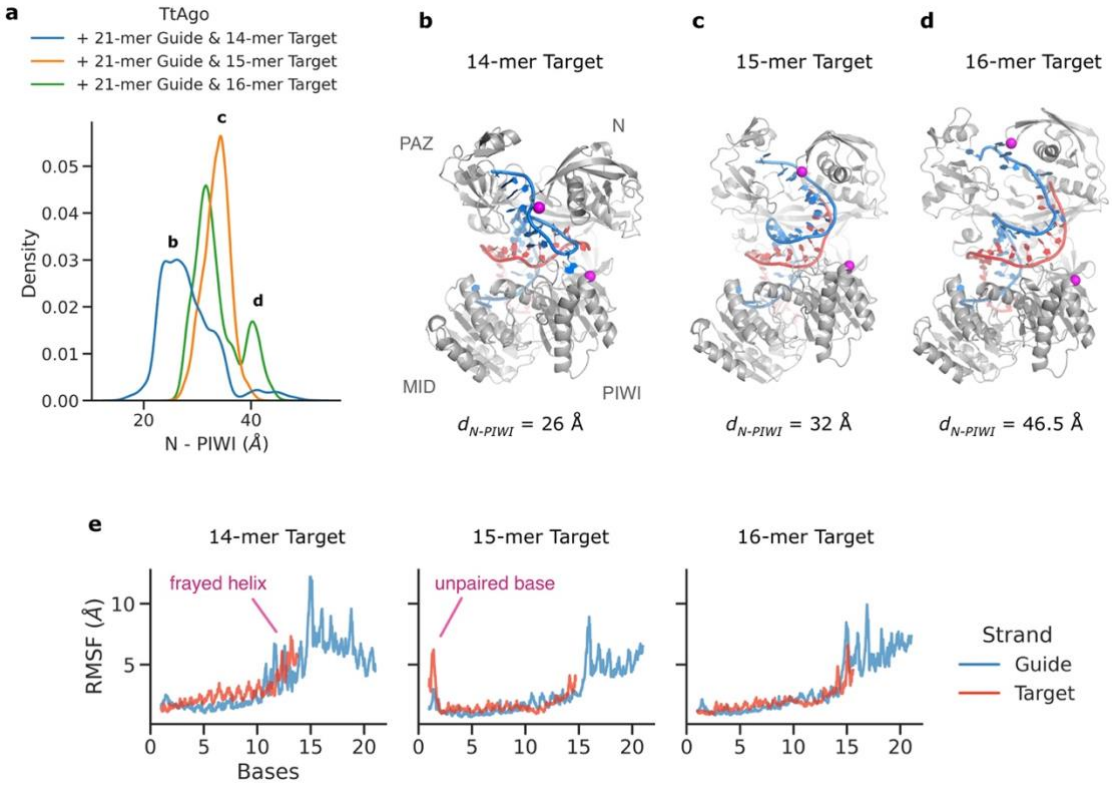

Figure 6-7. Widening in the N-PIWI cleft with extension of the DNA duplex base pairing.
a, distribution of the N-PIWI distances in three variants of Argonaute with varying target length. The distances are measured between α-carbons of the two residues M82 from N and D552 from PIWI domain, shown in magenta spheres. The distance distributions indicate a shift towards N-PIWI opening as the target length is increased to 16 bases. This is also evident in the representative structures taken from the simulations of 14, 15, and 16-mer target DNA ternary complex are shown in **b**-**d** respectively. The guide and target strands are shown in blue and red cartoons). **e**, root mean square fluctuations (RMSF) of each base for all the DNA duplexes. The RMSF values were calculated with respect to the average structure.

In the inactive TtAgo structures, the 3’ end of the guide strand is bound to the PAZ domain. It is known that with the extension of the guide/target duplex beyond position 15, the TtAgo protein adopts an active conformation in which the 3’ end is released from the PAZ domain and PAZ has moved towards the MID domain (174). Interestingly, we observed the dissociation of the 3' end when proceeding to 15- and 16-mer target strands. Cluster analysis showed that in 3% of the frames of the 15-mer-bound simulation the 3' end was released from PAZ, nevertheless still present in the gap between PAZ and N domains (Figure 6-6**b**). This fraction was increased to 6% in the 16-mer target-bound complex. The release of the 3' end observed in 15-mer and 16-mer complexes explains the increased RMSF values in the guide's 3'-end region in Figure 6-7**e** compared to the 14-mer target complex. Restarting the 16-mer-bound simulation from one of these frames resulted in a complete transition of the 3' end towards the N-PIWI gap. Snapshots of the 3' end's release from the PAZ domain are illustrated in Figure 6-8. The stacking of the aromatic rings of Y43 and P44 over the duplex bases at position 16 was strikingly

persistent throughout the release. We postulate that this stacking interaction is a prerequisite for the push back on the N domain, release of the 3' end from PAZ and the activation of the Argonaute protein.

The conformation with the released 3' end is a cleavage-incompatible conformation, as the catalytic tetrad in the PIWI domain was not formed. Nevertheless, it superimposes well with the active TtAgo complex structure (Figure 6-8**e**). Next, we measured the distance between a phosphate atom of the 3' end (OP2) and its interacting partner on PAZ (hydroxyl group of Y226). While in the 14-mer target DNA complex, the distance was consistently short (~2.8 Å), it drastically increased in the 15- and 16-mer target variants during the H-REMD simulations (Figure 6-8**f**). The associated free-energy calculations in Figure 6-8**f** shows that with the extension of the guide/target duplex to position 15 and 16, the energy landscape changes in favor of the 3'-end release.

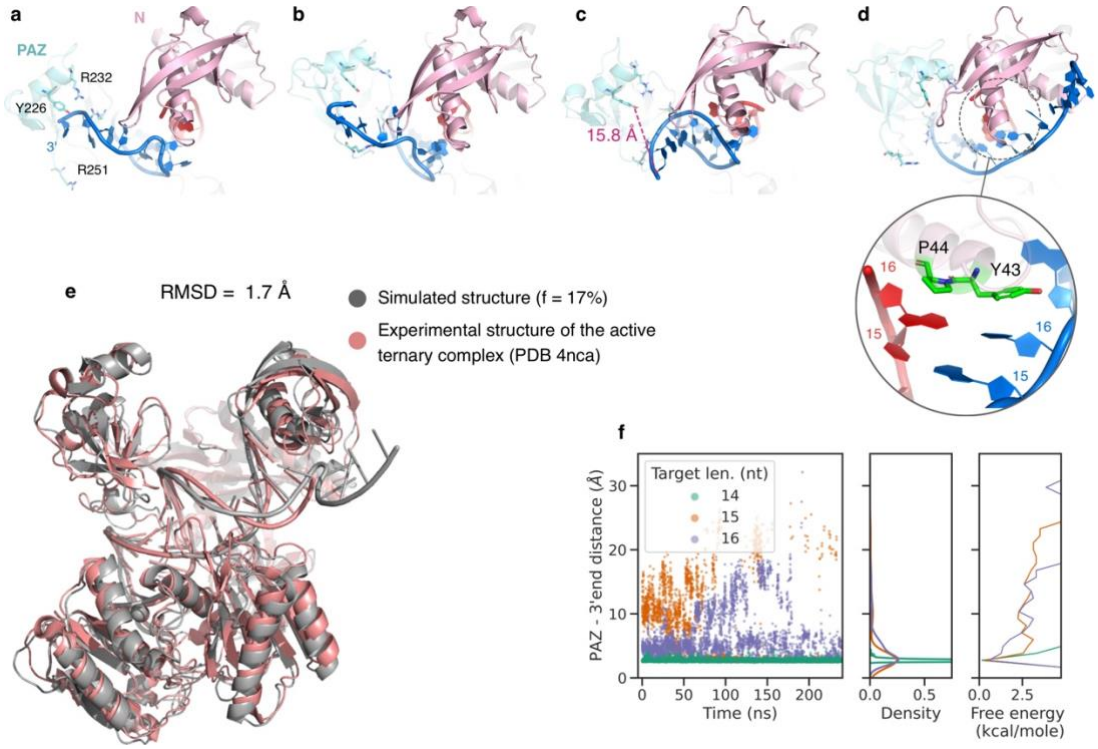

Figure 6-8. Release of the 3'-end from the PAZ domain in ternary structure with 16-mer target. **a-d**, simulation results showing four dynamical states of the release of guide 3'-end from the PAZ domain in the ternary complex of TtAgo with 21-mer guide and 16-mer target DNA strands. The guide is represented in blue, the target in red and the PAZ and N domains in cyan and pink respectively. The stacking interactions between Y43 & P44 (green sticks) and guide and target strands at position 16 are persistent throughout the release process. **e**, superimposition of a simulated structure (cluster representative with 17 % population) shown in grey cartoon, on the catalytically active TtAgo structure shown in pink. **f**, the distance between the hydroxyl group of Y226, located on PAZ, and the guide 3'-end (OP2) is plotted for TtAgo in complex with three different target lengths.

