

Molecular Dynamics and Interactions of Murine  
Guanylate Binding Proteins and the Bacterial  
Dynamin Like Protein



Heinrich-Heine-Universität Düsseldorf



**Molecular Dynamics and Interactions of Murine  
Guanylate Binding Proteins and the Bacterial  
Dynamin Like Protein**

Inaugural dissertation

presented to the Faculty of Mathematics and Natural Sciences  
of Heinrich-Heine-Universität Düsseldorf

for the degree of

Doctor rerum naturalium (Dr. rer. nat.)

by

**Wibke Schumann**

from Düsseldorf

January 18, 2023

aus dem Institut für Theoretische und Computerchemie  
Arbeitsgruppe Computergestützte Biochemie  
der Heinrich-Heine-Universität Düsseldorf

Gedruckt mit der Genehmigung der  
Mathematisch-Naturwissenschaftlichen Fakultät der  
Heinrich-Heine-Universität Düsseldorf

Berichtersteller:

1. Prof. Dr. Birgit Strodel
2. Prof. Dr. Sander Smits

Tag der mündlichen Prüfung: 10.03.2023

# Abstract

Dynamin superfamily proteins (DSPs) are mechanochemical enzymes working in highly oligomeric and highly cooperative superstructures to deform membranes. Although many static structures of DSPs have been resolved and pictures of the coarse-grained dynamics of membrane reshaping are published, the atomistic detail of the dynamin hinge motion remains obscure. In this work, we characterize the complete hinge motion of a dynamin-superfamily protein for the first time, using both unbiased molecular dynamics simulations and umbrella sampling. The bacterial dynamin like protein BDLP serves as our model protein, since both end states of the hinge motion (termed open and closed) are structurally resolved. Intriguingly, we show that the hinge motion is actually a shear motion. Using weighted histogram analysis, we affix an energetic barrier of 60 kJ/mol to the bacterial dynamin-like protein (BDLP) hinge motion, which is lowered to 30 kJ/mol by GTP binding alone. We explore how the GTP-loading state is communicated between the GTPase domain and stalk via allosteric effects mediated by saltbridges. This could explain the previously known cooperative effect, wherein GTP binding leads to dynamin polymerization, and GTP hydrolysis is facilitated in polymeric dynamins. Aided by coarse-grained simulations, we propose a speculative BDLP cycle of action, in which GTP-binding, membrane-binding and oligomerization stabilize the open conformation. Furthermore, we observe a previously unknown wide-open conformation of BDLP, reminiscent of human dynamin 1.

Guanylate binding proteins (GBPs) are another representant of DSPs, effective against an array of pathogens, among them *Toxoplasma gondii*. Their protective functions require oligomerization, but the actual oligomeric structures have not been resolved yet. The only exceptions are the dimer models of human GBP1 (hGBP1) and hGBP5. We provide dimer models for hGBP1 and the murine GBPs 2 and 7 (mGBP2 and mGBP7), and compare their sequences and dynamics to the monomers, as well as between apo and holo state. While hGBP1 and its close orthologue mGBP2 dimerize via their G domains, mGBP7 shows a variety of possible dimer structures, among them parallel and crossed-stalk motifs. The G domain is only partly involved in mGBP7 dimerization, which provides a rationale why mGBP7, unlike hGBP1 and mGBP2, can dimerize in the absence of GTP. Some of the dimer models for mGBP2, mGBP7 and mGBP9 are backed by data from small angle X-ray scattering (SAXS)

and crosslinking mass spectrometry (XL-MS) from our collaborators. For the mGBP9 dimer, sequential analysis and molecular dynamics led to the discovery of a unique *trans*-interaction of a switch 2 serine residue and GTP, which induces a more stable fit of the dimer. We also transfer our knowledge of the hinge motion of BDLP to GBPs, and again find that dimerization and membrane binding stabilize the open form.

We finally broaden our focus to the macroscopic effects. GBPs, as part of the interferon-stimulated immune response, need to recognize molecular patterns of *T.gondii*, and recruit the autophagy machinery to the pathogen. Among the proteins interacting with GBPs are a surface antigen of *T.gondii*, and several autophagy-related proteins. Using protein-protein docking, and energy decomposition of molecular dynamics, we predict the interacting domains, and support them with bioinformatic sequence analyses.

# Declaration

Ich, Wibke Schumann, versichere an Eides Statt, dass diese Dissertation von mir selbständig und ohne unzulässige fremde Hilfe unter Beachtung der Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf erstellt worden ist.

*Düsseldorf, January 18, 2023*

---

Wibke Schumann



# Contents

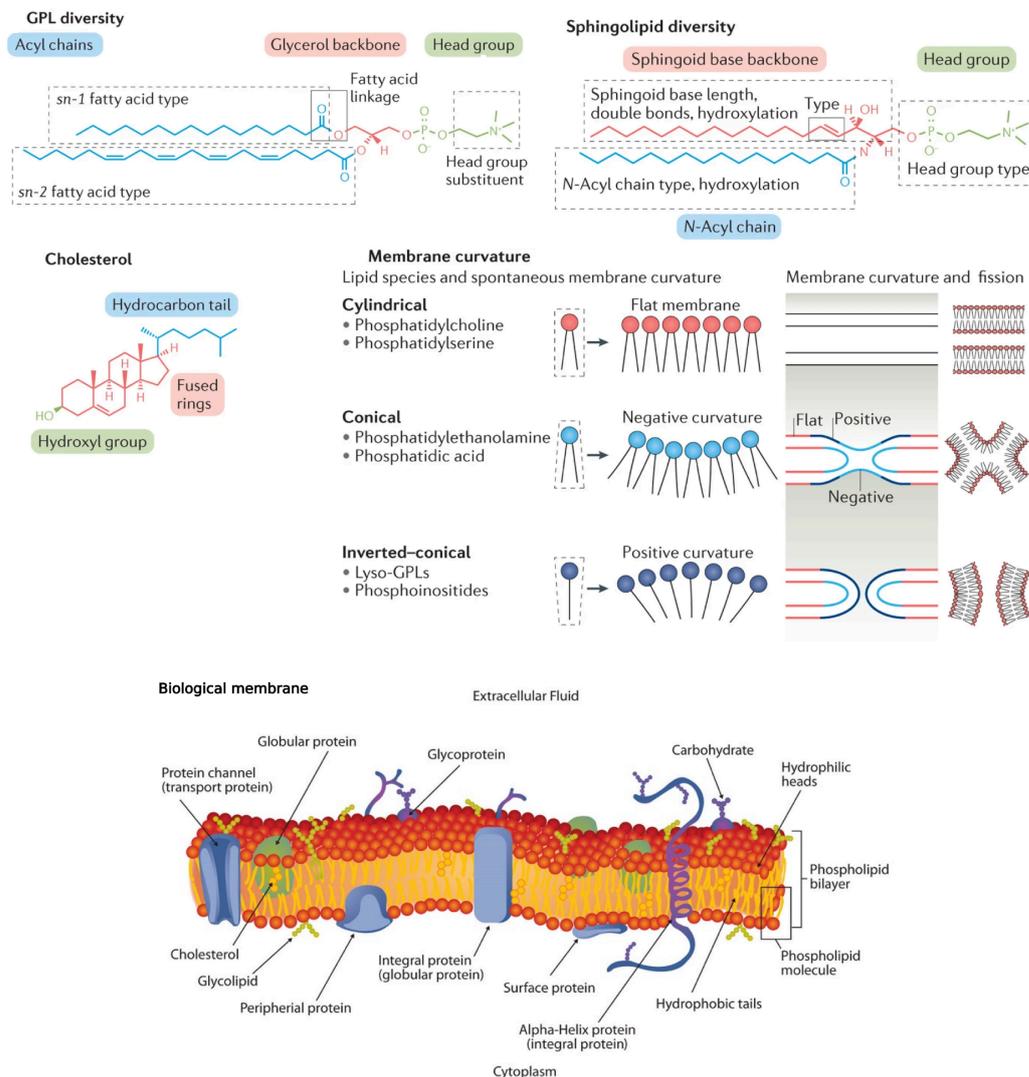
<b>1. Introduction</b>	<b>1</b>
1.1. Lipids and Membranes . . . . .	1
1.2. Membrane Proteins . . . . .	3
1.3. Dynamin Superfamily Proteins . . . . .	4
1.3.1. Energetics of GTP . . . . .	6
1.4. <i>Toxoplasma gondii</i> . . . . .	7
1.5. Mammalian Immune System . . . . .	9
1.6. Research Questions . . . . .	11
<b>2. Methods</b>	<b>13</b>
2.1. Molecular Dynamics . . . . .	13
2.1.1. The Force Field . . . . .	15
2.1.2. The AMBER Force Field . . . . .	17
2.1.3. Newtonian Equation of Motion . . . . .	18
2.1.4. Integration with the Leapfrog Algorithm . . . . .	18
2.1.5. Thermostat and Barostat . . . . .	19
2.1.6. Nonbonded Cutoffs and Periodic Boundary Conditions . . . . .	21
2.1.7. Water Models: TIP3P . . . . .	23
2.1.8. Virtual Sites and Timestep . . . . .	23
2.1.9. Enhanced Sampling . . . . .	23
2.2. Molecular Docking . . . . .	31
2.2.1. Protein-Ligand Docking . . . . .	31
2.2.2. Protein-Protein Docking . . . . .	33
2.3. Normal Mode Analysis . . . . .	34
2.4. Dynamic Cross-Correlation Analysis . . . . .	34
<b>3. Bacterial Dynamin Like Protein</b>	<b>37</b>
3.1. Hinge Motion . . . . .	37
3.2. Membrane Binding . . . . .	39
<b>4. Guanylate Binding Proteins</b>	<b>43</b>
4.1. Dimers . . . . .	43
4.2. Hinge Motion . . . . .	49
4.3. Membrane . . . . .	51

<b>5. Interaction partners</b>	<b>55</b>
5.1. Methods . . . . .	55
5.2. ISG15 . . . . .	56
5.3. SRS29C . . . . .	57
5.4. TOM1 . . . . .	59
5.5. Conclusion . . . . .	62
<b>6. Conclusion</b>	<b>65</b>
<b>References</b>	<b>67</b>
<b>List of Abbreviations</b>	<b>85</b>
<b>List of Figures</b>	<b>91</b>
<b>List of Tables</b>	<b>92</b>
<b>List of Listings</b>	<b>92</b>
<b>A. Publication 1</b>	<b>95</b>
<b>B. Publication 2</b>	<b>135</b>
<b>C. Appendix</b>	<b>217</b>
C.1. Further Manuscripts . . . . .	217
C.2. MD Parameters . . . . .	218
C.2.1. NVT Equilibration . . . . .	218
C.2.2. NpT Equilibration . . . . .	220
C.2.3. Production Run . . . . .	222
C.2.4. HREMD Parameter File . . . . .	224
C.2.5. Pulling and USMD Parameter File . . . . .	226
C.2.6. MARTNI Parameter Files . . . . .	228
C.3. Scripts . . . . .	229
C.3.1. Hinge Motion . . . . .	229
C.3.2. Plumed Pulling . . . . .	232

## 1.1 Lipids and Membranes

Biological membranes fulfill the important role of isolating living organisms from their surroundings, and yet enabling communication. They are inextricably linked with the further characteristics of life: metabolism, homeostasis and reproduction. The simplest prokaryotic organisms have only one encircling plasmamembrane, while eukaryotes (fungi, plants and animals) have additional intracellular compartments, the organelles. Each compartment's lipid composition and shape is adapted to its function [1]–[3]. Hundreds of different lipids can be represented in a membrane, structurally divergent, but unified by their hydrophobicity. Some examples can be seen in Fig. 1.1, as well as a model of a lipid bilayer. Due to the hydrophobic tails and polar head groups, lipids naturally form micelles and bilayers. In a cellular context, this bilayer is asymmetric due to differential lipid composition. For example, the cytoplasm-facing leaflet is enriched in phosphatidylserine, generating a negative charge [4]. Phospholipids (PLs) are the most prominent component of biological membranes, and it consists of two hydrophobic acyl chains - of variable length and saturation - linked by the glycerol backbone to a phosphate with a substitutable polar head group (Fig. 1.1). From the general glycerophospholipid (GPL) the common phosphatidylcholines (PCs) are derived, with a choline headgroup. An example is palmitoyl-oleoyl-phosphatidylcholine (POPC), where the acyl chains are substituted with palmitoyl and oleoyl fatty acid chains. Sphingolipids (Fig. 1.1) are based on a sphingosine backbone, which is an amino alcohol with a long hydrocarbon chain. Its hydroxyl group can be substituted with a head group, and its amino group with a second hydrophobic chain. Both PLs and sphingolipids can be further substituted at the head with saccharides, giving rise to glycopospholipids and glycosphingolipids. These are presented towards the outside extracellular fluid, and are responsible for mechanisms like apoptosis, immune response and blood groups. Another type of lipid, phosphoinositol, whose head groups can be detached from their lipid tails, functions as second messenger within the cell. The more complex, branched gangliosides can be found in the isolating myelin sheath of neurons. Lipopolysaccharides (LPS), on the other hand, are sugars directly linked to the fatty acids, and they are often the toxic agent in gram-negative bacterial infections. Cholesterol has a steroid

backbone with only a simple hydroxyl head group, and mediates membrane fluidity.



**Fig. 1.1.:** Biological lipids and membranes. Examples for lipids: GPL: glycerophospholipid (here: phosphatidyl-ethanolamine), sphingolipid (type can be sphingosine, sphinganine or 4-hydroxy-sphinganine), cholesterol, and their associated membrane curvature. Bottom: Schematic representation of a membrane. Figure composed from [5] and [6], reproduced with permission.

A further important characteristic of membranes is their semipermeability: while hydrophobic molecules (like steroid hormones) can diffuse across, polar ions, or large peptides cannot. Channels and transporters enable passive and active transport (with and against the concentration gradient), with the latter generating electrochemical potentials across membranes. Membranes are not static, but rather fluid, allowing

for diffusion within a layer, a characteristic best described by the fluid-mosaic model [7]. Additionally, a single membraneous compartment can display subcompartments, called lipid rafts, formed by liquid-liquid phase separation and collecting specific lipids and proteins in close proximity [8]–[10].

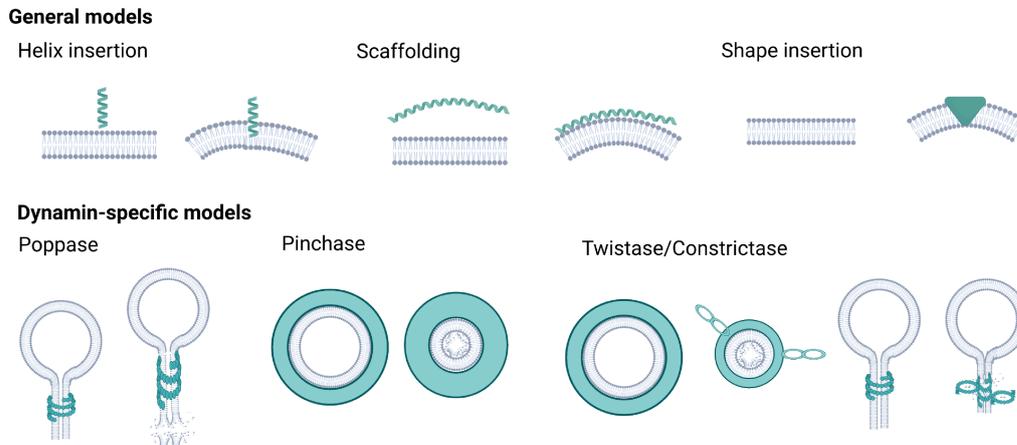
## 1.2 Membrane Proteins

About a third of the human proteome consists of membrane proteins, and they are overrepresented at 50% of all drug targets [11]. Membrane proteins can be integral, like transporters and channels (Fig.1.2), often with membrane-spanning alpha-helices, like the light-sensing rhodopsin, a member of the seven-transmembrane helix receptor class. Alternatively, they can be peripherally or superficially attached, with a membrane anchor, consisting of S-palmitoylation, isoprenylation, glycosyl-phosphatidyl-inositol (GPI), or alpha-helical tails.

Another important characteristic of membranes is their shape, mediated by their curvature. Lipid composition affects curvature, for example, phosphatidyl-ethanolamine's conical shape generates negative curvature (Fig. 1.1). Phosphatidylserine's negative charges can lead to head group repulsion and positive curvature, and additionally recruit Bin/Amphiphysin/Rvs (BAR) domain proteins. These can further function as a scaffold, imposing curvature [12]–[14]. Furthermore, proteins can influence curvature by helix insertion and helix tilt [15]. The cristae of mitochondria have regions of very high curvature, where the Fo/F1-ATPases are gathered [16]. Such an insertion of whole protein shapes is only one example of proteins shaping membranes, as illustrated in Fig.1.2. For the special case of DSPs, the poppase, pinchase and twistase models have been proposed. In the poppase model, the effect is achieved by extending the membrane stalk, thinning the bilayer, which then ruptures, while the pinchase model proposes a tightening of the inner protein layer and membrane stalk, which can then no longer form a bilayer (Fig. 1.2). The twistase works by forming supramolecular winches, which decrease the number of proteins per turn [17]–[19]. The true mechanism of action of DSPs may lie in a compromise of these models. Since DSPs are a main focus of this work, they will be more closely introduced in the following.

In order to deform a flat bilayer, one needs to overcome the lipid-lipid interaction or bending rigidity, which has been quantified at ca.  $20 k_B T$  [13], [20]. When taking the edge-preventing tendency of lipids into account, the bending energy of a vesicle

risers to  $500 k_B T$ . In other words, curves with a 10 nm radius can be achieved with forces in the tens of piconewtons [20]. Stowell et al. also give the force generated by a single dynamin (in a polymer) as 5.5 pN [17]. As we will see in the next section, this is a common order of magnitude for cellular forces.



**Fig. 1.2.:** Schematic representations of membrane shaping mechanisms. Top: More general cases: helix insertion, scaffolding and shape insertion. Bottom: Dynamin-specific poppase, pinchase and twistase/constrictase models. Figure created with Biorender.

## 1.3 Dynamin Superfamily Proteins

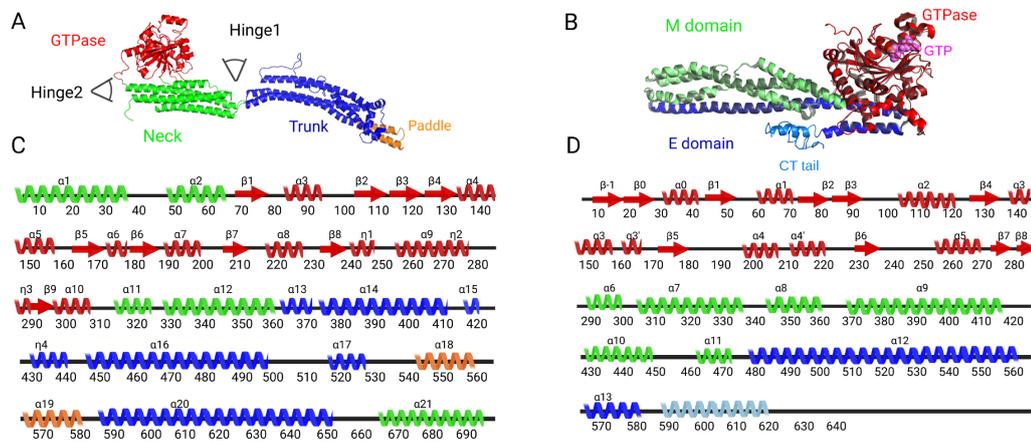
DSPs are mechanochemical enzymes, involved in critical cellular functions like endocytosis, cell division and immune response [21]–[24]. Next to flagellar proteins, they generate some of the highest torques known for proteins, in the range of a thousand piconewton-nanometers (or a few attojoule) [14], [21].

As shown in Fig. 1.3, DSPs have a common domain structure, here exemplified by the BDLP and murine guanylate binding protein 7 (mGBP7), which will be studied in this work. Their GTPase domain (GD) can easily be identified from its structural motifs (G1-G5), and fuels the mechanical mechanism by hydrolyzing GTP. G1, also called P-loop, is a GxxxxGKS/T sequence binding the phosphates of nucleotides. G2 or switch I (SWI) consists of a threonine, which binds  $Mg^{2+}$  and the  $\gamma$ -phosphate, together with the DxxG motif of switch II (G3). The G4 motif (N/TK/RxD) and G5 motif (or guanine cap) bind the nucleobase. A structural example of the GTP binding site will later be shown in Sec. 4.1.

The stalk domains are elongated and helical, and their nomenclature is not unified, owing to the separate discovery of dynamin family members. When there are two

helical domains, they can be referred to as "neck" and "trunk", separated by hinge1. Another common name for the neck is bundle signalling element (BSE). Hinge2 separates the GD from the "neck", and is the only hinge present in smaller DSPs (like atlastin). Its rotational state depends on the nucleotide loading (GTP vs. GDP). Guanylate binding proteins use the terms middle domain (M domain) and effector domain (E domain) instead, and it is not clear yet where hinge1 and hinge2 are located.

Membrane binding is mediated by different domains, depending on the DSP family representative. In human dynamins, a pleckstrin homology domain fulfills that role, while the yeast dynamin related protein 1 (Drp1) uses a B-insert. In BDLP the transmembrane helix is called "paddle", and guanylate binding proteins have either a C-terminal CaaX motif which gets isoprenylated, or a C-terminal transmembrane helix [25], [26].



**Fig. 1.3.:** Domain structure of dynamin superfamily proteins. A: BDLP (modeled after PDB ID 2W6D). B: mGBP7, as modeled after hGBP1 (PDB ID 1F5N). C: secondary structure sequence of BDLP. D: secondary structure sequence of mGBP7. Figure created with Biorender.

Unlike the smaller Ras-like GTPases, DSPs are guanine nucleotide-binding proteins (G proteins) activated by nucleotide-dependent dimerization. This means that the function of the otherwise accessory GTPase activating proteins (GAPs) is encoded in the dimer instead. The guanine nucleotide exchange factor (GEF) protein is unnecessary, because DSPs bind GTP with low affinity ( $K_m = 15 \mu\text{M}$ ) and hydrolyze it with a high basal rate ( $K_{cat} = 0.3 \text{ s}^{-1}$ ), which is stimulated up to 1000-fold upon membrane bound polymerization [26]. DSP polymers are highly ordered helices, which tubulate budding vesicles and contract under GTP addition. Nonhydrolyzable GTP analogues lead to elongated membrane stalks, proving that hydrolysis is what

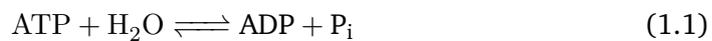
finally drives fission [27] (see also Sec. 3.2 for a generalized reaction cycle). Polymerization is generally driven by interactions of the helical domain, though the GD can also be involved [27], [28].

DSPs have various places of action in the cell: dynamin itself is involved in endocytosis, golgi vesiculation, and actin bundling, while Drp1, mitofusins, fuzzy onion protein 1 (Fzo1p), optic atrophy 1 (OPA1) and mitochondrial genome maintenance protein 1 (Mgm1p) are responsible for mitochondrial plasticity, atlastin for ER fusion, and vacuole sorting protein 1 (Vpsp1) for vacuole fusion and fission. A little apart are the immunity functions of myxovirus resistance protein A (MxA), MxB and GBPs, cytokinesis functions of BDLP and DynA, and toxin secretion by labile enterotoxin output (Leo)ABC [25], [29].

Unsurprisingly, such widely involved proteins can cause an array of diseases, among them microcephaly, Charcot-Marie Tooth disease and centronuclear myopathy [30].

### 1.3.1 Energetics of GTP

The existence of GTPases, next to ATPases, raises the question: Why do cells maintain two currencies? ATP is used widely as an energetic currency, and is the main product of glycolysis. Its dominance might be due to chance, but it has also been suggested that its synthesis from five molecules of HCN in abiotic conditions kickstarted its importance [31]. On the other hand, GTP has important roles in protein synthesis, signaling, but also in fueling dynamins. GTP has a similar energetic value to ATP [32], and the main difference may lie in protein conglomerates channeling the required nucleotide to the specific process. As such, dynamins may have evolved together with GTP producing partners [32]. In order to understand the energy generated, let us consider ATP hydrolysis:



where the first reaction converts ATP to ADP and inorganic phosphate, and the second ATP to AMP and pyrophosphate. The Gibbs free energies of these reactions are generally given as

$$\Delta G = -RT \ln(K) + RT \ln \left( \frac{[\text{ADP}] * [\text{P}_i]}{[\text{ATP}]} \right) = -30.5 \text{ kJ/mol} \quad (1.3)$$

$$\Delta G = -RT \ln(K) + RT \ln \left( \frac{[\text{AMP}] * [\text{PP}_i]}{[\text{ATP}]} \right) = -45.6 \text{ kJ/mol} \quad (1.4)$$

with conditions as  $K = 10^5$  M (obtained from equilibrium of ATP hydrolysis), pH 7, 10 mM  $\text{Mg}^{2+}$ , 10 mM  $\text{P}_i$ , and the displaced "standard concentrations" of  $\frac{[\text{ADP}] * [\text{P}_i]}{[\text{ATP}]} = 1$  M, which omits a very important fact: conditions in the cell are not standard conditions. In fact, cells maintain ratios of  $\frac{[\text{ADP}] * [\text{P}_i]}{[\text{ATP}]} = 10^{-3}$  M- $10^{-5}$  M which displaces the equilibrium by 10 orders of magnitude, and the free energy of  $\text{ATP} \rightarrow \text{ADP}$  hydrolysis is brought to -57 kJ/mol [33, p.36] [34, p.454]. This value varies by cell type and by cell compartment, so that slightly different numbers may be found in the literature.

It is important to note that the guanylate binding proteins studied in this work do not hydrolyze GTP to GMP directly, as indicated in Eq. 1.2, but dephosphorylate GTP in two successive steps.

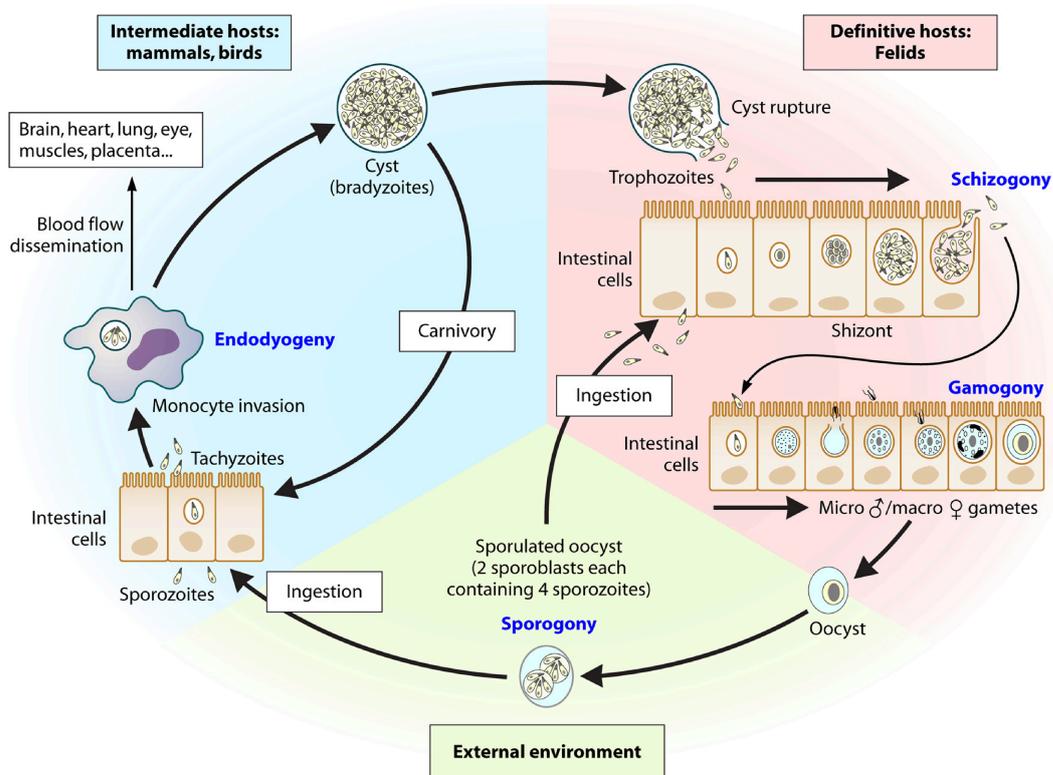
## 1.4 *Toxoplasma gondii*

The previously mentioned GBPs from the dynamin superfamily are a medically relevant class of proteins due to their action against a wide array of intracellular pathogens, such as *Chlamydia* [35], *Shigella flexneri* [36], zika virus, measles, influenza A [37], [38], cancer [39]–[41], hepatitis C [42], human immune deficiency virus (HIV) [43], [44], *Legionella* [45], *Francisella* [46], *Leishmania* [47] and *Burkholderia* [48]. For this work, their action against *Toxoplasma gondii* was chosen as the main focus.

*T. gondii* is an obligate intracellular parasite, belonging to the monocellular eukaryotes (protozoans), more specifically, to the apicomplexa. Its name was coined due to its bow-shaped form and first described isolation from gundi rodents [49]. Its life cycle includes sexual replication phases in the main host (cats) and asexual replication in almost all warmblooded vertebrates (intermediate hosts), as well as a phase where the parasite is exposed to the external environment [50], [51]. The slightly different morphology of the parasite during these phases has led to the names tachyzoites, bradyzoites, merozoites and sporozoites (Fig. 1.4).

A cat ingesting an infected mouse will have its intestinal epithelium invaded by the asexually replicating bradyzoites (Fig. 1.4). After self-limiting multiplication, the bradyzoites develop into tachyzoites, then merozoites. This asexual schizogony is followed by gamogony, the formation of female macrogametes and male microgametes. Their fusion yields the diploid zygote, which turns into the oocyste, and

is then excreted with the cat feces. Subsequent sporulation yields haploid sporozoites, which are infective for cats and intermediate hosts over a long time. Smear infection and subsequent hand-to-mouth contact transport the sporozoites to the intestines of another host. Once there, they invade epithelial cells again and develop into tachyzoites. These rapidly replicate via endodyogeny, filling the host cell up, until controlled lytic rupture occurs. The parasite can then spread to brain, eye and muscle tissue via the blood. Subsequent host immune response pressures the conversion to bradyzoites, the persistent, slow-growing infection stage, protected by a cyst shell [52]. Insufficiently cooked meat of infected animals can lead to infection of other intermediate hosts (e.g. humans), or the closing of the cycle to the main host (Fig. 1.4).



**Fig. 1.4.:** *T. gondii* life cycle: Morphology and replication for each of the three developmental stages (cat, environment and intermediate host) are shown. Figure taken from Gangneux et al. [52], reproduced with permission.

*T. gondii*, as a eukaryote, is in possession of organelles, like a nucleus, endoplasmic reticulum, Golgi, and a single mitochondrion [53]. Additionally, the name-giving apical complex, consisting of cytoskeleton and secretory organelles (rhoptries, micronemes) is essential for *T. gondii* survival [54]. After adhesion to the host cell using its surface antigens, *T. gondii* penetrates the cell using proteins secreted by

the apical complex, which leads to its inclusion in the parasitophorous vacuole membrane (PVM). The PVM is derived from the host plasma membrane, but fusion with the endocytotic or autophagic machinery is prevented, supplying the parasite with a protected niche for replication [55]. *T. gondii* supplies GPI-anchored proteins and pores to the PVM, and recruits host organelles to ensure nutrient flow. The apicoplast is another specialized organelle, which originates from secondary endosymbiosis of an alga and is thus surrounded by four membranes. Its genome has vital functions in apicomplexan metabolism and can be safely targeted by antibiotics without consequence to the host metabolism [56]. During the phases where *T. gondii* is exposed to the external environment, it is enclosed by a shell, the so-called pellicle, which is readily dissolved by gastric enzymes, freeing the infectious cells [57].

An estimated third of the world population, and half of the German population is infected by *T. gondii* during their life time [58]. An acute infection during pregnancy can be transplacentally transmitted to the fetus, leading to miscarriage or birth defects [59]. Immunocompromised individuals (like AIDS or transplantation patients) with an untreated *T. gondii* infection can die within weeks from toxoplasmosis encephalitis. Healthy individuals on the other hand will only experience flu-like symptoms [60], [61].

## 1.5 Mammalian Immune System

The reason for an uncomplicated course of disease as described in the previous section is the human immune system. It is commonly divided into innate (unspecific) immunity and adaptive (specific) immunity.