6.3. Discussion

The molecular mechanism of the Argonaute's transition from inactive to active conformation has not been addressed previously. Here, we proposed an accelerated-sampling scheme combining replica-exchange method with a self-learning algorithm for introducing a bias that rapidly explores the conformational space. The algorithm uses the Gaussian mixture model to detect low free-energy minima and adds the bias force along the CVs to fill the minima to avoid getting trapped. Since they are treated independently, any number of CVs can be simultaneously probed with no performance loss.

The accelerated-sampling scheme was demonstrated on the Argonaute protein, using center-of-mass (COM) distances between domains in a hierarchical manner as CVs. In the 2 μs-long unrestrained MD simulations, the apo protein immediately transitioned to a closed conformation, and afterwards its overall structure was only minimally changed. The accelerated-sampling results showed that the nucleic acid-free TtAgo protein has a broader conformational space available to it, which includes conformations characterized by an open substrate binding channel. The opening of the channel is necessary as the protein cannot readily accommodate the bulky nucleic acid duplexes. The energy required to impose the opening is compensated by the ATP hydrolysis, while the internal tension caused by it drives the subsequent duplex unwinding without consuming ATP (182).

The accelerated-sampling trajectories also provide considerable insight into the guide- and guide/target-bound complex structures. Notably, in the binary complex direct contacts between the PAZ and MID and PIWI domains are formed, which bury the bases 13-17 of the guide DNA in the central cleft. Nevertheless, the seed region of the guide remains exposed to the solvent, which allows probing the target candidates, while adopting a low-energy conformation. In the 14-mer target ternary complex, the cluster analysis results showed that the DNA duplex is frayed in the majority of the times, which is attributed to steric repulsive force coming from the N domain. With a frayed duplex, the guide strand lacks the required strain to dissociate from the PAZ pocket. This explains the observations that in the cleavage assays the truncation of the target strand from its 5' end to positions 15 and 14 (relative to the guide strand) sharply reduces the cleavage activity (169).

Duplex propagation in the 15-mer and 16-mer ternary complexes improves its stability, as indicated by the absence of transient fraying events, and more importantly, by the increased tendency of the guide 3' end to dissociate from PAZ and transition towards adopting the active conformations. In the crystal structure of the 19-mer ternary complex it has been revealed that the N domain blocks the DNA duplex by stacking the aromatic rings of Y43 and P44 on the DNA bases at position 16 (169). Based on our observations, the stacking interactions seem to

form a lever for the duplex to open the N-PIWI channel. In addition, stacking of Y43 on base 16 of the guide strand creates an anchor point for the dissociation of the 3'-end from PAZ and its rotation around the N domain. The structures with the released 3' end appeared to represent cleavage-incompatible conformations as the catalytic tetrad of DEDD amino-acids residues in the PIWI were not transitioned to the active arrangement. Such transition requires additional rearrangements in loops that were not included in our choice of the CVs, but are facilitated by higher temperatures (185).

Our results suggest that domain-domain COM distances are general, yet relevant metrics that drive specific biological processes in the Argonaute proteins. The technique can be also useful for a rapid investigation of conformational changes in any protein of interest that has similar domain structure.

6.4. Methods

6.4.1 Thermos Thermophilus Argonaute starting structure

The protein structure was taken from PDB 4n41 corresponding to the ternary TtAgo complex with 21-mer guide DNA and 15 nucleotides of target DNA (174). The DNA duplex in the crystal structure is resolved in positions 1-14 and 20, 21 in the guide strand and 1-14 in the target strands. All structurally resolved nucleotides were kept and restrained during the equilibration phases of the simulations. The missing bases–position 15-19 of the guide and 15 & 16 of the target strands–were added to the structures and the Watson-Crick base pairings were imposed using distance restraints during an equilibration phase of simulations. The missing residues were added partly from other TtAgo structures and partly using the MODELLER software (186). The ff14SB force field and TIP3P water model were used to model proteins and the explicit solvent molecules (141,187). The parameters for the phosphorylated 5' end of the guide strand were generated using the generalized Amber force field (gaff) (188). The OL15 force field refinements were employed for nucleic acids (189). The solvated box was then energy minimized (500 steps), followed by 25 ps of heating and 50 ps of density equilibration in an NPT ensemble with the pressure kept at 1 bar and temperature adjusted to 300 K. During these phases, the protein's heavy atoms, the nucleotide and the magnesium ions were restrained at their initial positions using a harmonic potential with a decreasing force constant, starting at 5.0 kcal·mol·$Å^{-2}$ and ending with 1.0 kcal·mol·$Å^{-2}$. The solvated box was equilibrated further in a restraint-free NPT ensemble at 300 K and1 bar for 1 ns. The temperature for the actual data-gathering production run was 315 K. The GPU-accelerated pmemd version of the Amber 18 software package was used implementing the

hydrogen mass repartitioning feature of the Parmed tool, which allows a simulation time step of 4 fs (7). Long range interactions were included using the particle mesh Ewald (PME) method combined with periodic boundary conditions and an 8.5 Å cut-off for non-bonded interactions. Figures were generated using the PyMol software package (145).

6.4.2 Replica-exchange simulation protocol

The simulation setup employed version 18 of the Amber software package and a python library. The Amber code was modified to accommodate a Gaussian-shaped bias potential between COM distances in the GPU version. The replica-exchange simulations were initialized using a batch file that contained all the input parameters, including the number of replicas, timesteps and exchange attempts per window, the residue IDs involved in CVs and the overall number of windows. The term "window" here refers to a cycle of i) replica-exchange sampling, ii) analysis of the CVs and iii) update of the biasing forces. Simulations started with eight structurally identical replicas. The first 1 ns of simulations ran without any added bias potentials, with the aim to have an initial approximation of the CV values in equilibrium. The trajectory files were updated every 8 ps, and the exchanges were attempted with the same time intervals. A Metropolis criterion was used to allow or reject the exchange attempts. The trajectories of the previous 10 ns from all eight replicas were read by the python library in 1 ns intervals to calculate the CV values. The calculation of the COM distances that define the CVs and other trajectory analysis were performed using Pytraj python library (165). Next, the library fitted the data to a Gaussian mixture model (GMM) using the scikit-learn machine learning package (190). The GMM assumes that the datapoints are collected from a mixture of K Gaussian distributions–called components–with unknown means, variances, and mixture component weights. Initial test runs indicated that using three components along each CV can efficiently reconstruct the arbitrary shape of the distribution i.e., $K = 3$. For a univariate GMM with K components, the k-th component has a mean of μ_k and variance of σ_k. The mixture component weights are defined as ϕ_k, with the condition that $\sum_{i=1}^{k} \phi_i = 1$, i.e., the total probability distribution sums up to one. Our implementation of the GMM fits the distribution of the CVs to a weighted sum of a three-component Gaussian density, given by the equation,

$$p(x) = \sum_{i=1}^{K} \phi_i \, \mathcal{N}(\mathrm{x} \mid \mu_i, \sigma_i),$$

$$\mathcal{N}(\mathrm{x} \mid \mu_i, \sigma_i) = \frac{1}{\sigma_i \sqrt{2\pi}} \, exp\left(- \frac{(x - \mu_i)^2}{2\sigma_i^2} \right).$$

The output of the GMM–the components means, variances, and weights–shapes the next window's biasing forces. The bias potential was incremented by 4 kcal·mol^{-1} in the replicas 2-8 while the reference replica ran without any added potential.

$$\sum_{r=1}^{8} B_r(\mathrm{CV}) = (r-1) * 4\,kCal/mol * p(x)$$

In doing so, each simulation is biased by an external potential $B_r(\mathrm{CV})$ that is built iteratively based on the sampling of the previous 10 ns in all replicas. This cycle goes on until the maximum number of windows is reached. The simulation procedure is illustrated in the figure below.