Innate immunity includes simple barriers like our skin, but also cellular responses from phagocytes (monocytes, macrophages, granulocytes and dendritic cells) and natural killer cell (NK cells) cells. Germline-encoded pattern recognition receptors (PRRs) on the surface of these cells can recognize pathogen associated molecular patterns (PAMPs), and then secrete reactive oxygen species or engulf the pathogenic cells. The latter, in turn, leads to secretion of chemokines and cytokines, activating further immune responses. In contrast to the large cellular components, the smaller chemokines, cytokines, acute phase proteins and the complement proteins are called humoral factors. In a positive feedback loop, these then enhance recruitment of inflammatory cells and phagocytosis [62, chapter Immunity].

On the other hand, adaptive immunity is a cell-scale weaponized evolution process.

B-cell receptors (BCRs) and T-cell receptors (TCRs) have an infinite protein variability due to somatic recombination of their hypervariable genetic segments. After culling of autoimmune-reactive cells in the thymus, the remaining B- and T-cells circulate in the lymphatic system. When a phagocyte presents the random fragments left over from a digested pathogen, and such an antigen matches one of the random antibodies presented on different T-cells, a clonal expansion of the effective cell is instigated. That way, the most fitting cells for defense against the pathogen increase in numbers [62, chapter Immunity].

T-cells can then differentiate to killer T-cells, which induce apoptosis in the pathogen, or helper T-cells, which secrete interleukins (ILs) and interferons (IFNs). B-cells are activated by the native antigen, and secrete soluble antibodies, which then agglutinate the pathogens, and further activate the innate immune response. The persistence of some B-cells as memory B-cells helps to respond quicker upon secondary infection. While adaptive immunity can overcome pathogens that evade innate immunity, it needs more time to take effect [62, chapter Immunity].

In the case of *T. gondii*, toll-like receptors recognize the parasitic profilin and GPI-anchored proteins, and induce IL-12 and IFN $\gamma$  secretion, while the adaptive immune system may recognize surface antigens (SAGs) [63]. IFN $\gamma$  stimulates removal of tachyzoites by NK cells and T-cells before invasion of the host cell [64], [65]. After invasion, and especially during chronic infection, several metabolic reactions to IFN $\gamma$  signals can starve the parasite by cutting off essential nutrients [66], [67]. IFN $\gamma$  also intracellularly induces the previously mentioned dynamin superfamily proteins, immunity related GTPase (IRGs) and GBPs. IRGs and GBPs recruit best to the PVM in tandem [68]–[70], and GBP complexes forming on the PVM lead to its rupture [70]–[74]. The specifics of this process are unknown, and will in part be addressed in this work.

Autophagy is the recycling process which allows cells to remove dysfunctional components via lysosomal degradation, and it plays an additional role in defense against *T. gondii*. Autophagy proteins (ATGs) are recruited to the defective component, followed by recruitment of ubiquitin ligases. A membrane derived from the endoplasmic reticulum, the phagophore, engulfs the faulty organelle, forming the autophagosome, which then fuses with a lysosome, whose acidic pH and hydrolytic enzymes degrade the contents. ATGs enhance recruitment of IRGs and GBPs, but a classic accumulation of autophagic membranes follows the GBP-induced rupture of the PVM. Thus, it seems that both processes are intertwined in defense against the

parasite [75]–[80].

The final line of defense against a pathogen is programmed cell death, apoptosis. *T. gondii* has evolved several ways to influence this process, possibly to prevent, but maybe also to induce it, in order to escape and infect new cells [81].

## 1.6 Research Questions

Apart from extensive experimental studies, membrane fusion/fission has lately been an object of coarse grained (CG) simulations [82]–[87]. Except for our all-atom work on GBPs [88], [89], dynamins have been little simulated in atomic detail. Experimental observation of the DSP hinge motion has been made by Chen et al. [90], finding that MxA assumes an open conformation (hinge2) in nucleotide free or GDP-bound state, and GTP triggers closing. The dwell times in each state appear in the range of ca. 1 s. Vöpel et. al observed that hGBP1 dimerizes in the presence of GTP, which triggers a rearrangement of the  $\alpha$ 13 helices [91]. We will use the available structures of *Nostoc punctiforme* BDLP as starting points, in order to characterize the complete dynamin hinge motion in all-atom detail for the first time, using both unbiased molecular dynamics (MD) simulations and umbrella sampling molecular dynamics (USMD) simulations. BDLP was chosen because it is the only DSP where both open and closed conformation are experimentally resolved. We will then use BDLP as a proxy for our proteins of interest, the GBPs. As part of the Collaborative Research Center 1208, especially the interaction of BDLP and GBP polymers with membranes is of interest. This will lead to a greater understanding of how the dynamin hinge motion causes membrane stalk formation, and thus membrane fusion, fission and rupture. Our goals are thus to:

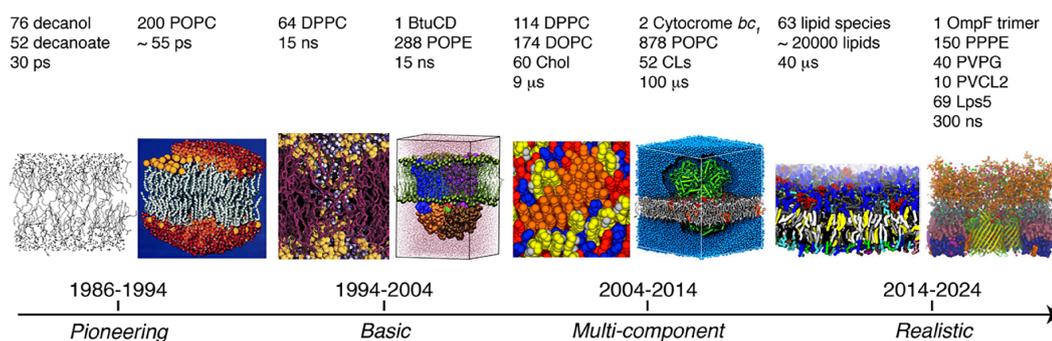
1. elucidate the complete dynamin-like hinge motion for BDLP and transfer it to GBPs,
2. develop models for GBP homo- and heterocomplexes,
3. and finally, to develop models of membrane damage.

This cumulative thesis is structured into the chapters "Bacterial Dynamin-Like Protein (BDLP)", containing the first publication, "Guanylate Binding Proteins (GBPs)", containing the second and third publication, and "Interaction Partners of mGBP7", an ongoing project, followed by an overarching conclusion.



## 2.1 Molecular Dynamics

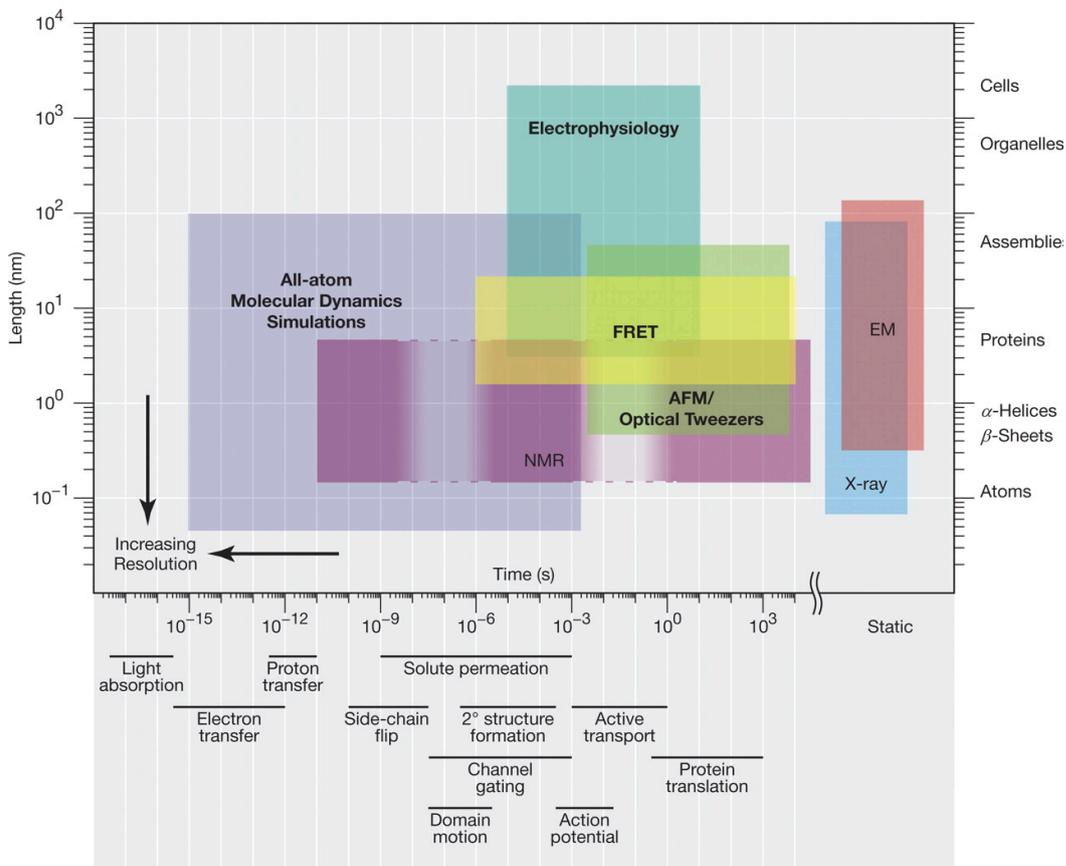
The Nobel Prize in Chemistry 2013 was awarded jointly to Martin Karplus, Michael Levitt and Arieh Warshel "for the development of multiscale models for complex chemical systems". The groundwork for this technique was laid in the 1970s with simulations of retinal, while the first simulation of a protein was achieved in 1976 [92]. The initial simulations of small proteins, simulated for picoseconds in vacuum, have since increased to large protein-membrane complexes, or even cellular-scale simulations, and the timescale can be counted in milliseconds by now [93], see also Fig. 1.1.



**Fig. 2.1.:** Scope of membrane simulations over time: From left to right, the increase of membrane system size, complexity and simulation time are shown. Abbreviations: POPC, 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine; DPPC, 1,2-dipalmitoyl- sn-glycero-3-phosphocholine; POPE; 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine; DOPC, 1,2-dioleoyl-sn-glycero-3-phosphocholine; Chol, cholesterol; CLs, cardiolipins; PPPE, 1-palmitoyl-2-palmitoleoyl-phosphatidylethanolamine; PVPG, 1-palmitoyl-2-vacenoyleoyl-phosphatidyl- glycerol; PVCL2, 1,10-palmitoyl-2,20-vacenoyleoyl cardiolipin; Lps5, *E.coli* R1 lipopolysaccharide core with repeating units of O6-antigen. Figure taken from Marrink et al. [94], reproduced with permission.

Protein structures are often elucidated by X-ray crystallography, which results in static snapshots of, in reality, flexible molecules. Cryo-electron microscopy (cryo-EM) and nuclear magnetic resonance (NMR) allow for the resolution of some motions, but are not always applicable. The cryo-EM size range encompasses proteins larger than

50 kDa, while this is the upper limit for NMR spectroscopy [95], [96]. Molecular dynamics simulation is an important tool over a variety of length- and timescales, where experimental biophysics methods leave gaps, as illustrated in Fig.2.2.



**Fig. 2.2.:** Biophysical methods and their temporal and spatial scope, taken from Dror et al. [97], reproduced with permission.

Modern software packages like GROMACS [98]–[104] or CHARMM-GUI [105]–[117], facilitate the setup of molecular dynamics simulations for the user. The Groningen machine for chemical simulations (GROMACS) contains linear methods which deal with the initial setup or final analysis of the system, and parallelized methods (originally written in C, with inner loops in Fortran, now moved to C++, with explicit message passing by MPI and OpenMP), which deal with the production MD runs.

The following sections "The Force Field", "Integration with the Leapfrog Algorithm" and "Replica Exchange MD" are taken largely unchanged from my master's thesis [118].

## 2.1.1 The Force Field

In simple terms, molecular dynamics works by applying physical forces to atoms, resulting in complex motions. These forces use simple classical mechanics, where a force  $\mathbf{F}$  is described as the derivative of an energy  $U$  by the location  $\mathbf{r}$ :

$$\mathbf{F} = -\frac{\partial U}{\partial \mathbf{r}} \quad (2.1)$$

where  $U$  is a function of the positions of all  $N$  atoms in the system:

$$U = f(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) \quad (2.2)$$

More specifically,  $U$  depends on  $(N - 1)$  bonds,  $(N - 2)$  angles,  $(N - 3)$  dihedral angles, and internal non-bonded interactions, as well as possibly applied external forces.

For pure molecular dynamics, a harmonic potential

$$U(r_{ij}) = \frac{k_{vib}}{2}(r_{ij} - r_0)^2 \quad (2.3)$$

which makes the bond unbreakable, is sufficient. More complicated applications such as quantum mechanics/molecular mechanics (QM/MM) allow for the breaking of bonds by introducing a small area that is modeled after quantum mechanical laws. Above,  $k_{vib}$  is the spring constant of the harmonic potential, and  $r_0$  is the equilibrium bond length.

The bond angles  $\theta_{ijk}$  can be similarly described by a harmonic potential, where  $k_{bend}$  is again the spring constant, and  $\theta_0$  the equilibrium angle:

$$U(\theta_{ijk}) = k_{bend}(\theta_{ijk} - \theta_0)^2 \quad (2.4)$$

Torsion angles  $\phi_{ijkl}$  can be described by the following potential:

$$U(\phi_{ijkl}) = \sum_n \frac{V_n}{2}(1 + \cos(n\phi_{ijkl} - \phi_0)) \quad (2.5)$$

where  $V_n$  is the height of the energy barrier and  $n$  the periodicity of the torsion, and  $\phi_0$  the equilibrium angle.

Improper torsion, which prevents out-of-plane-bending in planar systems (like aromatics or delocalised double bonds), is described as follows:

$$U(\omega_{ijkl}) = \frac{k_{imp}}{2}(\omega_{ijkl} - \omega_0)^2 \quad (2.6)$$

where  $\omega$  is the "improper angle" between the four atoms  $i$ ,  $j$ ,  $k$  and  $l$ ,  $\omega_0$  its equilibrium value, and  $k_{imp}$  the spring constant.

On the other hand, electrostatic potentials belong to the non-bonded interactions and are described by the following formula:

$$U(r) = \frac{q_i q_j}{4\pi\epsilon_0\epsilon_r r_{ij}} \quad (2.7)$$

where  $r$  is the distance between the two charges  $q_i$  and  $q_j$  of the interacting atoms.  $\epsilon_0$  is the vacuum permittivity, and  $\epsilon_r$  is a material-based, relative permittivity, which takes a value  $>1$  for almost all practical purposes and the exact value 1 for MD simulations with explicit solvent. Coulomb interactions are long-ranged and therefore care has to be taken when calculating them in combination with cutoff schemes that are usually employed to reduce computational cost. In periodic systems the the Particle-Mesh-Ewald method [119] is typically used for the calculation of the Coulomb interactions as it takes care of its long-range effects (see also Sec.2.1.6).

The Lennard-Jones potential is the second non-bonded interaction, describing Pauli-repulsion and van-der-Waals attraction between atoms. It has a shorter range than Coulomb interactions and is described as follows:

$$U(r) = U_m \left[ \left( \frac{r_{ij}^{vdw}}{r_{ij}} \right)^{12} - 2 \left( \frac{r_{ij}^{vdw}}{r_{ij}} \right)^6 \right] \quad (2.8)$$

where  $r_{ij}^{vdw}$  is the combination of the van-der-Waals radii of atoms  $i$  and  $j$  (a sum in AMBER and CHARMM) and  $r$  is the distance between atoms  $i$  and  $j$ .  $U_m$  (often referred to as well-depth  $\epsilon$ ) is the potential of the minimum at  $r_{ij}^{vdw}$ .

All previously mentioned equilibrium lengths/angles can be taken from theoretical or experimental studies (physics-based force fields) or from averaging over many known structures (knowledge-based force fields). This parametrization of force fields is essential for its realistic description of molecules and subject to constant verification.

Finally, all terms are then summed to form the force field:

$$\begin{aligned}
U(\mathbf{r}) = & \sum_{bonds} \frac{k_{vib}}{2} (r_{ij} - r_0)^2 + \sum_{angles} k_{bend} (\theta_{ijk} - \theta_0)^2 \\
& + \sum_{dihedrals} \frac{V_n}{2} (1 + \cos(n\phi_{ijkl} - \phi_0)) + \sum_{improper} \frac{k_{imp}}{2} (\omega_{ijkl} - \omega_0)^2 \quad (2.9) \\
& + \sum_i \sum_j \frac{q_i q_j}{4\pi\epsilon_0\epsilon_r r} + \sum_i \sum_j U_m \left[ \left( \frac{r_{ij}^{vdw}}{r} \right)^{12} - 2 \left( \frac{r_{ij}^{vdw}}{r} \right)^6 \right]
\end{aligned}$$

This makes the important but justifiable assumption of independence of these terms. Different force fields may differ in the parameters used here, which together with charges and water model, attempt to reproduce reality.

### 2.1.2 The AMBER Force Field

The first all-atom assisted model building with energy refinement (AMBER) force field was published in 1986 [120] and intended for protein and nucleic acid simulations, parametrized with a mix of empirical and quantum mechanics data. Charges are assigned using the restrained electrostatic surface potential method. Simulation outcomes are validated with crystallographic data, free energies of solvation and quantum mechanics (QM) calculations [121], [122].

Most AMBER force fields today are still based on the version developed in 1999. Misleadingly, AMBER99SB\*-ILDNP-Q, which is used in this work, has been updated as recently as 2012. [123]. Its modifications include better phi and psi torsions (denoted by the SB\*), chi dihedral revisions for the amino acids isoleucine, leucine, aspartate and asparagine (denoted by the ILDN) and better partial charges for charged amino acids (denoted with Q).

Amber is also the name for the suite of programs around the AMBER forcefields, which helps set up and analyse the simulations. The forcefields can however, be integrated into other software suites, like GROMACS. [124] Carbohydrates are handled by the GLYCAM forcefield in Amber [125]. Lipids can be handled by the Lipids14 (or later versions) [126], although CHARMM-GUI offers a greater selection of molecules for CHARMM forcefields than for AMBER forcefields, which will become important when we turn to complex glycolipids in the future. Both CHARMM and AMBER are well suited for biomolecules, but CHARMM has an additional Urey-Bradley correction term for better agreement with infrared spectra and a 2D dihedral energy grid correction map (CMAP) term, which is popular for intrinsically disordered proteins. When backmapping our coarse-grained MARTINI membrane systems, a

switch to Amber19 had to be made, due to the selection offered by CHARMM-GUI. In Amber19 the backbone profiles of all standard amino acids are overhauled with a CMAP approach, and it is better compatible with water models other than three-site transferrable intermolecular potential (TIP3P) [127].

### 2.1.3 Newtonian Equation of Motion

The second Newtonian equation correlates force  $\mathbf{F}$  and acceleration  $\mathbf{a}$  with each other:

$$\mathbf{F}_i = m_i \mathbf{a}_i = \frac{d\mathbf{p}_i}{dt} = m \frac{d^2 \mathbf{r}_i}{dt^2} \quad (2.10)$$

Note that this equation is given for a single atom  $i$ , which has three coordinates ( $xyz$ ), resulting in vector notation. As per equation (2.1),  $\mathbf{F}$  can be described by the force field introduced in section 2.1.1. The mass  $m$  of all atoms is known. However, for  $N$  atoms,  $3N$  coupled differential equations have to be solved numerically. This yields a trajectory with (vectorized) positions  $\mathbf{r}$  for every atom at every time  $t$ .

### 2.1.4 Integration with the Leapfrog Algorithm

In order to calculate velocities of atoms, different numeric integration algorithms can be used. In this work, the leapfrog algorithm will be used as implemented in GROMACS. It has the advantage of being fast with low long-term energy drift. It is derived from the Verlet algorithm, which in turn is based on Taylor expansions of  $\mathbf{r}(t \pm \Delta t)$ . The difference is that the leapfrog algorithm calculates  $\mathbf{r}(t)$  at whole time steps and  $\mathbf{v}(t)$  at half time steps:

$$\mathbf{v}\left(t + \frac{\Delta t}{2}\right) = \mathbf{v}\left(t - \frac{\Delta t}{2}\right) + \Delta t \frac{\mathbf{F}(t)}{m} + \mathcal{O}(\Delta t^3) \quad (2.11)$$

In this case,  $\mathcal{O}(\Delta t^3)$  is the error. Since it scales with  $\Delta t$ , small time steps (ca. 1 fs) have to be used. Once the new velocities have been calculated, the new positions can be determined:

$$\mathbf{r}(t + \Delta t) \approx \mathbf{r}(t) + \Delta t \mathbf{v}\left(t + \frac{\Delta t}{2}\right) \quad (2.12)$$

Another disadvantage of the leapfrog algorithm is that kinetic energies can only be calculated from half step velocities and are therefore slightly inaccurate. The potential energy, on the other hand, is calculated at whole time steps, resulting in a skewed total energy. However, for nearly all production simulations the leapfrog integrator is accurate enough [128]. In Alg. 1, a broad overview of the GROMACS algorithm is given. The computation of energies, application of thermostat and writing of output can all have different frequencies, not necessarily identical with the timestep.

### Algorithm 1: Pseudocode of GROMACS

**Data:**  $\mathbf{v}(t_0), \mathbf{v}(t_0 - \frac{\Delta t}{2}), \mathbf{r}(t_0), U(\mathbf{r})$   
**Result:**  $\mathbf{a}(t), \mathbf{v}(t), \mathbf{r}(t), U(\mathbf{r})$  for all timesteps  
**while** *timesteps left* **do**

1. pair list  $\leftarrow$  neighbor search;
2. compute nonbonded forces and bonded forces;
3. **if applicable then** compute energies, pressure, temperature, apply thermostat and barostat;
- ;
4. update  $\mathbf{r}(t)$  and  $\mathbf{v}(t)$ , apply LINCS;
5. **if applicable then** write output;
- ;
- go to 1.;

## 2.1.5 Thermostat and Barostat

Simulations have moved from the canonical (NVT) ensemble to the isothermic-isobaric (NpT) ensemble, to better describe *in vitro* conditions. To ensure that the correct average temperature is maintained, a thermostat is employed. During initialization, velocities of all particles are generated according to the equipartition theorem:

$$\frac{m_i}{2} \langle v_{i,a}^2 \rangle = \frac{k_B T}{2} \quad (2.13)$$

where  $i$  is still the atom index and  $a$  the coordinate index, so that  $v$  is no longer vectorized here. This means that kinetic energy is shared equally amongst all accessible quadratic degrees of freedom of the system. The velocity rescale thermostat below is the simplest implementation of this concept:

$$\begin{aligned}
v_{i,a} &= \text{rand}[-0.5, 0.5] \\
\sum v_{i,a} &= 0 \\
v_{new} &= \sqrt{\frac{T_{target}}{T(t)}} v_{old}
\end{aligned} \tag{2.14}$$

meaning that the  $xyz$  velocities of each particle are randomly drawn, fulfilling only the constraint that their overall sum is zero, resulting in a temperature  $T(t)$  calculated from those velocities. The velocities are then rescaled with  $\sqrt{\frac{T_{target}}{T(t)}}$  with the frequency set for the thermostat. After equilibration, the velocities converge towards a Maxwell-Boltzmann distribution, corresponding to the target temperature:

$$f(\mathbf{v}) = 4\pi \left( \frac{m}{2\pi k_B T} \right)^{\frac{3}{2}} |\mathbf{v}|^2 e^{-\frac{m|\mathbf{v}|^2}{2k_B T}} \tag{2.15}$$

For the production run, the more accurate Nosé-Hoover thermostat is applied, which is canonic, deterministic and time-reversible [129], [130]. The conserved quantity of the equation of motion,  $H_{NH}$ , is no longer the total energy but:

$$\begin{aligned}
H_{NH} = & \underbrace{\sum_i N \frac{\mathbf{p}_i^2}{2m_i}}_{E_{kin}} + \underbrace{U(\mathbf{r})}_{E_{pot}} + \underbrace{\frac{\zeta^2 Q}{2}}_{E_{kin}} + \underbrace{3Nk_B T \zeta}_{E_{pot}} \\
& \underbrace{\hspace{10em}}_{system} \qquad \underbrace{\hspace{10em}}_{reservoir}
\end{aligned} \tag{2.16}$$

where  $\zeta$  is a friction coefficient and  $Q$  the mass parameter of the external heat bath (determining coupling strength). The equation of motion for  $\zeta$  is:

$$\frac{dp\zeta}{dt} = (T(t) - T_{target}) \tag{2.17}$$

Disadvantages of the Nosé-Hoover thermostat include its slow relaxation time and nonergodicity (not the entire phase space is sampled). A possible solution is a nested heat bath construct, called Nosé-Hoover chain [131].

In this work, a combination of Nosé-Hoover thermostat and Parrinello-Rahman barostat is generally employed for production runs, while equilibrations may use the simpler v-rescale thermostat (2.14) and Berendsen barostat [132]. The latter is characterized by:

$$\frac{dP}{dt} = \frac{P_{target} - P(t)}{\tau_p} \quad (2.18)$$

where at each step the MD cell volume is scaled by by a factor  $\eta$  and the coordinates (and cell vectors) by its cubed root:

$$\eta(t) = 1 - \frac{\Delta t}{\tau_p} \gamma (P_{target} - P(t)) \quad (2.19)$$

Here,  $\gamma$  is the isothermal compressibility of the system, and  $\tau$  the rise-time or coupling efficiency to the external pressure bath. In theory, the Parrinello-Rahman barostat [133], [134] yields a true NpT ensemble, and is similar to the Nosé-Hoover approach. Anisotropic scaling is also possible with the Parrinello-Rahman barostat, making it more suited to protein-membrane systems. The box vectors  $\mathbf{b}$  are described by:

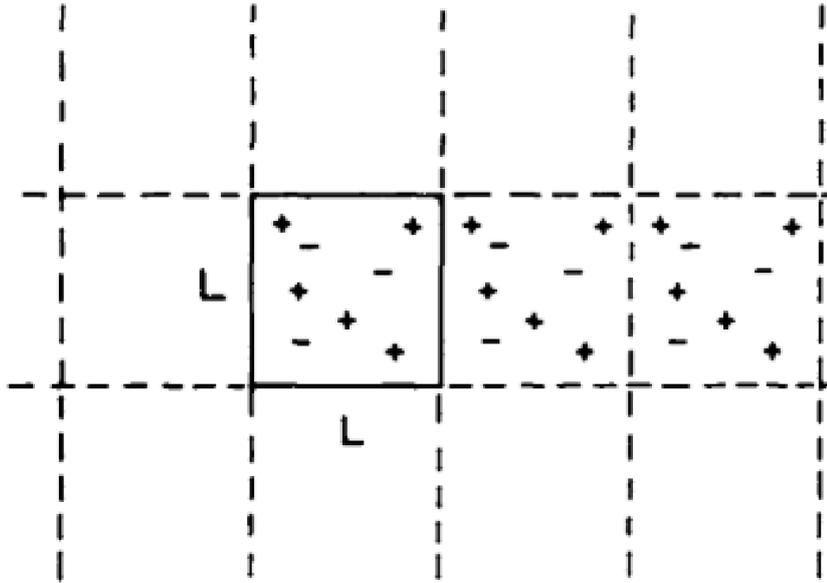
$$\frac{d^2\mathbf{b}}{dt^2} = \frac{V}{\mathbf{W}\mathbf{b}'} (P(t) - P_{target}) \quad (2.20)$$

where  $V$  is the volume of the box and  $\mathbf{W}$  is a matrix determining the coupling strength.

## 2.1.6 Nonbonded Cutoffs and Periodic Boundary Conditions

The calculation of nonbonded forces scales with  $N^2$  for  $N$  particles, since all pairs of particles have to be considered. To increase performance, a twin-range cutoff-scheme can be employed. Pairs can only interact within a certain radius,  $R_{short}$ . This introduces a neighbor list to the computational overhead, which needs to be updated every few time steps. A naive spherical cutoff creates a nondifferentiable potential. Therefore, switching and shifting methods can be used instead. Switching introduces a second,  $R_{long}$  cutoff, to make the falloff to zero less abrupt. Shifting changes the entire potential function instead of the near-cutoff region.

A popular method for dealing with the tail end of the nonbonded potentials is the Particle-Mesh-Ewald method [119], [135]. For this scheme, we first have to understand that MD is mostly used under periodic boundary conditions (PBC), meaning images of the simulation box are propagated infinitely in  $xyz$  direction [136]. This prevents unrealistic wall potentials and instead introduces a crowded environment similar to *in vitro* conditions. A particle can now interact with the nearest copy of another particle, but should not be able to interact with itself.



**Fig. 2.3.:** Periodic boundary conditions. Taken from Gunsteren et al. [136], reproduced with permission.

Remembering Eqn. (2.7), we see that the Coulomb energy is of the form  $\frac{1}{r}$ , which, combined with point charges, does not yield an estimable  $U_{tail}$  for an infinite system:

$$U_{tail} = \frac{N\rho}{2} \int_{R_{long}}^{\infty} U(r)4\pi r^2 dr \quad (2.21)$$

with  $N$  being the number of particles,  $\rho$  their density, and  $r$  their distance to the reference particle. Paul Ewald cleverly added two canceling terms to the point charges: one to represent each point charge as a Gaussian density, and one counteracting long-range Gaussian charge cloud. The long-range term can then be Fourier-transformed into crystallographic reciprocal space, solved and retransformed. The short-range term can be directly solved [137].

By optimizing the Gaussian width parameter and representing the charges on a grid the Particle-Mesh-Ewald procedure, as it is used now, can be calculated in  $\mathcal{O}(N \log(N))$  instead of  $\mathcal{O}(N^2)$  time [137].

## 2.1.7 Water Models: TIP3P

TIP3P water is an explicit water model using three point charges to model interactions [138]. Among the parameters that need to be modeled are the OH distance, HOH angle, and the charges of O and H. The agreement with neutron diffraction structures, diffusion coefficients, density and phase transitions are used for verification. TIP3P is derived from the older TIPS3 [139] model, but better optimized for the liquid phase. TIP4P, which features an additional charged pseudoatom to simulate the lone electron pairs, outperforms TIP3P in accuracy. However, TIP3P is computationally less expensive, and force fields have long been optimized for usage with this water model [140]. GROMACS uses holonomic constraints, fixing bond lengths, to save computing time, especially for solvent. To deal with different velocities calculated for the three atoms of water, the SETTLE algorithm is used [141]. All other atoms are reset to their bond lengths after each timestep using the LINCS algorithm [142] (or the older SHAKE method [143]).

## 2.1.8 Virtual Sites and Timestep

Fixing bond lengths, as above, increases the allowed timestep, which is limited by the process with the shortest oscillation period, from 0.5 fs to 2 fs. The next-quickest process is hydrogen bond-angle motion, which can be eliminated by introducing virtual sites. A massless pseudoatom is introduced, and the extra mass redistributed to the closest heavy atoms. The pseudoatoms keep fixed bond lengths, and inherit charge and van der Waals parameters from the hydrogen. This further increases the possible time step to 4 fs [100].

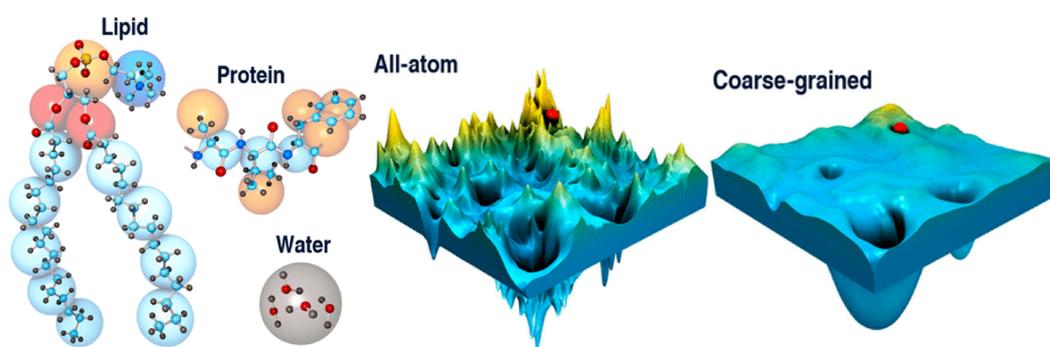
## 2.1.9 Enhanced Sampling

As introduced above, a molecular dynamics simulation is a computationally expensive process, so how can we use it to study processes like protein folding, which take seconds to minutes in the wet lab, but months on the supercomputer? While computing power and optimization have steadily increased the available spatiotemporal range (Fig.2.1), the exhaustive sampling continues to be a problem. The sampling problem occurs when a system is not simulated long enough, and low-energy states are sampled often and high-energy states rarely, so that the state remains trapped in a local minimum. In the following, different approaches to overcome this limitation, as used in this work, are presented.