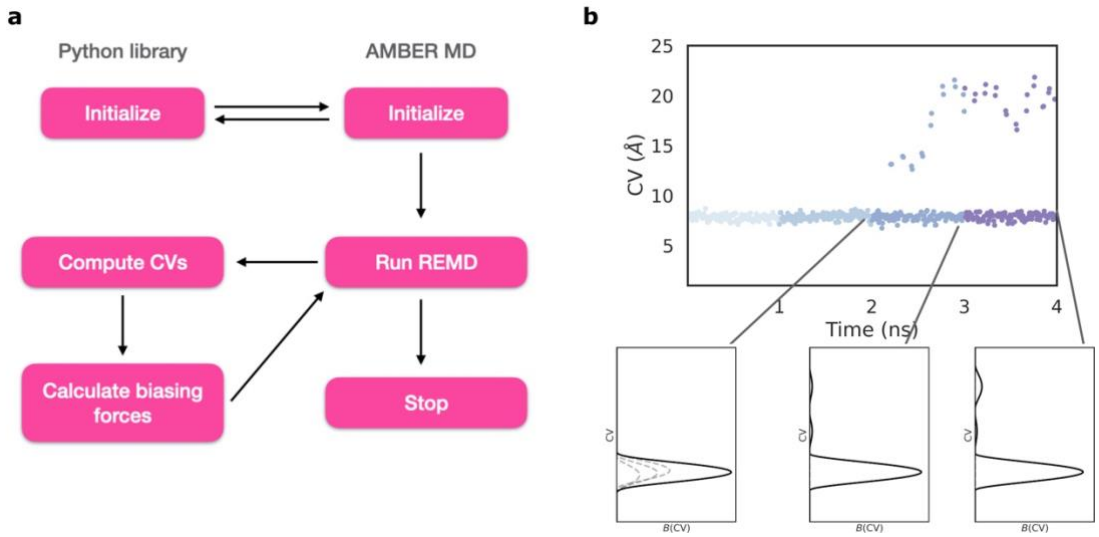

Figure 6-9. Illustration of the advanced sampling algorithm.
a, the accelerated-dynamics simulation algorithm. The python library and the Amber MD package are initialized using a batch file that contains input parameters. There is no biasing force inserted to the system in the first window. At the end of each window the trajectories are passed to the library and CVs are calculated. Based on their values in the previous runs the biasing force is updated. The accelerated-dynamics simulation then continues with the updated biasing forces introduced in the replicas. The cycle goes on until the simulation is stopped. **b**, an illustration of the biasing potential acting on an exemplary CV vs. time. The value of CV is coloured based on the simulation window. In the first two windows the CV is trapped, while in the third and fourth window with the addition of the biasing potential it moves to higher values. The overall bias potential (solid black line) is the sum of three Gaussians (dashed grey lines), each of which represents an energy minimum.

Clustering analysis. Trajectories were processed and analyzed to find similar conformations using the CPPTRAJ tool and the DBSCAN clustering algorithm (24). We used RMSD of the protein's heavy atoms as the distance metric and used every fifth frame of the trajectories to reduce memory consumption. A minimum of 4 conformations with the distance cutoff of 1.25

Å were required to form a cluster. The initial "sieve" value was set to 10 random frames to form the initial clusters. The sieved frames were then added to the clusters as an additional step.

7. Conclusion and Perspective

Protein-docking platforms and accelerated molecular dynamics simulations are two powerful tools that offer a deep understanding of the biomolecular assemblies, as described in this work. In Chapter 3, a thorough overview of the current physics-based methods used to predict protein-protein complex structures was presented. We saw how the protein-docking tools account for flexibility of the components and take advantage of the available experimental data to refine and score the predicted results. A possible future direction in this field would be to develop models for predicting larger protein complex structures and protein assemblies that are composed of multiple components.

Chapter 4 illustrated that covalent modification of a single amino-acid residue, Ser111, in small GTPase Rab8a protein can influence the interaction with the exchange factor Rabin8. It was shown that the modification perturbs a favorable intermolecular salt-bridge contact between R79 in Rab8a and aspartate 187 in Rabin8 and therefore, decreases the binding affinity. The results indicated that post-translational modifications of the residues that are not directly part of the protein-protein interface, can interfere with the sidechain networks of the interacting residues and thereby, influence the complex formation.

In Chapter 5, a dihedral-angle-biasing enhanced-sampling technique was presented and used to study the conformational transitions in Rab1b proteins triggered by GTP/GDP loading. The technique successfully enhanced the sampling and captured events that are rarely observed using conventional MD methods, namely unfolding of the switch regions. Provided the enhanced sampling, the influence of phosphorylation of Ser111 in Rab1b was scrutinized and it was found that it stabilizes the active state in the presence of GDP but has negligible influence on the GTP-bound complex. The technique can be further applied to other PTMs and GTPase proteins as it enhances the sampling at a limited computational cost.

Finally, in Chapter 6, an enhance-sampling algorithm coupled with the replica-exchange methodology was presented, which identifies and compensates for low-energy conformations using a bias potential in a self-learning manner. As the test case, the method was successfully tested on investigation of large domain motions in bacterial Argonaute protein and its conformational changes upon activation. This algorithm could be further used for other large multi-domain proteins and in the specific case of Argonaute protein, it can be coupled with

quantum mechanics/molecular mechanics (QM/MM) methods to fully capture the activation of the bacterial Argonaute and the cleavage of the target strand, which would complete the description of the Argonaute gene-silencing activity.

For decades, experimental and computational scientists studying biological systems have been driven by the classical structural biology paradigm, which states that molecular structure prompts biological function. Structures reveal the three-dimensional organization of the components and the molecular composition of the complex surface. Additionally, they help identify structural motifs or amino-acid residue arrangements that underpin certain functions of the protein aggregate. In spite of the advancements in the experimental characterization methods, a comprehensive study of protein assemblies still remains a challenging task, not only due to their startling abundance and variety, but also owing to the transient nature of some interactions, which hinders a high atomic resolution. Molecular dynamic simulations have a significant advantage in this respect.

Both hardware and software developments, such as supercomputers and graphics processing unit (GPU)-based acceleration algorithms, have increased the reliability and efficiency of biomolecular simulations and have expanded the system size and timescales within their reach. Microsecond simulation of 12-base pair B-DNA (191), millisecond simulation of protein folding (192) and massive all-atom simulation of HIV-1 capsid (193) are among the landmark achievements of the past 20 years that reflect those advancements. The merits of MD simulations became even more evident during the recent global pandemic by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), as it provided a detailed mechanistic insight into the spike protein as a target for vaccine and therapeutic agents (194,195). Today, scientists routinely use computer simulations in combination with other experimental methods to investigate various aspects of biological systems, in the same way the light microscopes were used in the seventeenth century.

The previous advances of computational biology offer overwhelming and undeniable evidence of its brighter future. Refinement, parametrization, and validation of the force fields will continue and constantly improve their accuracy in the future. Moreover, their applicability to various biomolecular systems will be enhanced. Novel algorithmic developments will contribute to solve insufficient sampling of large macromolecules. Finally, artificial intelligence platforms will continue to incorporate the growing body of experimentally gathered data to predict structures, mechanisms, and functions of these incredibly important systems.

What a privilege to have glimpsed at the universe within the living cell.

List of Figures

List of Tables

Acknowledgements

I would like to thank the brilliant members of the Biomolecular Dynamics Group and especially, Prof. Martin Zacharias for his trust, positive attitude, and numerous inspirational discussions. A special thanks to Dr. Jonathan Coles for his amazing management of our high-performance computing cluster, and Sonja Ortner for her commendable handling of the administrative work. This work would not have been possible without funding from Deutsche Forschungsgemeinschaft through the SFB 1035 project, and the facilities provided by the Leibniz Supercomputing Centre of the Bavarian Academy of Sciences and Humanities. During our collaborative research work on Rab proteins I worked closely with Dr. Sergey Savitskiy and Dr. Sophie Vieweg and got help from many others at the Technical University of Munich. I have been blessed throughout my time as a PhD student with wonderfully supportive parents, family, and friends too many to mention individually. Here, I would like to express my gratitude to Stella, whose care and understanding enabled the closure of this chapter of my life.
From my heart thanks.

References

1. Fulyani F, Schuurman-Wolters GK, Žagar AV, Guskov A, Slotboom D-J, Poolman B. Functional Diversity of Tandem Substrate-Binding Domains in ABC Transporters from Pathogenic Bacteria. Structure. 2013 Oct 8;21(10):1879–88.