## Coarse-Graining with MARTINI

One solution is coarse-graining the studied system, which enables larger timesteps and also smoothes the energetic landscape. In this work, the general-purpose MARTINI forcefield was used. MARTINI was developed in 2003 for lipid systems, and since then the original parameters have been improved and new classes of molecules (proteins, carbohydrates and nucleic acids) have been added [115], [144]–[147].

In general, four heavy atoms and their associated hydrogens are mapped into one coarse-grain bead (Notable exceptions: ions and ring systems), with pseudobonds between the beads (Fig. 2.4).



**Fig. 2.4.:** The MARTINI coarse graining principle. Left: Mapping of atoms to beads. Right: Free energy surfaces. Adapted from Kmiecik et al. [148], reproduced with permission.

There are four bead types in MARTINI 2 (1-4 in the following), and seven types in MARTINI 3 (1-7 in the following list):

1. polar (P)
2. intermediate/non-polar (N)
3. apolar (C)
4. monovalent ions (Q)
5. halo-compounds (X)
6. divalent ions (D)
7. water (W)

These are all (except W and D) further split into subclasses of beads, according to their polarity and hydrogen donor/acceptor status. Beads are available in sizes regular (R), small (S) and tiny (T) in MARTINI 2 and 3, while MARTINI 1 had big (B) beads as well. Lennard-Jones potentials are weighted by 10 discrete well-depths,

depending on which pair of beads interacts. Coulomb interactions are screened with a dielectric constant of 15 (MARTINI 2) and charges are generally inherited from the all-atom charges. With the introduction of MARTINI 3, a new type of bead for water was introduced, to replace the previous P-type water (which had required BP4 antifreeze particles). Compared to MARTINI 2, MARTINI 3 is less likely to have proteins sticking together, because protein-water and protein-protein interactions have been reweighted.

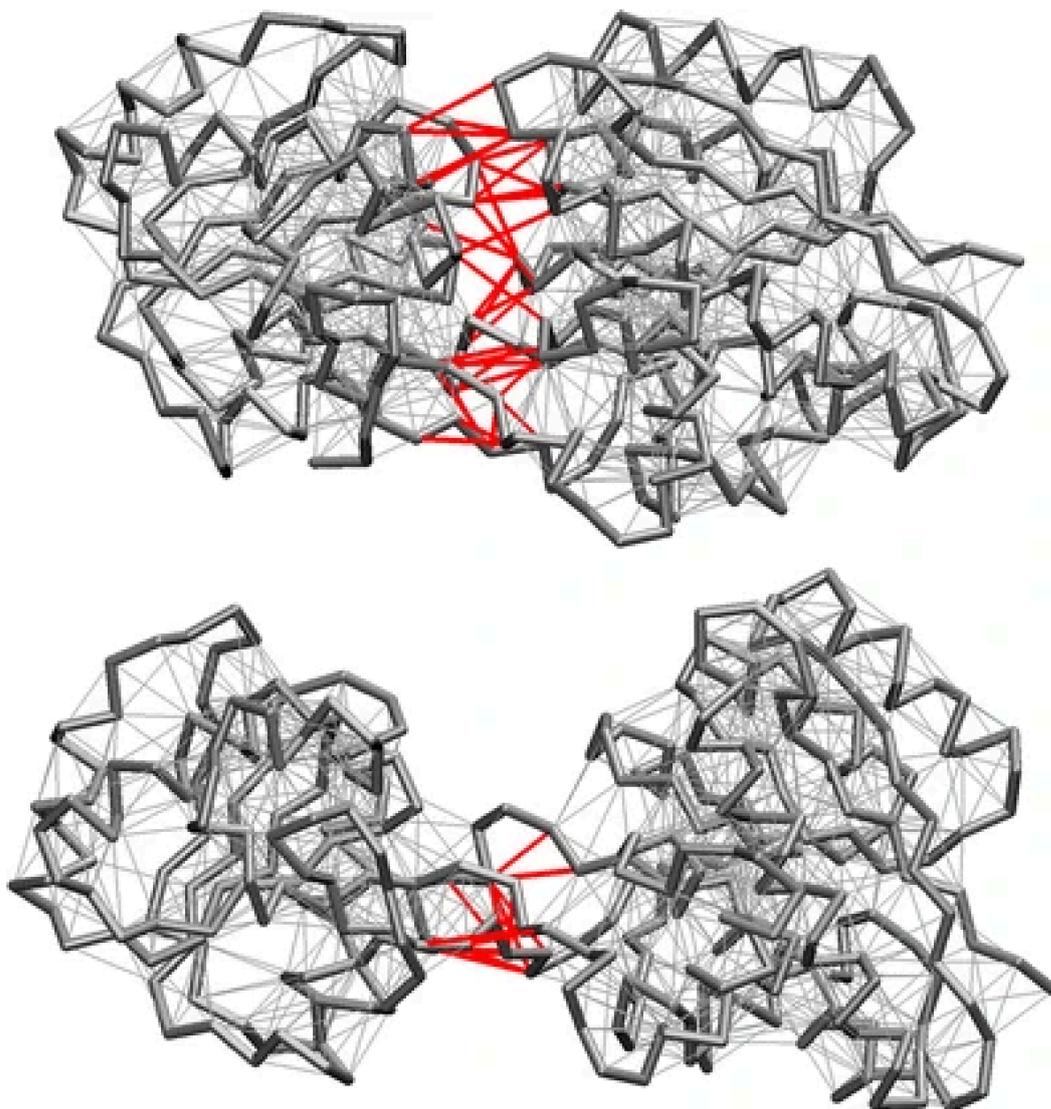
MARTINI parameters are verified by oil/water partitioning, comparisons to all-atom behavior and free energies of phase changes.

Simulation timesteps can be as high as 50 fs, though 20-40 fs are more frequently used. This already reduces computing times by a factor of at least 10 compared to all-atom simulations. Additionally, the reduced number of bonds and angles smoothes the energetic landscape (Fig. 2.4), resulting in 3-6 times larger diffusion constants (i.e. faster dynamics). Finally, the reduced number of pairs and the short interaction range speeds up energy calculations between timesteps, resulting in a total speedup of 1,000-10,000 for the user. For example, a large all-atom system simulated on several computing cores on a supercomputer could achieve a performance slightly under a hundred ns/d, while the same system, but coarse-grained, can easily reach a performance in the low  $\mu\text{s}/\text{d}$ , or alternatively still hundreds of ns/d on a normal computer.

A weakness of the basic MARTINI model can be the tendency to denature proteins, which is why it is often used in tandem with elastic networks [149]. To this end, CG backbone beads are linked with a spring ( $K_{spring}$ ) if they are close enough ( $R_{cutoff}$ ) in the initial structure, and if they are separated by at least two residues in the protein sequence. For one simulation,  $R_{cutoff}$  and  $K_{spring}$  are constant across the whole protein. The equilibrium length of a given spring is set to the experimentally observed distance between the two  $C_\alpha$  atoms that it connects. The resulting ELNEDIN model can reproduce both the global and local deformations of a protein, its residue fluctuations, and its large-amplitude collective motions, as observed in atomistic models. Values ranging from 0.8 to 1.0 nm for  $R_{cutoff}$  and from 500 to 1000  $\text{kJ mol}^{-1} \text{nm}^{-2}$  for  $K_{spring}$  are generally good starting points.

ELNEDIN allows for microsecond CG simulations, with occasional secondary structure changes, but it tends to overstabilize the initial conformation. This problem can be resolved by using the domELNEDIN script. In this model, the user defines

protein domains based on the residue indices, and all interdomain ELNEDIN links are deleted, to ensure interdomain flexibility (Fig. 2.5) [150].



**Fig. 2.5.:** ELNEDIN vs. domELNEDIN. Top: ELNEDIN, which includes interdomain springs. Bottom: domELNEDIN, which includes only intradomain springs. Figure taken from Siuda et al. [150], reproduced with permission.

### Backmapping from MARTINI

While coarse-grained simulations allow access to larger time-scales, they also filter out atomic dynamics. A common solution is the transformation of coarse-grained models back to all-atom resolution, called backmapping. The *backward* algorithm used in this work was developed as a general-purpose algorithm and works well with

MARTINI [151]. Mapping an AA model to CG representation is a simple process, because the mapping is unambiguous. For the reverse projection, no unique solution exists due to the degeneracy of AA states within one CG state. Backmapping consists of

1. projection using *backward.py*
2. correction using *initram.sh*
3. relaxation using *grompp&mdrun*

The general idea of *backward* is that any valid, probable enough, projected starting structure can be used, so backbone atoms can be placed along the CG backbone vector, and hydrogen or sidechain atoms at small random offsets from it. The script *initram.sh* is used to correct the geometry of e.g. double bonds, aromatics or chiral centers. Next, the structure's energy is minimized, once with nonbonded contributions turned off, and with the full contributions. For equilibration, a series of short MD simulations with time steps increasing from 0.2 fs to 2 fs, and harmonic position restraints. The resulting structure can be inspected as a snapshot, or the simulation can be continued at higher resolution for an added 100 ns. For ease of use, both the coarse-graining and backmapping process can be achieved with CHARMM-GUI [112], which provides the subsequently needed GROMACS input at the end.

## Replica Exchange MD

Another strategy against the sampling problem is replica exchange molecular dynamics simulation (REMD) (also called parallel tempering). Several instances of the same simulation setup are started at different temperatures and the conformations sampled at neighboring temperatures exchanged at regular intervals. Higher temperatures enable the crossing of energy barriers in shorter times  $\tau$ , according to the Arrhenius equation:

$$\tau = \tau_0 e^{\Delta G^\ddagger \beta} \quad (2.22)$$

where  $\Delta G^\ddagger$  is the Gibbs energy of the barrier,  $\tau_0 \approx 10^{-12}$  s and  $\beta = \frac{1}{k_B T}$ , with  $k_B$  being the Boltzmann constant.

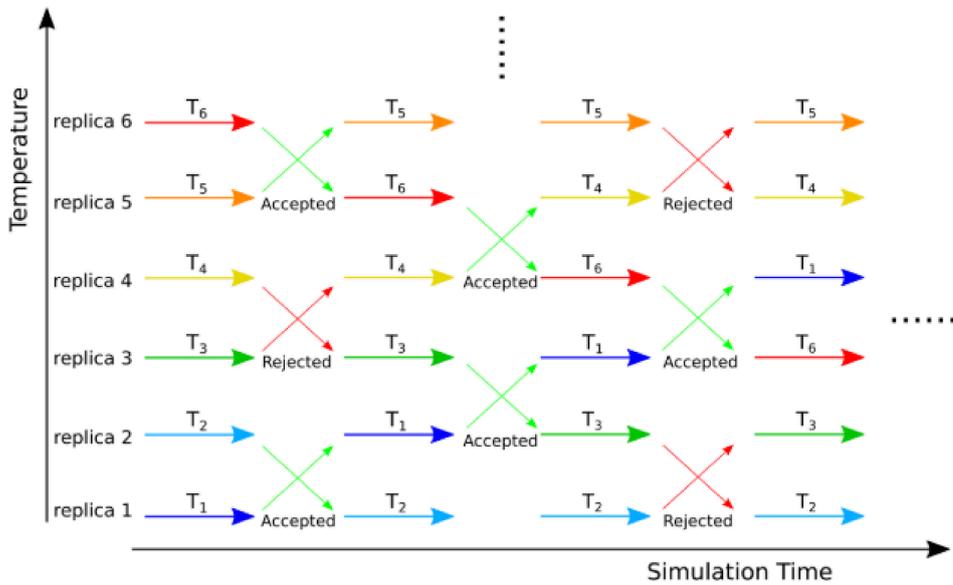
The final analysis of the system happens only at the originally intended temperature, since the higher temperatures are only the means for crossing energy barriers. These energy barriers exist between different stable conformations of the protein and

represent energetically unfavorable conformations. Using the higher temperature replicas as boosters, the system can explore new local minima.

To ensure that the detailed balance relation (microscopic reversibility) is fulfilled, the Metropolis criterion is used to check if an exchange attempt from replica  $i$  to replica  $j$  should be successful. The probability for this exchange is given by  $P_{i \rightarrow j}$  in the following formula:

$$P_{i \rightarrow j} = \begin{cases} 1 & \text{if } \Delta \leq 0 \\ e^{-\Delta} & \text{if } \Delta > 0 \end{cases} \quad \text{with } \Delta = (\beta_i - \beta_j)(U_j - U_i) \quad (2.23)$$

where  $\beta$  is again  $\frac{1}{k_B T}$  and  $U$  is the potential energy from equations (2.3)– (2.8). Thus, an exchange to an energetically lower state is always permitted, and an exchange to an energetically higher state only with a small probability (Fig. 2.6).



**Fig. 2.6.:** Replica exchange molecular dynamics principle. Here, six replicas are stacked from low to high temperature ( $T_1$  to  $T_6$ ). With a predefined frequency, attempts to swap replicas are made (marked with angled arrows). Acceptance and rejection according to the Metropolis criterion are also labeled. Figure adapted from Liao [152], reproduced with permission.

The original replica exchange method was improved upon by the Hamiltonian replica exchange molecular dynamics simulation (HREMD) [153]–[155]. One change is that  $\frac{U}{k_B T}$  (an extensive property) is scaled instead of the (intensive) temperature. The factor lambda serves for scaling the for Lennard-Jones, electrostatic and dihedral

energies. The user can define "hot" and "cold" regions of the system, and the scaling factor is only applied to the "hot" region. The "hot" regions then have an effective temperature of  $\frac{T}{\lambda}$  while the cold regions stay at  $T$ . Accelerating only, for example, the protein and not the bulk water, increases the computational efficiency. The swapping procedure always consists of an unconditional swap, to calculate the necessary energies for the Metropolis criterion (similar to equation (2.23)), followed by the swap back and the conditional swap.

## Umbrella Sampling MD

REMD-style methods belong to the collective-variable-free methods, which allow for a more naive, explorative approach. Previous knowledge of the system can be exploited in collective-variable-based methods, like metadynamics, steered MD and umbrella sampling. The theoretical background presented here is compiled from several sources [152], [156], [157].

In chemistry, the Gibbs free energy of a macrostate is related to its probability by:

$$G(\xi_0) = -k_B T \ln (\langle \delta(\xi - \xi_0) \rangle) \quad (2.24)$$

where the Dirac delta  $\delta$ , together with the ensemble average  $\langle \dots \rangle$  ensures that only states with  $\xi = \xi_0$  are considered.

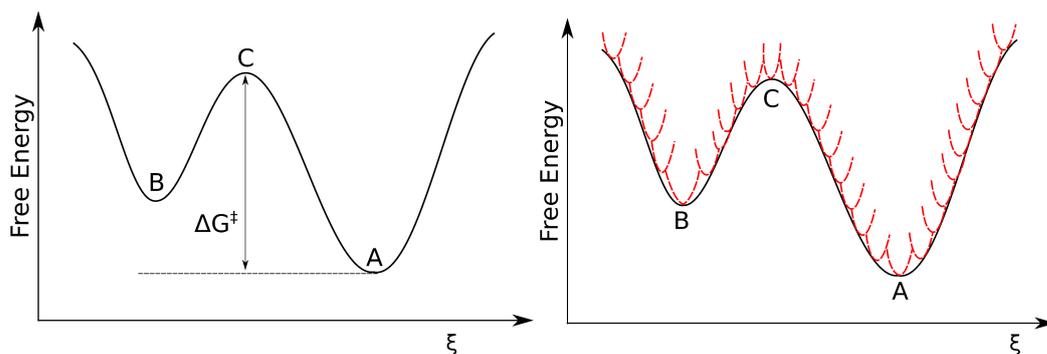
In other words, we separate our chosen collective variable  $\xi$  from all other variables of the system by integrating them out.

When we want to determine the height of an energetic barrier,  $\Delta G^\ddagger$ , we turn to

$$\Delta G^\ddagger = \Delta G_C - \Delta G_A = -k_B T \ln \left( \frac{\langle \delta(\xi - \xi_C) \rangle}{\langle \delta(\xi - \xi_A) \rangle} \right) \quad (2.25)$$

But when state C is not well sampled, this calculation becomes impossible. Therefore, we introduce additional harmonic potentials along our collective variable  $\xi$ , each centered around a given  $\xi_0$ . These additional potentials have to be considered when calculating  $\Delta G$ , also called "potential of mean force":

$$\Delta G_{\xi_0} = -k_B T \ln (\langle \delta(\xi - \xi_0) \rangle_{E+V}) - V(\xi_0) + kT \ln \langle e^{\beta V} \rangle_{E+V} \quad (2.26)$$



**Fig. 2.7.:** Umbrella Sampling. Left: Schematic free energy landscape along collective variable  $\xi$  Right: Added harmonic potentials for umbrella sampling. Adapted from Liao [152], reproduced with permission.

This implies that, we can get the unbiased potential for a single  $\xi_0$  by observing our biased (E+V) ensemble, then simply subtracting the added potential  $V(\xi_0)$  and adding a corrective factor  $kT \ln \langle e^{\beta V} \rangle_{E+V}$ .

The most popular method for getting the unbiased probabilities ( $P_\xi$ ) of the entire range of  $\xi$  is the weighted histogram analysis method (WHAM). It consists of a system of two coupled equations (easily obtained from Eq. (2.25)), which are solved iteratively:

$$P(\xi) = \frac{\sum_{l=1}^N h_l(\xi)}{\sum_{k=1}^N n_k e^{\frac{F_k - V(\xi)}{k_B T}}} \quad (2.27)$$

$$F_k = -k_B T \ln \left( \sum \xi P(\xi) e^{\frac{V(\xi)}{k_B T}} \right) \quad (2.28)$$

where indices  $l$  and  $k$  run over the  $N$  umbrella windows independently, calculating the fraction of macrostates in the current bin compared to the data points in all windows  $\frac{h_l}{n_k}$ , weighted with the corrective factors introduced in Eq. (2.26). The corrective factors  $F_k$  are sampled from the unbiased probabilities  $P_\xi$ . While the equations above employ  $\xi$  for simplicity's sake, the user can define the WHAM bins to be different from the windows used during simulation.

In MD simulations, the ensemble averages can be replaced with time averages, if the system has been simulated long enough. Finally, to affix error bars to the potential of mean force, a bootstrap technique can be employed, as is implemented in `gmx_wham`. In this technique, the total data are split into subsets several times, while making sure the subsets are still covering the range of  $\xi$ . For each subset, the

potential of mean force is calculated independently, and thus, an uncertainty can be derived [158].

In this work, the existence of two discrete protein conformations offers itself to the application of, first, a steered MD simulation to generate the intermediate conformations, and then, an USMD to affix an energetic cost to the barrier. Conformations along the collective variable (CV) can be generated and restrained using either the gromacs pull-code or the plugin plumed [159]–[161].

## 2.2 Molecular Docking

The simulation of dynamics is much more computationally expensive than molecular docking, which is used to predict binding conformations and binding affinities. Small ligand docking has become a standard tool in computer aided drug design, and can be supported by many protein-ligand complexes available in the Protein Data Bank (PDB). Protein-protein complexes are naturally harder to resolve structurally, as evidenced by their only 25% contribution to all PDB structures (queried on Nov. 16th, 2022 at [www.ebi.ac.uk/pdbe/](http://www.ebi.ac.uk/pdbe/)), only a few hundred of which are suitable for benchmarking [162]. While AlphaFold [163] has recently revolutionized structure prediction, the prediction of protein complexes is still in its infancy due to the small training set. Nonetheless, several protein-protein docking services exist and annually compete in the critical assessment of predicted interactions (CAPRI) competition.

### 2.2.1 Protein-Ligand Docking

In this work, we use AutoDock [164]–[166] to perform a redocking of GTP into BDLP. The binding site of GTP being known, this is only done for automatization purposes of the ca. 100 umbrella sampling windows. AutoDock is a software suite consisting of

1. AutoDockTools to prepare the structures
2. AutoGrid (or Vina) to generate the grid and precalculate affinities
3. AutoDock 4 or Vina to perform the docking
4. AutoDockTools to analyze the results

Precalculating the affinities results in faster evaluation later. AutoGrid embeds the protein in a 3D grid and places a probe atom at each grid point. Each point is assigned an interaction energy, for each atom type in the ligand. This grid is

dependent on the ligand, but the user can later refine which  $xyz$  interval should be used for docking. Docking again consists of two subprocesses: the sampling of conformations for the ligand, and the scoring of these conformations.

The scoring function of AutoDock Vina depends on all  $i - i + 3$  atomic pairs, both inter- and intramolecular, which are assigned empiric energies. The function includes two steric terms, a repulsion term, a hydrophobic term, hydrogen bonding (up to 8 Angstrom) and rotational degrees of freedom. AutoDock 4 has a more complex scoring function that also includes desolvation and electrostatics, but takes longer to compute than the Vina version. Minimization of the score happens with an Iterated Local Search Global Optimizer, meaning that conformations are locally mutated and optimized, and then accepted according to the Metropolis criterion (2.23).

Local optimization follows the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm, which is a quasi-Newton method where the Hesse matrix  $\mathbf{H}_{ij}$  is approximated. In Alg. 2), each point in the high-dimensional space ( $\mathbf{x}$ ) of protein conformations is mapped to an energy value ( $f(\mathbf{x})$ ). The search direction  $\mathbf{p}_k$  is set opposite the gradient  $\mathbf{g}_i(\mathbf{x})$ , additionally using the second derivative  $\mathbf{H}_{ij}$  for higher accuracy. The step size  $\lambda_k$  is set to satisfy sufficient decrease of  $f$ , and sufficient directional derivative reduction (step 2 below), where  $\alpha$  and  $\beta$  are accuracy parameters set by the user.

#### Algorithm 2: Pseudocode of Minimization

**Data:**  $f(\mathbf{x}), \mathbf{g}_i(\mathbf{x}) = \frac{\delta f(\mathbf{x})}{\delta \mathbf{x}_i}, \mathbf{H}_{ij}(\mathbf{x}) = \frac{\delta^2 f(\mathbf{x})}{\delta \mathbf{x}_i \delta \mathbf{x}_j}$   
**Result:**  $\min_{\mathbf{x}}(f(\mathbf{x}))$   
**while** not converged & steps  $k < k_{max}$  **do**

1. set direction  $\mathbf{p}_k \leftarrow -\mathbf{H}_k^{-1} \mathbf{g}_k$ ;
2. set stepsize  $\lambda_k$  so that  $f(x_{k+1}) \leq f(x_k) + \alpha \lambda_k \mathbf{g}_k^T \mathbf{p}_k$  and  $|\mathbf{g}_k^T \mathbf{p}_k| \leq \beta |\mathbf{g}_k^T \mathbf{p}_k|$  with  $0 < \alpha < \beta < 1$ ;
3.  $\mathbf{x}_{k+1} \leftarrow \mathbf{x}_k + \lambda_k \mathbf{p}_k$ ;
4. go to 1.;

New conformations are sampled in a highly dimensional space (denoted  $x$  here) by treating the protein receptor as rigid and the ligand as flexible.

AutoDock 4, on the other hand, uses a Lamarckian genetic algorithm. For this method, the degrees of freedom of the ligand (translation, rotation, internal angles, etc.) are equated to the genetic code, and phenotypes are generated by setting specific values for these genes. The individuals  $i$  of a generation are evaluated according to:

$$n_0 = \frac{f_w - f_i}{f_w - \langle f \rangle} \quad (2.29)$$

This means that individual  $i$  will have  $n$  offspring in generation 0, according to how its fitness  $f_i$  (or binding energy) compares to the worst ( $f_w$ ) and average  $\langle f \rangle$  fitness of the last 10 generations.

During either minimization scheme, inter- and intramolecular energies are considered, but final free energies of binding depend only on the intermolecular contributions.

## 2.2.2 Protein-Protein Docking

Protein-protein docking is more computationally expensive than protein-ligand docking due to the increased search space of two potentially flexible, large molecules.

When docking the interaction partners of mGBP7, the servers PatchDock [167], Zdock [168], [169], Swarmdock [170]–[172], Firedock [173], ClusPro [174] and HADDOCK [175]–[177] were used, to explore different algorithms. For the homodimer docking of GBPs, the ClusPro server was chosen because it explicitly offered multimer docking and the inclusion of SAXS or XL-MS restraints.

PatchDock uses a shape complementarity approach by matching concave/convex/flat surfaces, instead of brute-forcing a 6D space search. Matching conformations are evaluated further with desolvation energy, and then clustered via RMSD. ZDock works similarly, with an added electrostatics term. Both PatchDock and ZDock are rigid-body docking programs, while SwarmDock can model the flexibility of both proteins. SwarmDock uses a local docking supported by normal modes (see Sec.2.3), then a particle swarm optimization, and clusters the results. Conformations are generated by moving protein atoms, along linear combinations of normal modes, randomly away from their starting positions. The score is then based on the CHARMM force field, using only Coulomb and van der Waals terms with a 9 Å cutoff. Firedock is offered as a refinement service after ZDock, to reorient side chains, which happens with a rotamer library and Monte-Carlo minimization, followed by a physics-based scoring function. ClusPro performs first a brute force, Fast Fourier Transform, rigid body docking, followed by root mean square deviation (RMSD) based clustering of the best structures, and finally the largest clusters are refined with a short CHARMM simulation to resolve clashes. This puts ClusPro in the medium flexibility category of docking programs, and ClusPro also offers optional restraints instead of ab initio

docking. While HADDOCK can now be used without restraints, its strength is the highly localized flexible docking. To generate the restraints needed, *a priori* experimental information of the complex is useful. In this work, the results of the other docking servers were used to generate restraints. Structures are ranked according to their intermolecular van der Waals, electrostatic and restraint energies. Note that none of the servers used in this work are template-based, i.e. none rely on homology search, all are physics-based. Shortly after our docking results were summarized in a manuscript (attached in App.B), Alphafold won the critical assessment of techniques for protein structure prediction (CASP) competition, and might in the future revolutionize docking as well.

## 2.3 Normal Mode Analysis

As briefly mentioned for SwarmDock and MARTINI, proteins can also be modeled as elastic networks. For this, it is sufficient to consider only a subset of backbone beads, placed on the initial structure's centers of mass for the corresponding residues. These positions are assumed as the minima of harmonic potential wells, modeled as simple Hookean springs. After constructing and diagonalising the Hessian matrix (second derivative of the potential with respect to mass-weighted coordinates), we obtain eigenvectors as a new base to describe the motions of the protein. The eigenvector with the largest eigenvalue can have a significant contribution towards the protein's dynamics, and might require significantly less than the original  $3N$  coordinates. As such, elastic network models (ENM) are similar to principal component analysis (PCA), with the added benefit that normal modes can be calculated from a single starting structure, while PCA requires a trajectory. Elastic network models of proteins can thus circumvent the computationally expensive molecular dynamics simulations and predict protein motions [171], [178], [179]. Apart from the applications named above, normal mode analysis will later be used to analyse BDLP motions.

## 2.4 Dynamic Cross-Correlation Analysis

However, Dynamic Cross-Correlation (DCC) is the more accurate method, working not just with a single frame, but the whole trajectory. According to the formula:

$$C_{ij} = \frac{\langle \Delta \mathbf{r}_i \cdot \Delta \mathbf{r}_j \rangle}{\sqrt{\langle \Delta \mathbf{r}_i^2 \rangle} \sqrt{\langle \Delta \mathbf{r}_j^2 \rangle}} \quad (2.30)$$

the correlation coefficient between atoms  $i$  and  $j$ ,  $C_{ij}$  can assume values of 1 (correlated) to -1 (anticorrelated), with 0 meaning no correlation.  $\Delta\mathbf{r}(t) = \mathbf{r}(t) - \langle \mathbf{r}(t) \rangle$  represents the displacement from the average position of the respective atom, and the brackets denote a time average. More specifically, the displacement is calculated in vector form  $\Delta\mathbf{r}$  and only afterwards the magnitude of the vector  $\sqrt{\Delta\mathbf{r}^2}$  is used in the equation. Here, the implementation of DCC packaged in MD-Task will be used [180].



# Bacterial Dynamin Like Protein

## Introduction

The bacterial dynamin-like protein from *Nostoc punctiforme* (BDLP) is a representative of the dynamin-superfamily proteins. It consists of a globular GTPase domain (head) sitting atop the helical neck and trunk domains. At the trunk base, a paddle domain mediates membrane interaction with the outer leaflet of the inner plasmamembrane (see Fig. 1.3) [181], [182]. In the following sections, we explore the conformational change between the closed and open form of BDLP, as well as smaller GTP-induced changes, and finally, its membrane-bound behaviour.

### 3.1 Hinge Motion

The hinge motion of BDLP is discussed in detail in publication I:

**Allosteric communication induced by GTP binding sets off a closed-to-open transition in a bacterial dynamin-like protein**

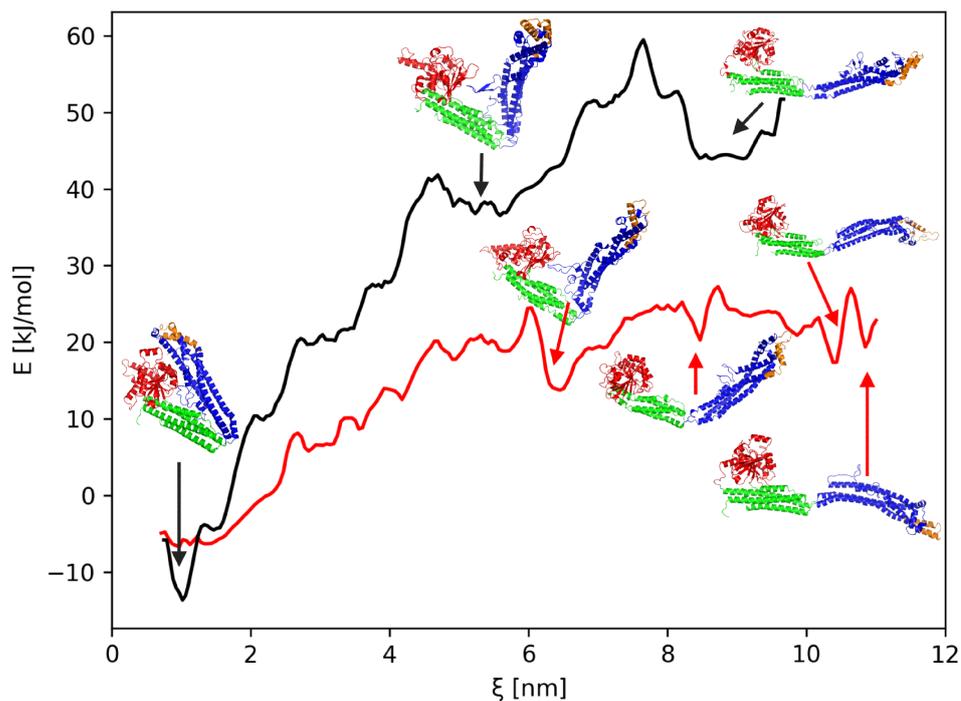
*Wibke Schumann, Birgit Strodel*

bioRxiv, DOI:10.1101/2023.01.16.524228 (2023)

which is an original publication, attached in Appendix A, to which I contributed the execution and analysis of molecular dynamics simulations, producing the figures, and writing of the first draft.

The key points are summarized here for convenience: First, we have simulated the apo- and GTP- bound form for  $1\ \mu\text{s}$  to understand the unbiased dynamics and effects of GTP bound to BDLP. While the closed form of BDLP is stable for at least 600 ns, both in apo- and GTP-Mg<sup>2+</sup>-bound form, the open conformation is very flexible and spontaneously starts to close. This is also reflected in the potential of

mean force (PMF) obtained from a 88x100 ns USMD, where the open state is only metastable, see Fig. 3.1. GTP binding leads to a lowering of the energetic barrier, and a stabilization of the open state, and the emergence of a previously unknown, wide-open state. When investigating possible allosteric effects of GTP that might explain this behavior, we first found two flaps in the G-domain, closing to trap GTP. Secondly, we found that this releases the membrane paddle from its interaction with the G-domain. This is the first major contribution towards the energetic barrier. With the paddle abstracted, six salt bridges remain to prevent opening. We have shown that these do not break all at once, but rather in a shearing motion, where some are initially strengthened when GTP binds.



**Fig. 3.1.:** Potential of mean force obtained from USMD simulations of BDLP. The x-axis represents the restrained parameter, the distance between neck and trunk. From left to right, BDLP is first closed, then half-open, then open. The y-axis represents the free energy determined via WHAM analysis, with the black curve for the apo state and the red curve for the GTP-Mg<sup>2+</sup> (holo) state. Figure created with Biorender.