2. Scrima N, Lepault J, Boulard Y, Pasdeloup D, Bressanelli S, Roche S. Insights into Herpesvirus Tegument Organization from Structural Analyses of the 970 Central Residues of HSV-1 UL36 Protein. Journal of Biological Chemistry. 2015 Apr 3;290(14):8820–33.

3. Dill KA, Ozkan SB, Shell MS, Weikl TR. The Protein Folding Problem. Annu Rev Biophys. 2008 May 7;37(1):289–316.

4. Harrington L, Fletcher JM, Heermann T, Woolfson DN, Schwille P. De novo design of a reversible phosphorylation-dependent switch for membrane targeting. Nature Communications. 2021 Mar 5;12(1):1472.

5. Müller MM. Post-Translational Modifications of Protein Backbones: Unique Functions, Mechanisms, and Challenges. Biochemistry. 2018 Jan 16;57(2):177–85.

6. Nooren IMA, Thornton JM. Diversity of protein–protein interactions. The EMBO Journal. 2003 Jul 15;22(14):3486–92.

7. D.A. Case, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, D. Ghoreishi, M.K. Gilson, H. Gohlke, A.W. Goetz, D. Greene, R Harris, N. Homeyer, Y. Huang, S. Izadi, A. Kovalenko, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D.J. Mermelstein, K.M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L. Simmerling, J. Smith, R. Salomon- Ferrer, J. Swails, R.C. Walker, J. Wang, H. Wei, R.M. Wolf, X. Wu, L. Xiao, D.M. York and P.A. Kollman. AMBER 2018. University of California, San Francisco; 2018.

8. Verlet L. Computer “Experiments” on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. Phys Rev. 1967 Jul 5;159(1):98–103.

9. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. The Journal of Chemical Physics. 1984 Oct 15;81(8):3684–90.

10. Hoover WG. Canonical dynamics: Equilibrium phase-space distributions. Phys Rev A. 1985 Mar 1;31(3):1695–7.

11. Parrinello M, Rahman A. Polymorphic transitions in single crystals: A new molecular dynamics method. Journal of Applied Physics. 1981 Dec;52(12):7182–90.

12. Laio A, Parrinello M. Escaping free-energy minima. Proceedings of the National Academy of Sciences. 2002 Oct 1;99(20):12562–6.

13. Bussi G, Laio A. Using metadynamics to explore complex free-energy landscapes. Nature Reviews Physics. 2020;1–13.

14. Hošek P, Toulcová D, Bortolato A, Spiwok V. Altruistic Metadynamics: Multisystem Biased Simulation. J Phys Chem B. 2016 Mar 10;120(9):2209–15.

15. Fu H, Zhang H, Chen H, Shao X, Chipot C, Cai W. Zooming across the Free-Energy Landscape: Shaving Barriers, and Flooding Valleys. J Phys Chem Lett. 2018 Aug 16;9(16):4738–45.

16. Fukunishi H, Watanabe O, Takada S. On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction. The Journal of Chemical Physics. 2002 May 22;116(20):9058–67.

17. Woods CJ, Essex JW, King MA. The Development of Replica-Exchange-Based Free-Energy Methods. J Phys Chem B. 2003 Dec;107(49):13703–10.

18. Khavrutskii IV, Wallqvist A. Improved Binding Free Energy Predictions from Single-Reference Thermodynamic Integration Augmented with Hamiltonian Replica Exchange. J Chem Theory Comput. 2011 Sep 13;7(9):3001–11.

19. Zwanzig RW. High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. J Chem Phys. 1954 Aug 1;22(8):1420–6.

20. Meng Y, Sabri Dashti D, Roitberg AE. Computing Alchemical Free Energy Differences with Hamiltonian Replica Exchange Molecular Dynamics (H-REMD) Simulations. J Chem Theory Comput. 2011 Sep 13;7(9):2721–7.

21. Ostermeir K, Zacharias M. Rapid Alchemical Free Energy Calculation Employing a Generalized Born Implicit Solvent Model. J Phys Chem B. 2015 Jan 22;119(3):968–75.

22. Nooren IMA. NEW EMBO MEMBER'S REVIEW: Diversity of protein-protein interactions. The EMBO Journal. 2003 Jul 15;22(14):3486–92.

23. Marsh JA, Teichmann SA. Structure, Dynamics, Assembly, and Evolution of Protein Complexes. Annual Review of Biochemistry. 2015;84(1):551–75.

24. Mosca R, Céol A, Aloy P. Interactome3D: adding structural details to protein networks. Nature Methods. 2013 Jan;10(1):47–53.

25. Elmlund D, Le SN, Elmlund H. High-resolution cryo-EM: the nuts and bolts. Current Opinion in Structural Biology. 2017 Oct 1;46:1–6.

26. Güntert P. Automated structure determination from NMR spectra. Eur Biophys J. 2008 Sep 20;38(2):129.

27. Clore GM, Gronenborn AM. Determining the structures of large proteins and protein complexes by NMR. Trends in Biotechnology. 1998 Jan 1;16(1):22–34.

28. Soni N, Madhusudhan MS. Computational modeling of protein assemblies. Current Opinion in Structural Biology. 2017 Jun;44:179–89.

29. Im W, Liang J, Olson A, Zhou H-X, Vajda S, Vakser IA. Challenges in structural approaches to cell modeling. Journal of Molecular Biology. 2016 Jul 31;428(15):2943–64.

30. Johnson GT, Autin L, Al-Alusi M, Goodsell DS, Sanner MF, Olson AJ. cellPACK: a virtual mesoscope to model and visualize structural systems biology. Nature Methods. 2015 Jan;12(1):85–91.

31. Zacharias M. Accounting for conformational changes during protein–protein docking. Current Opinion in Structural Biology. 2010 Apr 1;20(2):180–6.

32. Bonvin AM. Flexible protein–protein docking. Current Opinion in Structural Biology. 2006 Apr 1;16(2):194–200.

33. Kundrotas PJ, Anishchenko I, Dauzhenka T, Kotthoff I, Mnevets D, Copeland MM, et al. Dockground: A comprehensive data resource for modeling of protein complexes. Protein Science. 2018;27(1):172–81.

34. Kundrotas PJ, Zhu Z, Janin J, Vakser IA. Templates are available to model nearly all complexes of structurally characterized proteins. PNAS. 2012 Jun 12;109(24):9438–41.

35. Porter KA, Desta I, Kozakov D, Vajda S. What method to use for protein-protein docking? Curr Opin Struct Biol. 2019 Jan 31;55:1–7.

36. Gromiha MM, Yugandhar K, Jemimah S. Protein–protein interactions: scoring schemes and binding affinity. Current Opinion in Structural Biology. 2017 Jun;44:31–8.

37. Koshland DE. Das Schüssel-Schloß-Prinzip und die Induced-fit-Theorie. Angew Chem. 1994 Dec 19;106(23–24):2468–72.

38. Csermely P, Palotai R, Nussinov R. Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. Trends in Biochemical Sciences. 2010 Oct 1;35(10):539–46.

39. Geng C, Xue LC, Roel-Touris J, Bonvin AMJJ. Finding the ΔΔG spot: Are predictors of binding affinity changes upon mutations in protein–protein interactions ready for it? Wiley Interdisciplinary Reviews: Computational Molecular Science. 2019 Jun 14;0(0):e1410.

40. Moal IH, Fernández-Recio J. SKEMPI: a Structural Kinetic and Energetic database of Mutant Protein Interactions and its use in empirical models. Bioinformatics. 2012 Oct 15;28(20):2600–7.

41. S. Moreira I, M. Martins J, S. Coimbra JT, J. Ramos M, A. Fernandes P. A new scoring function for protein–protein docking that identifies native structures with unprecedented accuracy. Physical Chemistry Chemical Physics. 2015;17(4):2378–87.

42. Andrusier N, Mashiach E, Nussinov R, Wolfson HJ. Principles of flexible protein–protein docking. Proteins: Structure, Function, and Bioinformatics. 2008;73(2):271–89.

43. Katchalski-Katzir E, Shariv I, Eisenstein M, Friesem AA, Aflalo C, Vakser IA. Molecular surface recognition: determination of geometric fit between proteins and

their ligands by correlation techniques. Proc Natl Acad Sci USA. 1992 Mar 15;89(6):2195–9.

44. Mashiach E, Schneidman-Duhovny D, Peri A, Shavit Y, Nussinov R, Wolfson HJ. An Integrated Suite of Fast Docking Algorithms. Proteins. 2010 Nov 15;78(15):3197–204.