## 3.2 Membrane Binding

After our study of the BDLP hinge motion, the half-open and open state minima seemed not stable enough to justify their occurrence. Therefore, we studied BDLP dimers and BDLP membrane systems additionally.

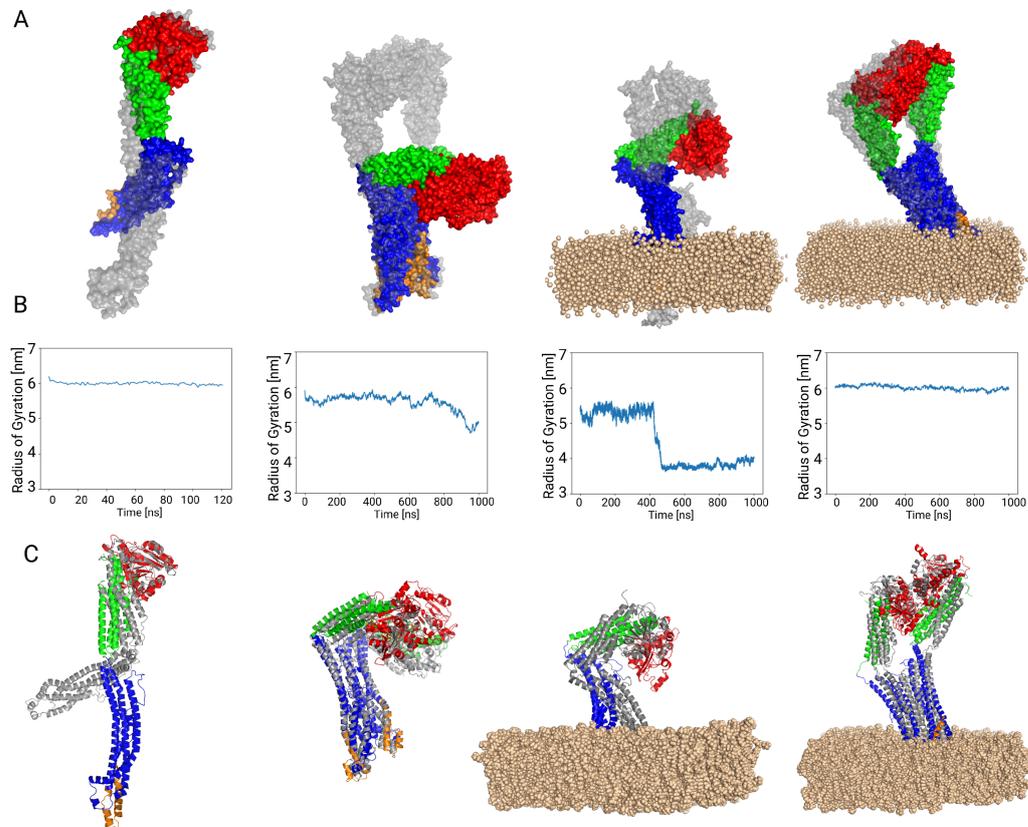
### Methods

This was accomplished with initial CG simulations, for which the all-purpose force-field MARTINI 2 was used, with the setup achieved in CHARMM-GUI. The output of CHARMM-GUI was then modified with the `domElNeDyn` script [150], to ensure that the protein was restrained by elastic networks, but only within domains. Since the `domElNeDyn` script only works with MARTINI 2, a test simulation of the dimers in MARTINI 3 was performed with `ElNeDyn`, to rule out undue stabilization of protein-protein-interaction. Since the dimer did not dissociate in MARTINI 3 either, the better interdomain flexibility was judged worth the downgrade to MARTINI 2. The energy minimization, equilibration and simulation were carried out according to the `mdp` files provided by CHARMM-GUI. After 1  $\mu$ s simulation, the backmapping to the AA level was performed using the CHARMM-GUI all-atom converter, followed by a minimization in vacuum, and 100 ns simulation under conditions analogous to the respective CG system. In absence of knowledge about *N.punctiforme* membrane composition, POPC was used as a generalized membrane model. The dimer structures were obtained from ClusPro, but selected for their similarity to the known crystal structure 2W6D (open) and to the closed dimer featured in [181].

### Results

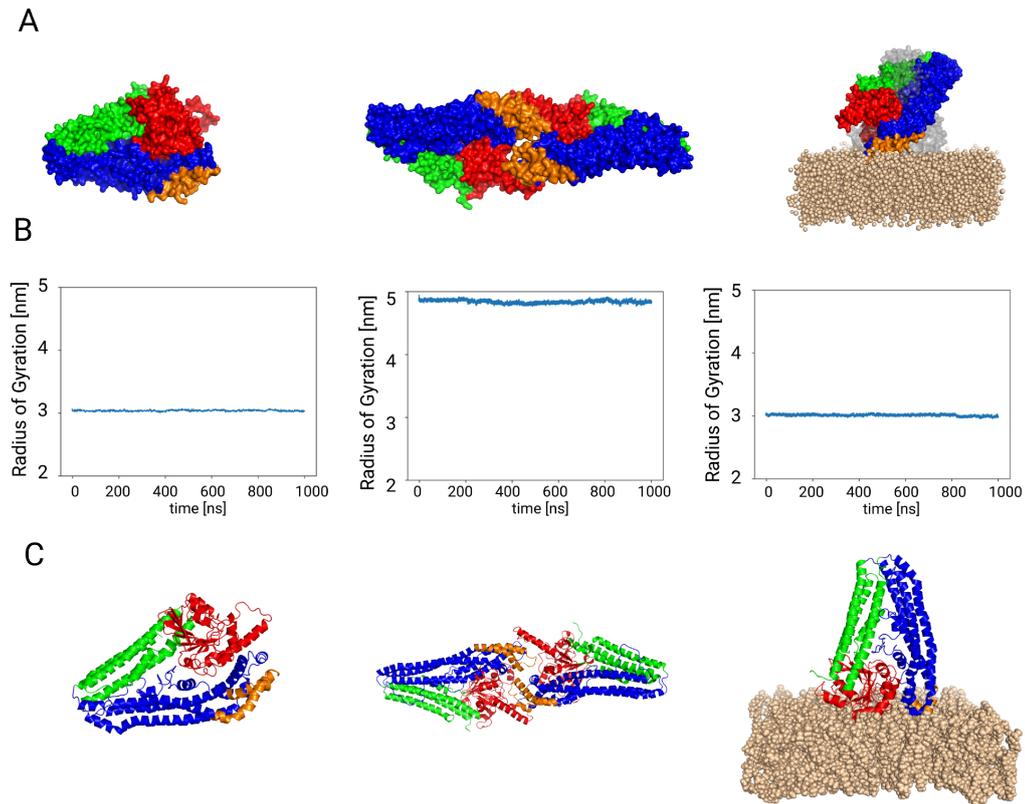
Initially, we tested the dynamics of the open conformation of BDLP monomer and dimer in solution and on the membrane. As shown in Fig. 3.2, the CG systems differ in their mobility. The open conformation of BDLP tends towards the wide-open structure, as observed before in the all-atom (AA) simulations with GTP. The BDLP dimer in solution assumes the half-open structure that was identified as metastable in the USMD simulation, the same is true for the monomer in the membrane environment. An exception to this is the membrane-bound dimer, which stays mostly stable. After backmapping, the wide-open structure immediately reverses to the open conformation. This is probably explained by the MARTINI forcefield smoothing an important conformational change which in AA would only happen with GTP

bound. The dimer in solution stays in its half-open form, as does the monomer on the membrane. The membrane-bound dimer is stabilized in the open conformation.



**Fig. 3.2.:** CG simulation of BDLP systems. A: CG systems of BDLP simulated for  $1\mu s$ , open (solution), dimer (solution), open (membrane), dimer (membrane). Gray: starting structure. Colored: end structure. B: Time development of the radius of gyration of the CG systems. C: 100 ns backmapping simulation in AA. Gray: starting structure. Colored: End structure. Figure created with Biorender.

Next, we simulated the closed conformation of BDLP ( Fig. 3.3), which is stable in solution even when simulated in CG for  $1\mu s$ . The same is true of the closed dimer. The closed conformation also seems less favorable for interacting with the membrane and is observed dissociating from it in the CG simulation. After backmapping, the membrane-bound closed BDLP monomer does not dissociate, instead we can see that the GD helices also have transmembrane tendencies ( Fig. 3.3C).

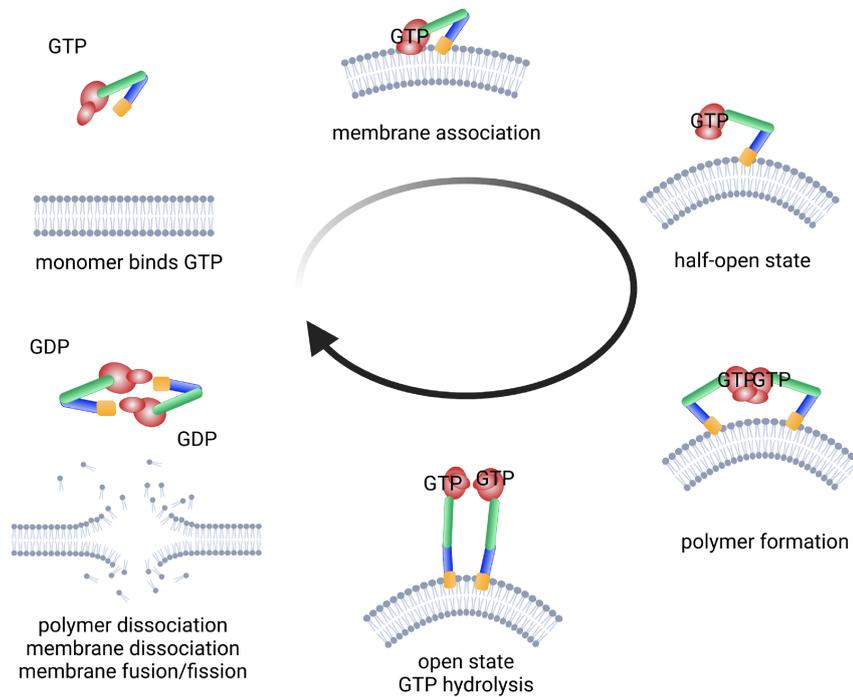


**Fig. 3.3.:** CG simulation of BDLP systems. A: CG systems of BDLP simulated for  $1 \mu s$ . Gray: starting structure. Colored: end structure B: Time development of the radius of gyration of the CG systems. C: 100 ns backmapping simulation in AA. Gray: starting structure. Colored: End structure. Figure created with Biorender.

We take this as evidence that the energetic profile will change upon membrane association and dimerization, likely by deepening the minimum associated with the half-open structure.

A speculative reaction cycle as shown in Fig. 3.4 would start with GTP binding to the closed monomer, loosening the paddle-GD interaction. It should be noted that in the membrane bound state of the closed BDLP monomer shown in Fig. 3.3, GTP would be unable to enter into the GD. The paddle is then able to insert into the membrane, while flap2 (residues 251-271), closes upon GTP and emerges from the membrane. The half-open state is assumed more easily when the paddle is occluded by the membrane, since only the breaking of the stalk-saltbridges is needed. The half-open form can further be coaxed into the open form by dimerization at the membrane. Here, the shearing motion is crucial for enabling the hinge even in the dimer without collision. These membrane-bound polymers are then stable as long as GTP is not hydrolyzed, after which the process is reversed. When the closed BDLP dissociates from the membrane, perhaps by forming an unproductive closed dimer,

the thinned, highly curved membrane is unstable and ready for fusion or fission. Continued membrane binding with less conformational change would favor fusion, as proposed by Rennie et al. [183].



**Fig. 3.4.:** BDLP GTPase cycle  
Proposed interplay of GTP binding, membrane binding and polymerization and its effect on membrane shape. Figure created with Biorender.

## Conclusion

In conclusion, we have integrated the small-scale GTPase flap and paddle motions of BDLP with the energetic profile for the large-scale motion provided by USMD simulation, and with the help of CG simulations a tentative cycle of action is proposed. In the next chapter, we will explore how to transfer this knowledge to guanylate binding proteins.

# Guanylate Binding Proteins

Guanylate binding proteins are more distant family members of the dynamin superfamily. In place of the bundle signalling element, they feature an M-domain, and their E-domain stretches back along the whole length of the latter. GTP binding leads to liberation of the membrane binding motif [184], which can be a posttranslational isoprenylation at the C-terminus, or a C-terminal transmembrane helix. GBPs localize at cell-invading pathogens like influenza, *Chlamydia* or *T.gondii*, and then recruit autophagy proteins as part of the immune response. GBPs are unique in their ability to hydrolyze GTP all the way to GMP, which could potentially prolong the lifetime of their polymeric, membrane-bound structures. In this chapter, we discuss several dimerization modes and their functional implications, followed by an exploration of the hinge motion, and finally, membrane interaction.

## 4.1 Dimers

To achieve oligomerization, DSPs use up to four interfaces in the helical (stalk) region, and the previously loaded GTP is stable until two G-domains interact. The result is a conformational change, and the release of GDP, as well as dissociation of the oligomer [28], [183], [185]. GBPs are able to hydrolyze GDP further to GMP, which suggests that they remain in an oligomeric state even after the first hydrolysis. The second hydrolysis takes place either in a tetrameric state, or in a drastically extended conformation of the dimer [186], [187].

We know of two structures for GBPs: Ghosh et al. proposed a dimer formed by the G-domains of hGBP1, while in the structure reported by Cui et al. two hGBP5 molecules interact lengthwise, with a cross at the GD base [188], [189].

In order to bridge the gap towards the macromolecular observation of thousands of GBPs on the membrane [74], we studied dimerization modes. Their structure could yield important hints about the encoded function, whether it is a destabilization of pathogenic membranes, or fusion of the PVM with the autophagic machinery.

## Results

The results can be found in in publication II:

**Integrative modeling of guanylate binding protein dimers** *Wibke Schumann, Jennifer Loschwitz, Daniel Degrandi, Klaus Pfeffer, Jens Reiners, Sander H.J. Smits, Gereon Poschmann, Kai Stühler, and Birgit Strodel*

bioRxiv, DOI:10.1101/2022.12.20.521180 (2022)

which is an original publication, attached in Appendix B, to which I contributed the docking, execution and analysis of molecular dynamics simulations of the dimer systems, producing several figures, and writing the first draft of the dimer section, as well as abstract, introduction and conclusion.

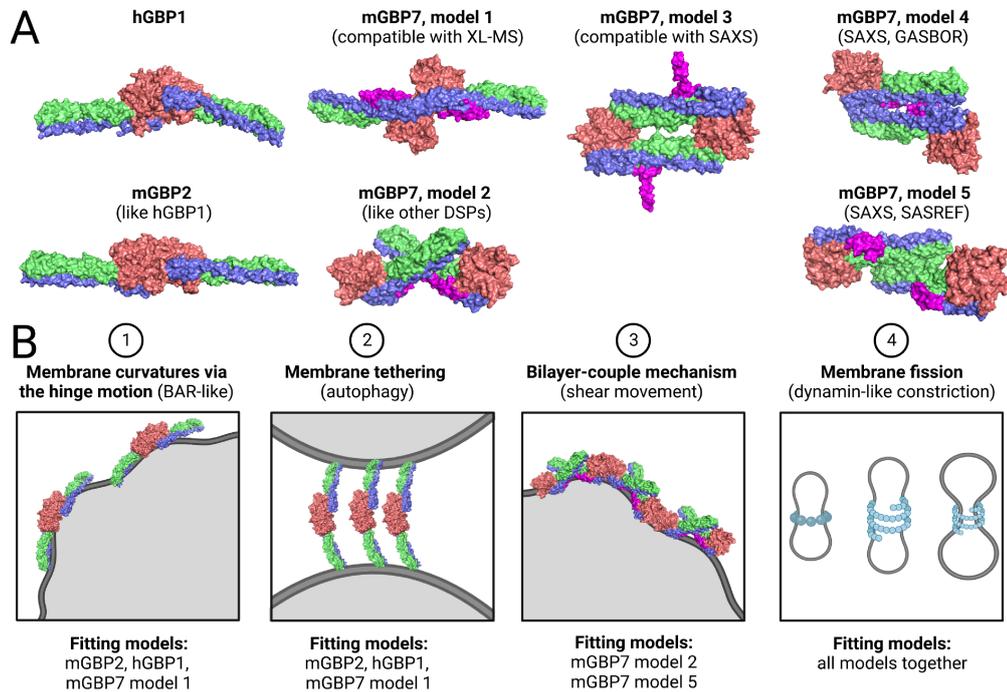
The key points are summarized here for convenience:

hGBP1 and mGBP2 were predicted as GD dimers, even with ClusPro being a physics-based (i.e. not knowledge-based) server. The proposed dimer models for mGBP7 do not include a GD dimer, and mGBP7 differs in one important aspect from hGBP1 and mGBP2: it can form dimers without GTP [190]. In order to compare with experimental data, supplied by Jens Reiners and Sander Smits (SAXS), and Gereon Porschmann and Kai Stühler (XL-MS), we simulated 5 dimer models of mGBP7. One of them agrees well with the XL-MS data, while three have shapes that agree with the SAXS data. One other interesting model shows a crossed-stalks motif (Fig. 4.1), reminiscent of other DSPs.