45. Chen R, Li L, Weng Z. ZDOCK: An initial-stage protein-docking algorithm. Proteins: Structure, Function, and Bioinformatics. 2003;52(1):80–7.

46. Kozakov D, Hall DR, Xia B, Porter KA, Padhorny D, Yueh C, et al. The ClusPro web server for protein–protein docking. Nature Protocols. 2017 Feb;12(2):255–78.

47. Ritchie DW, Kemp GJL. Protein docking using spherical polar Fourier correlations. Proteins: Structure, Function, and Bioinformatics. 2000;39(2):178–94.

48. Padhorny D, Kazennov A, Zerbe BS, Porter KA, Xia B, Mottarella SE, et al. Protein–protein docking by fast generalized Fourier transforms on 5D rotational manifolds. PNAS. 2016 Jul 26;113(30):E4286–93.

49. Venkatraman V, Sael L, Kihara D. Potential for Protein Surface Shape Analysis Using Spherical Harmonics and 3D Zernike Descriptors. Cell Biochem Biophys. 2009 Jul 1;54(1):23–32.

50. Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ. PatchDock and SymmDock: servers for rigid and symmetric docking. Nucleic Acids Res. 2005 Jul 1;33(suppl_2):W363–7.

51. Zacharias M. ATTRACT: Protein–protein docking in CAPRI using a reduced protein model. Proteins: Structure, Function, and Bioinformatics. 2005;60(2):252–6.

52. Ruiz MEE, de Beauchêne IC, Ritchie DW. EROS-DOCK: Protein-Protein Docking Using Exhaustive Branch-and-Bound Rotational Search. Bioinformatics [Internet]. 2019 May 29 [cited 2019 May 29]; Available from: https://academic.oup.com/bioinformatics/advance-article/doi/10.1093/bioinformatics/btz434/5498285

53. Protein–protein docking with a reduced protein model accounting for side-chain flexibility - Zacharias - 2003 - Protein Science - Wiley Online Library [Internet]. [cited 2019 Jun 26]. Available from: https://onlinelibrary.wiley.com/doi/full/10.1110/ps.0239303

54. Schneider S, Saladin A, Fiorucci S, Prévost C, Zacharias M. ATTRACT and PTOOLS: Open Source Programs for Protein–Protein Docking. In: Baron R, editor. Computational Drug Discovery and Design [Internet]. New York, NY: Springer New York; 2012 [cited 2019 Jun 26]. p. 221–32. (Methods in Molecular Biology). Available from: https://doi.org/10.1007/978-1-61779-465-0_15

55. Chaudhury S, Berrondo M, Weitzner BD, Muthu P, Bergman H, Gray JJ. Benchmarking and Analysis of Protein Docking Performance in Rosetta v3.2. PLOS ONE. 2011 Aug 2;6(8):e22477.

56. Kuroda D, Gray JJ. Pushing the Backbone in Protein-Protein Docking. Structure. 2016 Oct 4;24(10):1821–9.

57. Moal IH, Bates PA. SwarmDock and the Use of Normal Modes in Protein-Protein Docking. International Journal of Molecular Sciences. 2010 Oct;11(10):3623–48.

58. Dominguez C, Boelens R, Bonvin AMJJ. HADDOCK: A Protein−Protein Docking Approach Based on Biochemical or Biophysical Information. J Am Chem Soc. 2003 Feb 19;125(7):1731–7.

59. Deplazes E, Davies J, Bonvin AMJJ, King GF, Mark AE. Combination of Ambiguous and Unambiguous Data in the Restraint-driven Docking of Flexible Peptides with HADDOCK: The Binding of the Spider Toxin PcTx1 to the Acid Sensing Ion Channel (ASIC) 1a. J Chem Inf Model. 2016 Jan 25;56(1):127–38.

60. Janin J. Assessing predictions of protein–protein interaction: The CAPRI experiment. Protein Science. 2005;14(2):278–83.

61. Lensink MF, Méndez R, Wodak SJ. Docking and scoring protein complexes: CAPRI 3rd Edition. Proteins: Structure, Function, and Bioinformatics. 2007;69(4):704–18.

62. Lensink MF, Wodak SJ. Docking and scoring protein interactions: CAPRI 2009. Proteins: Structure, Function, and Bioinformatics. 2010;78(15):3073–84.

63. Lensink MF, Velankar S, Wodak SJ. Modeling protein–protein and protein–peptide complexes: CAPRI 6th edition. Proteins: Structure, Function, and Bioinformatics. 2017;85(3):359–77.

64. Carter P, Lesk VI, Islam SA, Sternberg MJE. Protein–protein docking using 3D-Dock in rounds 3, 4, and 5 of CAPRI. Proteins: Structure, Function, and Bioinformatics. 2005;60(2):281–8.

65. Mandell JG, Roberts VA, Pique ME, Kotlovyi V, Mitchell JC, Nelson E, et al. Protein docking using continuum electrostatics and geometric fit. Protein Eng Des Sel. 2001 Feb 1;14(2):105–13.

66. Tovchigrechko A, Vakser IA. GRAMM-X public web server for protein–protein docking. Nucleic Acids Res. 2006 Jul 1;34(suppl_2):W310–4.

67. Macindoe G, Mavridis L, Venkatraman V, Devignes M-D, Ritchie DW. HexServer: an FFT-based protein docking server powered by graphics processors. Nucleic Acids Res. 2010 Jul 1;38(suppl_2):W445–9.

68. Fernández-Recio J, Totrov M, Abagyan R. ICM-DISCO docking by global energy optimization with fully flexible side-chains. Proteins: Structure, Function, and Bioinformatics. 2003;52(1):113–7.

69. Heifetz A, Katchalski-Katzir E, Eisenstein M. Electrostatics in protein–protein docking. Protein Science. 2002;11(3):571–87.

70. Ohue M, Shimoda T, Suzuki S, Matsuzaki Y, Ishida T, Akiyama Y. MEGADOCK 4.0: an ultra–high-performance protein–protein docking software for heterogeneous supercomputers. Bioinformatics. 2014 Nov 15;30(22):3281–3.

71. Chowdhury R, Rasheed M, Keidel D, Moussalem M, Olson A, Sanner M, et al. Protein-Protein Docking with F2Dock 2.0 and GB-Rerank. PLOS ONE. 2013 Mar 6;8(3):e51307.

72. Ramírez-Aportela E, López-Blanco JR, Chacón P. FRODOCK 2.0: fast protein–protein docking server. Bioinformatics. 2016 Aug 1;32(15):2386–8.

73. Jiménez-García B, Pons C, Fernández-Recio J. pyDockWEB: a web server for rigid-body protein–protein docking using electrostatics and desolvation scoring. Bioinformatics. 2013 Jul 1;29(13):1698–9.

74. Mitra P, Pal D. PRUNE and PROBE--two modular web services for protein-protein docking. Nucleic Acids Res. 2011 Jul;39(Web Server issue):W229-234.

75. Pallara C, Jiménez-García B, Romero M, Moal IH, Fernández-Recio J. pyDock scoring for the new modeling challenges in docking: Protein–peptide, homo-multimers, and domain–domain interactions [Internet]. Proteins: Structure, Function, and Bioinformatics. 2017 [cited 2019 Jun 24]. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/prot.25184

76. Lasker K, Phillips JL, Russel D, Velázquez-Muriel J, Schneidman-Duhovny D, Tjioe E, et al. Integrative Structure Modeling of Macromolecular Assemblies from Proteomics Data. Molecular & Cellular Proteomics. 2010 Aug 1;9(8):1689–702.

77. Schmidt C, Macpherson JA, Lau AM, Tan KW, Fraternali F, Politis A. Surface Accessibility and Dynamics of Macromolecular Assemblies Probed by Covalent Labeling Mass Spectrometry and Integrative Modeling. Anal Chem. 2017 Feb 7;89(3):1459–68.

78. Vreven T, Schweppe DK, Chavez JD, Weisbrod CR, Shibata S, Zheng C, et al. Integrating Cross-Linking Experiments with Ab Initio Protein–Protein Docking. Journal of Molecular Biology. 2018 Jun 8;430(12):1814–28.

79. Orbán-Németh Z, Beveridge R, Hollenstein DM, Rampler E, Stranzl T, Hudecz O, et al. Structural prediction of protein models using distance restraints derived from cross-linking mass spectrometry data. Nat Protoc. 2018;13(3):478–94.

80. de Vries SJ, Schindler CEM, Chauvot de Beauchêne I, Zacharias M. A Web Interface for Easy Flexible Protein-Protein Docking with ATTRACT. Biophysical Journal. 2015 Feb 3;108(3):462–5.

81. Schneidman-Duhovny D, Hammel M, Tainer JA, Sali A. FoXS, FoXSDock and MultiFoXS: Single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles. Nucleic Acids Res. 2016 Jul 8;44(W1):W424–9.

82. Jiménez-García B, Pons C, Svergun DI, Bernadó P, Fernández-Recio J. pyDockSAXS: protein-protein complex structure by SAXS and computational docking. Nucleic Acids Res. 2015 Jul 1;43(W1):W356-361.

83. Xia B, Mamonov A, Leysen S, Allen KN, Strelkov SV, Paschalidis IC, et al. Accounting for observed small angle X-ray scattering profile in the protein–protein docking server cluspro. Journal of Computational Chemistry. 2015;36(20):1568–72.

84. Sønderby P, Rinnan Å, Madsen JJ, Harris P, Bukrinski JT, Peters GHJ. Small-Angle X-ray Scattering Data in Combination with RosettaDock Improves the Docking Energy Landscape. J Chem Inf Model. 2017 Oct 23;57(10):2463–75.

85. Schindler CEM, de Vries SJ, Sasse A, Zacharias M. SAXS Data Alone can Generate High-Quality Models of Protein-Protein Complexes. Structure. 2016 Aug 2;24(8):1387–97.

86. van Zundert GCP, Melquiond ASJ, Bonvin AMJJ. Integrative Modeling of Biomolecular Complexes: HADDOCKing with Cryo-Electron Microscopy Data. Structure. 2015 May 5;23(5):949–60.

87. Vries SJ de, Zacharias M. ATTRACT-EM: A New Method for the Computational Assembly of Large Molecular Machines Using Cryo-EM Maps. PLOS ONE. 2012 Dec 14;7(12):e49733.

88. de Vries SJ, Chauvot de Beauchêne I, Schindler CEM, Zacharias M. Cryo-EM Data Are Superior to Contact and Interface Information in Integrative Modeling. Biophysical Journal. 2016 Feb 23;110(4):785–97.

89. Pons C, Grosdidier S, Solernou A, Pérez-Cano L, Fernández-Recio J. Present and future challenges and limitations in protein-protein docking. Proteins. 2010 Jan;78(1):95–108.

90. Mashiach E, Schneidman-Duhovny D, Andrusier N, Nussinov R, Wolfson HJ. FireDock: a web server for fast interaction refinement in molecular docking. Nucleic Acids Res. 2008 Jul 1;36(suppl_2):W229–32.

91. Vries SJ de, Dijk ADJ van, Krzeminski M, Dijk M van, Thureau A, Hsu V, et al. HADDOCK versus HADDOCK: New features and performance of HADDOCK2.0 on the CAPRI targets. Proteins: Structure, Function, and Bioinformatics. 2007;69(4):726–33.

92. Li L, Chen R, Weng Z. RDOCK: Refinement of rigid-body protein docking predictions. Proteins: Structure, Function, and Bioinformatics. 2003;53(3):693–707.

93. Comeau SR, Gatchell DW, Vajda S, Camacho CJ. ClusPro: an automated docking and discrimination method for the prediction of protein complexes. Bioinformatics. 2004 Jan 1;20(1):45–50.

94. Wang C, Bradley P, Baker D. Protein–Protein Docking with Backbone Flexibility. Journal of Molecular Biology. 2007 Oct 19;373(2):503–19.

95. Chaudhury S, Gray JJ. Conformer Selection and Induced Fit in Flexible Backbone Protein–Protein Docking Using Computational and NMR Ensembles. Journal of Molecular Biology. 2008 Sep 12;381(4):1068–87.

96. Schindler CEM, Vries SJ de, Zacharias M. iATTRACT: Simultaneous global and local interface optimization for protein–protein docking refinement. Proteins: Structure, Function, and Bioinformatics. 2015;83(2):248–58.

97. Martin MG, Siepmann JI. Transferable Potentials for Phase Equilibria. 1. United-Atom Description of n-Alkanes. J Phys Chem B. 1998 Apr 1;102(14):2569–77.

98. Wang T, Wade RC. Implicit solvent models for flexible protein–protein docking by molecular dynamics simulation. Proteins: Structure, Function, and Bioinformatics. 2003;50(1):158–69.

99. Perilla JR, Goh BC, Cassidy CK, Liu B, Bernardi RC, Rudack T, et al. Molecular dynamics simulations of large macromolecular complexes. Current Opinion in Structural Biology. 2015 Apr 1;31:64–74.

100. Pan AC, Jacobson D, Yatsenko K, Sritharan D, Weinreich TM, Shaw DE. Atomic-level characterization of protein–protein association. PNAS. 2019 Mar 5;116(10):4244–9.

101. Luitz MP, Zacharias M. Protein–Ligand Docking Using Hamiltonian Replica Exchange Simulations with Soft Core Potentials. J Chem Inf Model. 2014 Jun 23;54(6):1669–75.

102. Ostermeir K, Zacharias M. Accelerated flexible protein-ligand docking using Hamiltonian replica exchange with a repulsive biasing potential. PLOS ONE. 2017 Feb 16;12(2):e0172072.

103. Zhang C, Liu S, Zhu Q, Zhou Y. A Knowledge-Based Energy Function for Protein−Ligand, Protein−Protein, and Protein−DNA Complexes. J Med Chem. 2005 Apr 1;48(7):2325–35.

104. Shoemaker BA, Panchenko AR. Deciphering Protein–Protein Interactions. Part II. Computational Methods to Predict Protein and Domain Interaction Partners. PLOS Computational Biology. 2007 Apr 27;3(4):e43.

105. Chuang G-Y, Kozakov D, Brenke R, Comeau SR, Vajda S. DARS (Decoys As the Reference State) Potentials for Protein-Protein Docking. Biophysical Journal. 2008 Nov 1;95(9):4217–27.

106. Liu S, Zhang C, Zhou H, Zhou Y. A physical reference state unifies the structure-derived potential of mean force for protein folding and binding. Proteins: Structure, Function, and Bioinformatics. 2004;56(1):93–101.

107. Ballester PJ, Mitchell JBO. A machine learning approach to predicting protein–ligand binding affinity with applications to molecular docking. Bioinformatics. 2010 May 1;26(9):1169–75.

108. Feliu E, Aloy P, Oliva B. On the analysis of protein–protein interactions via knowledge-based potentials for the prediction of protein–protein docking. Protein Science. 2011;20(3):529–41.

109. Wang C, Greene D, Xiao L, Qi R, Luo R. Recent Developments and Applications of the MMPBSA Method. Front Mol Biosci [Internet]. 2018 [cited 2019 May 15];4. Available from: https://www.frontiersin.org/articles/10.3389/fmolb.2017.00087/full

110. Chen F, Liu H, Sun H, Pan P, Li Y, Li D, et al. Assessing the performance of the MM/PBSA and MM/GBSA methods. 6. Capability to predict protein–protein binding free energies and re-rank binding poses generated by protein–protein docking. Physical Chemistry Chemical Physics. 2016;18(32):22129–39.

111. Spiliotopoulos D, Kastritis PL, Melquiond ASJ, Bonvin AMJJ, Musco G, Rocchia W, et al. dMM-PBSA: A New HADDOCK Scoring Function for Protein-Peptide Docking. Front Mol Biosci [Internet]. 2016 [cited 2019 Jun 27];3. Available from: https://www.frontiersin.org/articles/10.3389/fmolb.2016.00046/full

112. Gilson MK, Given JA, Bush BL, McCammon JA. The statistical-thermodynamic basis for computation of binding affinities: a critical review. Biophysical Journal. 1997 März;72(3):1047–69.

113. Mobley DL, Gilson MK. Predicting Binding Free Energies: Frontiers and Benchmarks. Annual Review of Biophysics. 2017;46(1):531–58.

114. Woo H-J, Roux B. Calculation of absolute protein-ligand binding free energy from computer simulations. Proceedings of the National Academy of Sciences. 2005 Mai;102(19):6825–30.

115. Gumbart JC, Roux B, Chipot C. Efficient determination of protein–protein standard binding free energies from first principles. Journal of chemical theory and computation. 2013;9(8):3789–98.

116. May A, Pool R, van Dijk E, Bijlard J, Abeln S, Heringa J, et al. Coarse-grained versus atomistic simulations: realistic interaction free energies for real proteins. Bioinformatics. 2014 Feb 1;30(3):326–34.

117. Siebenmorgen T, Zacharias M. Evaluation of Predicted Protein–Protein Complexes by Binding Free Energy Simulations. J Chem Theory Comput. 2019 Mar 12;15(3):2071–86.

118. Blümer J, Rey J, Dehmelt L, Mazel T, Wu Y-W, Bastiaens P, et al. RabGEFs are a major determinant for specific Rab membrane targeting. Journal of Cell Biology. 2013 Feb 4;200(3):287–300.

119. Cherfils J, Zeghouf M. Regulation of Small GTPases by GEFs, GAPs, and GDIs. Physiological Reviews. 2013 Jan;93(1):269–309.

120. Guo Z, Hou X, Goody RS, Itzen A. Intermediates in the Guanine Nucleotide Exchange Reaction of Rab8 Protein Catalyzed by Guanine Nucleotide Exchange Factors Rabin8 and GRAB. Journal of Biological Chemistry. 2013 Nov;288(45):32466–74.

121. Hattula K, Furuhjelm J, Arffman A, Peränen J. A Rab8-specific GDP/GTP Exchange Factor Is Involved in Actin Remodeling and Polarized Membrane Transport. Mostov K, editor. MBoC. 2002 Sep;13(9):3268–80.

122. Hutagalung AH, Novick PJ. Role of Rab GTPases in Membrane Traffic and Cell Physiology. Physiological Reviews. 2011 Jan;91(1):119–49.

123. Müller MP, Goody RS. Molecular control of Rab activity by GEFs, GAPs and GDI. Small GTPases. 2018 Mar 4;9(1–2):5–21.

124. Zhen Y, Stenmark H. Cellular functions of Rab GTPases at a glance. Journal of Cell Science. 2015 Sep 1;128(17):3171–6.

125. Lai Y, Kondapalli C, Lehneck R, Procter JB, Dill BD, Woodroof HI, et al. Phosphoproteomic screening identifies Rab GTP ases as novel downstream targets of PINK 1. EMBO J. 2015 Nov 12;34(22):2840–61.

126. Shinde SR, Maddika S. A modification switch on a molecular switch: Phosphoregulation of Rab7 during endosome maturation. Small GTPases. 2016 Jul 2;7(3):164–7.

127. Shinde SR, Maddika S. Post translational modifications of Rab GTPases. Small GTPases. 2018 Mar 4;9(1–2):49–56.

128. Steger M, Tonelli F, Ito G, Davies P, Trost M, Vetter M, et al. Phosphoproteomics reveals that Parkinson's disease kinase LRRK2 regulates a subset of Rab GTPases. eLife. 2016 Jan 29;5:e12813.

129. Miller BR, McGee TD, Swails JM, Homeyer N, Gohlke H, Roitberg AE. MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. J Chem Theory Comput. 2012 Sep 11;8(9):3314–21.

130. Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, et al. Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. Accounts of Chemical Research. 2000 Dezember;33(12):889–97.

131. Pourjafar-Dehkordi D, Vieweg S, Itzen A, Zacharias M. Phosphorylation of Ser111 in Rab8a Modulates Rabin8-Dependent Activation by Perturbation of Side Chain Interaction Networks. Biochemistry. 2019 Aug 20;58(33):3546–54.

132. Ong ST, Freeley M, Skubis-Zegadło J, Fazil MHUT, Kelleher D, Fresser F, et al. Phosphorylation of Rab5a Protein by Protein Kinase Cϵ Is Crucial for T-cell Migration*. Journal of Biological Chemistry. 2014 Jul;289(28):19420–34.

133. Xiang S, Gapsys V, Kim H-Y, Bessonov S, Hsiao H-H, Möhlmann S, et al. Phosphorylation Drives a Dynamic Switch in Serine/Arginine-Rich Proteins. Structure. 2013 Dec;21(12):2162–74.

134. Kano Y, Gebregiworgis T, Marshall CB, Radulovich N, Poon BPK, St-Germain J, et al. Tyrosyl phosphorylation of KRAS stalls GTPase cycle via alteration of switch I and II conformation. Nature Communications. 2019 Jan 15;10(1):224.

135. Bunda S, Heir P, Srikumar T, Cook JD, Burrell K, Kano Y, et al. Src promotes GTPase activity of Ras via tyrosine 32 phosphorylation. Proceedings of the National Academy of Sciences. 2014 Sep 9;111(36):E3785–94.

136. Pos W, Sethi DK, Wucherpfennig KW. Mechanisms of peptide repertoire selection by HLA-DM. Trends in Immunology. 2013 Oct;34(10):495–501.

137. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. J Chem Theory Comput. 2015 Aug 11;11(8):3696–713.

138. Homeyer N, Horn AHC, Lanig H, Sticht H. AMBER force-field parameters for phosphorylated amino acids in different protonation states: phosphoserine, phosphothreonine, phosphotyrosine, and phosphohistidine. J Mol Model. 2006 Feb;12(3):281–9.

139. Meagher KL, Redman LT, Carlson HA. Development of polyphosphate parameters for use with the AMBER force field. J Comput Chem. 2003 Jul 15;24(9):1016–25.

140. Mann D, Höweler U, Kötting C, Gerwert K. Elucidation of Single Hydrogen Bonds in GTPases via Experimental and Theoretical Infrared Spectroscopy. Biophysical Journal. 2017 Jan;112(1):66–77.

141. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. The Journal of Chemical Physics. 1983 Jul 15;79(2):926–35.

142. Jorgensen WL, Jenson C. Temperature dependence of TIP3P, SPC, and TIP4P water from NPT Monte Carlo simulations: Seeking temperatures of maximum density. JOURNAL OF COMPUTATIONAL CHEMISTRY. 19(10):8.

143. D.A. Case et. al. AMBER 2016. University of California, San Francisco.

144. Hopkins CW, Le Grand S, Walker RC, Roitberg AE. Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. J Chem Theory Comput. 2015 Apr 14;11(4):1864–74.

145. The PyMOL Molecular Graphics System, Version 2.4 Schrödinger, LLC.

146. Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, et al. Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. Acc Chem Res. 2000 Dec 1;33(12):889–97.

147. Stenmark H. Rab GTPases as coordinators of vesicle traffic. Nat Rev Mol Cell Biol. 2009 Aug;10(8):513–25.

148. Mukherjee S, Liu X, Arasaki K, McDonough J, Galán JE, Roy CR. Modulation of Rab GTPase function by a protein phosphocholine transferase. Nature. 2011 Sep;477(7362):103–6.

149. Goody PR. Reversible phosphocholination of Rab proteins by Legionella pneumophila effector proteins. 2012;11.

150. Muller MP, Peters H, Blumer J, Blankenfeldt W, Goody RS, Itzen A. The Legionella Effector Protein DrrA AMPylates the Membrane Traffic Regulator Rab1b. Science. 2010 Aug 20;329(5994):946–9.

151. Singh PK, Muqit MMK. Parkinson's: A Disease of Aberrant Vesicle Trafficking. Annu Rev Cell Dev Biol. 2020 Oct 6;36(1):237–64.

152. Vieweg S, Mulholland K, Bräuning B, Kachariya N, Lai Y-C, Toth R, et al. PINK1-dependent phosphorylation of Serine111 within the SF3 motif of Rab GTPases impairs effector interactions and LRRK2-mediated phosphorylation at Threonine72. Biochemical Journal. 2020;18.

153. Valente EM. Hereditary Early-Onset Parkinson's Disease Caused by Mutations in PINK1. Science. 2004 May 21;304(5674):1158–60.

154. Rudack T, Xia F, Schlitter J, Kötting C, Gerwert K. Ras and GTPase-activating protein (GAP) drive GTP into a precatalytic state as revealed by combining FTIR and biomolecular simulations. Proc Natl Acad Sci USA. 2012 Sep 18;109(38):15295–300.

155. Kobayashi C, Saito S. Relation between the Conformational Heterogeneity and Reaction Cycle of Ras: Molecular Simulation of Ras. Biophysical Journal. 2010 Dec;99(11):3726–34.

156. Matsumoto K, Shima F, Muraoka S, Araki M, Hu L, Ijiri Y, et al. Critical Roles of Interactions among Switch I-preceding Residues and between Switch II and Its Neighboring α-Helix in Conformational Dynamics of the GTP-bound Ras Family Small GTPases. Journal of Biological Chemistry. 2011 Apr;286(17):15403–12.

157. Luitz MP. Adenylylation of Tyr77 stabilizes Rab1b GTPase in an active state: A molecular dynamics simulation analysis. Scientific Reports. :11.

158. Kannan S, Zacharias M. Enhanced sampling of peptide and protein conformations using replica exchange simulations with a peptide backbone biasing-potential. Proteins. 2006 Nov 21;66(3):697–706.

159. Ostermeir K, Zacharias M. Hamiltonian replica-exchange simulations with adaptive biasing of peptide backbone and side chain dihedral angles. J Comput Chem. 2014 Jan 15;35(2):150–8.

160. Kumawat A, Chakrabarty S, Kulkarni K. Nucleotide Dependent Switching in Rho GTPase: Conformational Heterogeneity and Competing Molecular Interactions. Sci Rep. 2017 Apr;7(1):45829.

161. Grant BJ, Gorfe AA, McCammon JA. Ras Conformational Switching: Simulating Nucleotide-Dependent Conformational Transitions with Accelerated Molecular Dynamics. Briggs JM, editor. PLoS Comput Biol. 2009 Mar 20;5(3):e1000325.

162. Yu L, Li D-W, Brüschweiler R. Balanced Amino-Acid-Specific Molecular Dynamics Force Field for the Realistic Simulation of Both Folded and Disordered Proteins. J Chem Theory Comput. 2020 Feb 11;16(2):1311–8.

163. Yu L, Li D-W, Brüschweiler R. Systematic Differences between Current Molecular Dynamics Force Fields To Represent Local Properties of Intrinsically Disordered Proteins. J Phys Chem B. 2021 Jan 28;125(3):798–804.

164. Phosphorylation of Ser111 in Rab8a modulates Rabin8 dependent activation by perturbation of side chain interaction networks. :23.

165. Roe DR, Cheatham TE. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. J Chem Theory Comput. 2013 Jul 9;9(7):3084–95.

166. Wu J, Yang J, Cho WC, Zheng Y. Argonaute proteins: Structural features, functions and emerging roles. Journal of Advanced Research. 2020 Jul;24:317–24.

167. Swarts DC, Jore MM, Westra ER, Zhu Y, Janssen JH, Snijders AP, et al. DNA-guided DNA interference by a prokaryotic Argonaute. Nature. 2014 Mar;507(7491):258–61.

168. Hur JK, Olovnikov I, Aravin AA. Prokaryotic Argonautes defend genomes against invasive DNA. Trends in Biochemical Sciences. 2014 Jun;39(6):257–9.

169. Wang Y, Juranek S, Li H, Sheng G, Wardle GS, Tuschl T, et al. Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes. Nature. 2009;461(7265):754–61.

170. Kwak PB, Tomari Y. The N domain of Argonaute drives duplex unwinding during RISC assembly. Nat Struct Mol Biol. 2012 Feb;19(2):145–51.

171. Wang Y, Sheng G, Juranek S, Tuschl T, Patel DJ. Structure of the guide-strand-containing argonaute silencing complex. Nature. 2008 Nov;456(7219):209–13.

172. Wang Y, Juranek S, Li H, Sheng G, Tuschl T, Patel DJ. Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex. Nature. 2008 Dec;456(7224):921–6.

173. Elkayam E, Kuhn C-D, Tocilj A, Haase AD, Greene EM, Hannon GJ, et al. The Structure of Human Argonaute-2 in Complex with miR-20a. Cell. 2012 Jul;150(1):100–10.

174. Sheng G, Zhao H, Wang J, Rao Y, Tian W, Swarts DC, et al. Structure-based cleavage mechanism of Thermus thermophilus Argonaute DNA guide strand-mediated DNA target cleavage. Proceedings of the National Academy of Sciences. 2014 Jan 14;111(2):652–7.

175. Zander A, Holzmeister P, Klose D, Tinnefeld P, Grohmann D. Single-molecule FRET supports the two-state model of Argonaute action. RNA Biology. 2014 Jan 1;11(1):45–56.

176. Jung S-R, Kim E, Hwang W, Shin S, Song J-J, Hohng S. Dynamic anchoring of the 3′-end of the guide strand controls the target dissociation of Argonaute–guide complex. Journal of the American Chemical Society. 2013;135(45):16865–71.

177. Deerberg A, Willkomm S, Restle T. Minimal mechanistic model of siRNA-dependent target RNA slicing by recombinant human Argonaute 2 protein. Proceedings of the National Academy of Sciences. 2013 Oct 29;110(44):17850–5.

178. Kong R, Xu L, Piao L, Zhang D, Hou T-J, Chang S. Exploring the RNA-bound and RNA-free human Argonaute-2 by molecular dynamics simulation method. Chemical Biology & Drug Design. 2017;90(5):753–63.

179. Wang Y, Li Y, Ma Z, Yang W, Ai C. Mechanism of microRNA-target interaction: molecular dynamics simulations and thermodynamics analysis. PLoS Comput Biol. 2010;6(7):e1000866.

180. Zhu L, Jiang H, Sheong FK, Cui X, Gao X, Wang Y, et al. A Flexible Domain-Domain Hinge Promotes an Induced-fit Dominant Mechanism for the Loading of Guide-DNA into Argonaute Protein in *Thermus thermophilus*. J Phys Chem B. 2016 Mar 17;120(10):2709–20.

181. Jiang H, Sheong FK, Zhu L, Gao X, Bernauer J, Huang X. Markov State Models Reveal a Two-Step Mechanism of miRNA Loading into the Human Argonaute Protein: Selective Binding followed by Structural Re-arrangement. Chen S-J, editor. PLoS Comput Biol. 2015 Jul 16;11(7):e1004404.

182. Iwasaki S, Kobayashi M, Yoda M, Sakaguchi Y, Katsuma S, Suzuki T, et al. Hsc70/Hsp90 Chaperone Machinery Mediates ATP-Dependent RISC Loading of Small RNA Duplexes. Molecular Cell. 2010 Jul;39(2):292–9.

183. Kawamata T, Tomari Y. Making RISC. Trends in Biochemical Sciences. 2010 Jul;35(7):368–76.

184. Liu Y, Esyunina D, Olovnikov I, Teplova M, Kulbachinskiy A, Aravin AA, et al. Accommodation of Helical Imperfections in Rhodobacter sphaeroides Argonaute Ternary Complexes with Guide RNA and Target DNA. Cell Reports. 2018 Jul;24(2):453–62.

185. Liu Y, Yu Z, Zhu J, Wang S, Xu D, Han W. Why Is a High Temperature Needed by Thermus thermophilus Argonaute During mRNA Silencing: A Theoretical Study. Front Chem. 2018 Jun 14;6:223.

186. Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen M, et al. Comparative Protein Structure Modeling Using Modeller. Current Protocols in Bioinformatics [Internet]. 2006 Sep [cited 2021 Apr 17];15(1). Available from: https://onlinelibrary.wiley.com/doi/10.1002/0471250953.bi0506s15

187. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. J Chem Theory Comput. 2015 Aug 11;11(8):3696–713.

188. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and testing of a general amber force field. J Comput Chem. 2004 Jul 15;25(9):1157–74.

189. Zgarbová M, Šponer J, Otyepka M, Cheatham TE, Galindo-Murillo R, Jurečka P. Refinement of the Sugar–Phosphate Backbone Torsion Beta for AMBER Force Fields Improves the Description of Z- and B-DNA. J Chem Theory Comput. 2015 Dec 8;11(12):5723–36.

190. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. MACHINE LEARNING IN PYTHON. :6.

191. Pérez A, Luque FJ, Orozco M. Dynamics of B-DNA on the Microsecond Time Scale. J Am Chem Soc. 2007 Nov 28;129(47):14739–45.

192. Lindorff-Larsen K, Piana S, Dror RO, Shaw DE. How Fast-Folding Proteins Fold. Science. 2011 Oct 28;334(6055):517–20.

193. Perilla JR, Schulten K. Physical properties of the HIV-1 capsid from all-atom molecular dynamics simulations. Nature Communications. 2017 Jul 19;8(1):15959.

194. Arantes PR, Saha A, Palermo G. Fighting COVID-19 Using Molecular Dynamics Simulations. ACS Cent Sci. 2020 Oct 28;6(10):1654–6.

195. Kalita P, Padhi AK, Zhang KYJ, Tripathi T. Design of a peptide-based subunit vaccine against novel coronavirus SARS-CoV-2. Microbial Pathogenesis. 2020 Aug 1;145:104236.