The extended length of a GD dimer and its concerted hinge motion could introduce more stress onto a membrane than a monomer can, comparable to the observations made for BAR (Bin/Amphiphysin/Rvs) domain proteins [12]–[14]. Another possible mechanism of introducing membrane stress/shearing force would require both the C-terminal (CT) tail and the stalk tip of a highly angled dimer to be inserted into the membrane. Additionally, the insertion of a single helix can already result in a hydrophobic mismatch that facilitates membrane fission [191], comparable to a wedge driven into the lipid bilayer.

It is also possible that the structures spatially complement each other to form a lattice, a multimer consisting of several different dimers. On the other hand, the dimers may also temporally complement each other during the GTPase cycle, with rearrangements triggered by nucleotide binding or hydrolysis. Finally, the dimerization and membrane binding mode might change in response to different membrane

compositions (like the host-derived PVM vs. the *T.gondii*-derived membrane).



**Fig. 4.1.:** Dimer models studied (A) and their potential mechanism of action (B). Figure created together with Jennifer Loschwitz using Biorender.

We then extended the work on dimerization models to mGBP9. mGBP9 is not as strongly upregulated as mGBP2 and mGBP7 upon *T.gondii* infection, and is largely unstudied. It does not colocalize with mGBP1 and mGBP2, but it targets the *T.gondii* PVM [73]. It is also effective against *Chlamydia trachomatis* infections [192]. Our study investigated how the dynamics change from the mGBP9 monomer to the dimer, and from the apo- to the holo-form, and which sequence differences might cause this, and resulted in manuscript III:

### Effect of GTP binding on Switch II conformation in mGBP9 dimers

Jennifer Loschwitz, Wibke Schumann, Jens Reiners, Jens Lichte, Daniel Degrandi, Klaus Peffer, Birgit Strodel, Sander H. J. Smits

which is an original manuscript, to which I contributed the docking, execution and analysis of molecular dynamics simulations of the dimer systems, producing several figures, and writing of the dimer section. In the following sections (Methods and

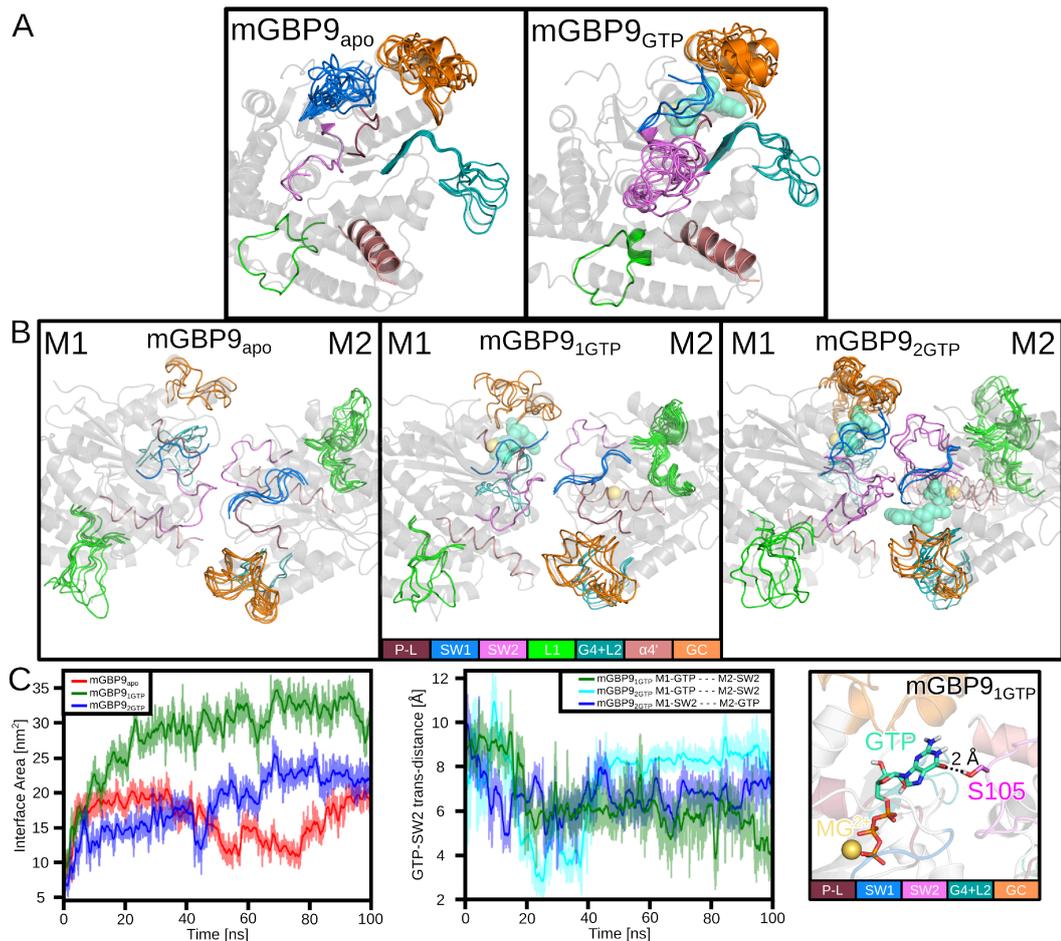
Results), excerpts from the manuscript will be included.

## Methods

For the MD simulations, the Amber99SB\*-ILDNP [193]–[195] force field, TIP3P [138] water, and GROMACS 2018.6 [104] were used. All simulations were performed at the temperature of 310 K and the pressure of 1 bar. For analysis, we used the Daura clustering algorithm as implemented in GROMACS [196], applied after aligning the trajectory to (each) GD. The  $C_{\alpha}$  atoms of the motifs G1-G5 and of helix  $\alpha 4'$  were each clustered with a cutoff of 0.25 nm and option "nofit". For the PCA, one index group was created for all motifs, including all atoms, and the GROMACS methods covar and ana eig were used to extract the first two principal motions. For the extent of the hinge motion, the spatial distribution function (gmx\_sdf) of the stalk tip (amino acid 480) was calculated from the aligned trajectory above. The dimer interaction area was calculated as described in [197]. Finally, the minimum distance between the GTP molecule and the opposing switch 2 (SW2) region was calculated.

## Results

An initial superposition of the SAXS-resolved dimer of mGBP9 with the hGBP1 dimer (provided by Jens Reiniers and Sander Smits) showed a conformational change in the SW2 region, warranting further investigation by MD simulation. The GD-GD-dimer mGBP9 dimer structure, as resolved by SAXS, was simulated in apo (mGBP9<sub>apo</sub>), 1GTP/2MG2+ (mGBP9<sub>1GTP</sub>) and 2GTP/2MG2+ (mGBP9<sub>2GTP</sub>) states, in order to observe changes between monomer and dimer, as well as changes in structure upon GTP binding. Simulating both mGBP9<sub>1GTP</sub> and mGBP9<sub>2GTP</sub> structures allows for judging *cis* (on the chain binding GTP) and *trans* (on the opposing chain) effects of GTP. All three simulations show an increase in dimer interface area during the simulation (Fig.4.2C). This is accompanied by an induced-fit scenario, in which the two G $\alpha$  domains adapt to each other. However, the apo trajectory displays a marked dip in interface area at 50 ns. The 1GTP trajectory reaches the highest interaction area of 35 nm<sup>2</sup>, followed by the 2GTP trajectory with 20 nm<sup>2</sup>. It seems that GTP plays a dimer-stabilizing role in mGBP9. The interface area is roughly anti-correlated with the minimum *trans*-distance of GTP and SW2 (Fig.4.2C).



**Fig. 4.2.:** Comparison of mGBP9 apo-dimer, 1GTP-dimer and 2GTP-dimer. AB: Clustering of structural motifs in the apo- and GTP-monomer and dimers. C: Interface area of the dimers during the simulation, distance between SW2 and GTP *in trans*, close-up of the relevant residue. Figure created together with Jennifer Loschwitz.

Specifically, the PCA reveals that this minimum distance is synonymous with the O6-carbonyl-Ser104-hydroxyl, which, at 5 Å get just close enough for a hydrogen bridge after 100 ns (Fig.4.2C). This type of *trans*-interaction, although visible in the hGBP1 dimer structure (PDB ID 2B92 [188]), has not been described before. Furthermore, our sequence alignment reveals this serine as a unique feature of mGBP9, compared to other GBPs. Our root mean square fluctuation (RMSF) and clustering analysis (Fig.4.2AB) reveal that switch 1 (SW1) and SW2 become more flexible in both dimeric GTP-bound states, compared to the dimeric mGBP9<sub>apo</sub>. This is more marked in the RMSF for mGBP9<sub>1GTP</sub>, but more marked in the clustering for mGBP9<sub>2GTP</sub>. In the clustering analysis, compared to the monomer, the GC is slightly stabilized in the dimer, while L1 is destabilized (Fig.4.2AB). We know from the study by Wehner et al.[198] that the guanine cap (GC) is involved in dimerization.

However, upon binding of GTP, the GC is more flexible in the dimer than in the monomer. Upon binding of GTP, loop 1 (L1) and helix  $\alpha 4$  are destabilized, as evidenced both by RMSF and clustering. This corroborates the theory of Ince et al. [199], that in the GTP-bound state,  $\alpha 4$  releases its lock on  $\alpha 12$ , allowing the latter, long helix to swing free. Comparing the monomeric and dimeric state, the E domain and CT tail are more rigid in the dimer structures. Additionally, the M domain is also stabilized compared to the monomer, except for the tip regions. Also, L1 and SW1 are very flexible in the mGBP9<sub>1GTP</sub> and mGBP9<sub>2GTP</sub> dimer, in comparison to the monomer. The GC keeps its flexibility even in the dimer, but is reduced compared to the monomer. GTP also causes a stronger hinge motion, which is true for both monomer and dimer. One monomer has a stronger hinge motion in all three dimers, however, in mGBP9<sub>2GTP</sub>, the strongest hinge motion was found. In total, the amplitudes and directions of the hinge motions are comparable across dimers and monomers. GTP on the other hand, acts as an allosteric effector, likely in a motion relayed via SW2 and  $\alpha 4$ . The SW2 in mGBP9 contains two major difference to mGBP2 and mGBP7: (i) G→S105 and (ii) D→N 106, while the rest is like mGBP7. Thus, the discussed residue S105 in mGBP9, observed to mediate the *trans*-GTP-interaction, is unique in mGBP9, and in mGBP2/hGBP1 the distance between SW2 and GTP in the dimer is too high for it to occur.

## Discussion

The SAXS structure of the mGBP9 monomer provided by Jens Reiners and Sander Smits is in good agreement with the AlphaFold prediction. Dimerization of mGBP9 was observed only when adding GTP [200, p.190], and the average radius of gyration increased with GTP concentration (Jens Reiners, unpublished work). Similar to hGBP1 and mGBP2, a GD-GD dimer is assumed by mGBP9, although the stalks are more angled and less linear. A possible explanation for the GTP-induced dimerization is the highly negatively charged electrostatic potential surface (calculated by Jennifer Loschwitz), where GTP might have a screening effect.

Dimerization and GTP-loading especially affect the flexibility of SW2 and the GC. When measuring the distance between GTP and the opposing SW2 region, the two can be observed approaching each other. While GTP might stabilize some motifs, it also destabilizes other GD motifs, resulting in an induced fit situation. (Fig. 4.2D)

mGBP9 is similar structurally and sequentially to mGBP7, but the dimer structure is more similar to the hGBP1/mGBP2 GD-GD dimer. mGBP2 also dimerizes only in

presence of GTP, while mGBP7 can dimerize in its absence [190]. For many other GTPases, polymerization stabilizes the GD, enabling hydrolysis [25], [201]–[203].

In summary, our work on mGBP9 is in agreement with experimental observations, and fits into our previous studies of mGBP2 and mGBP7, while underlining the great diversity of the seemingly so similar mGBPs. The unique role of the S105 and potential cooperativity of GTP hydrolysis, as well as mGBP9 colocalization patterns remain interesting objects for further study.

## 4.2 Hinge Motion

In order to transfer our knowledge of the BDLP hinge motion to GBPs, we started a steered MD simulation with the goal of pulling the known GBP structure towards more closed or more open conformations. This was partly inspired by an electron density provided by Jens Reiners and Sander Smits (Fig. 4.3A).

### Methods

GROMACS offers a pull code that can be invoked via the mdp file, however, the box size requirements are quite large. When using the plugin plumed [159]–[161], the setup, energy minimization and equilibration still go according to GROMACS protocol. In the production run, a normal mdp file can be used as well, but the simulation has to be run with

```
1 | gmx mdrun -s md.tpr -deffnm pull -plumed plumed.dat
```

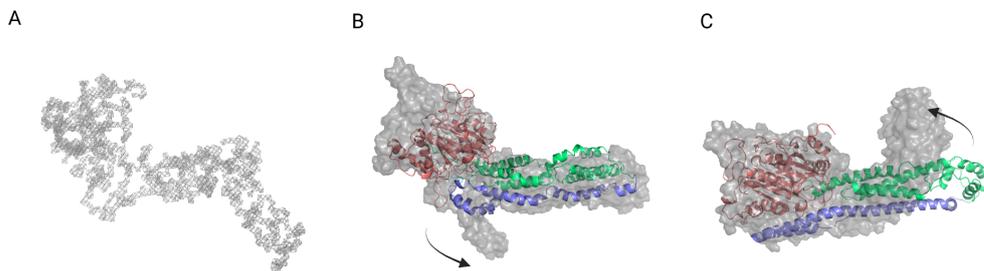
where the file plumed.dat needs to be in the same directory, and an example for its contents can be found in App. C.3.2.

In this case, hGBP1 was simulated in MARTINI 2, with domain-internal ElNeDyn [150] to avoid overstabilization across domains. One pull vector was defined between the stalk tip and the 80 C-terminal residues (Fig. 4.3B), with the goal to obtain a fully open conformation. Even with forces of up to 150,000 kJ/mol/nm, the  $\alpha$ 12 helix did not fully detach. For the future, a new order parameter (closer to the stalk tip) will be needed to loosen the entire M-E-domain-interactions like a zipper. An analogous setup with a different vector, defined between the distal stalk portion

and the G-domain closest to it is shown in Fig. 4.3C, had the goal to generate a fully closed form. Here, forces of 100,000 kJ/mol/nm were reached, and a fully closed form was assumed.

## Results

As seen in Fig.4.3B, the pulling towards an open did not yield the latter. Furthermore, during backmapping, the detached C-terminal residues reattach within 100 ns. However, after the backmapped simulation, we can still observe the expected detachment of the  $\alpha 4'$ - $\alpha 12$  interaction, based more in a hinge2 motion at the GD. Also, a break in the  $\alpha 12$ -helix forms at N513, possibly facilitating a hinge1 motion towards the GD. The pulling simulation towards an extreme closed form also reverts back towards the X-ray structure during the AA simulation. Here, the distal stalk end curves down between residues 409 and 502, which is distinct from the conformation before pulling. Additionally, the region 180-200 increases its distance to the  $\alpha 4'$ - $\alpha 12$  stub. One takeaway from this is that the  $\alpha 4'$ - $\alpha 12$  saltbridge loosening, and the hinge motion observed by us [88], [89], [197] are connected phenomena, as previously suggested by Sistemich et al. [204], [205].



**Fig. 4.3.:** A: unrefined SAXS electron density provided by Jens Reiners. B: Structure after CG pulling along vector1, in grey surface. Structure after backmapping to AA and 100 ns simulation, in colored cartoon. C: Structure after CG pulling along vector2, in grey surface. Structure after backmapping to AA and 100 ns simulation, in colored cartoon.

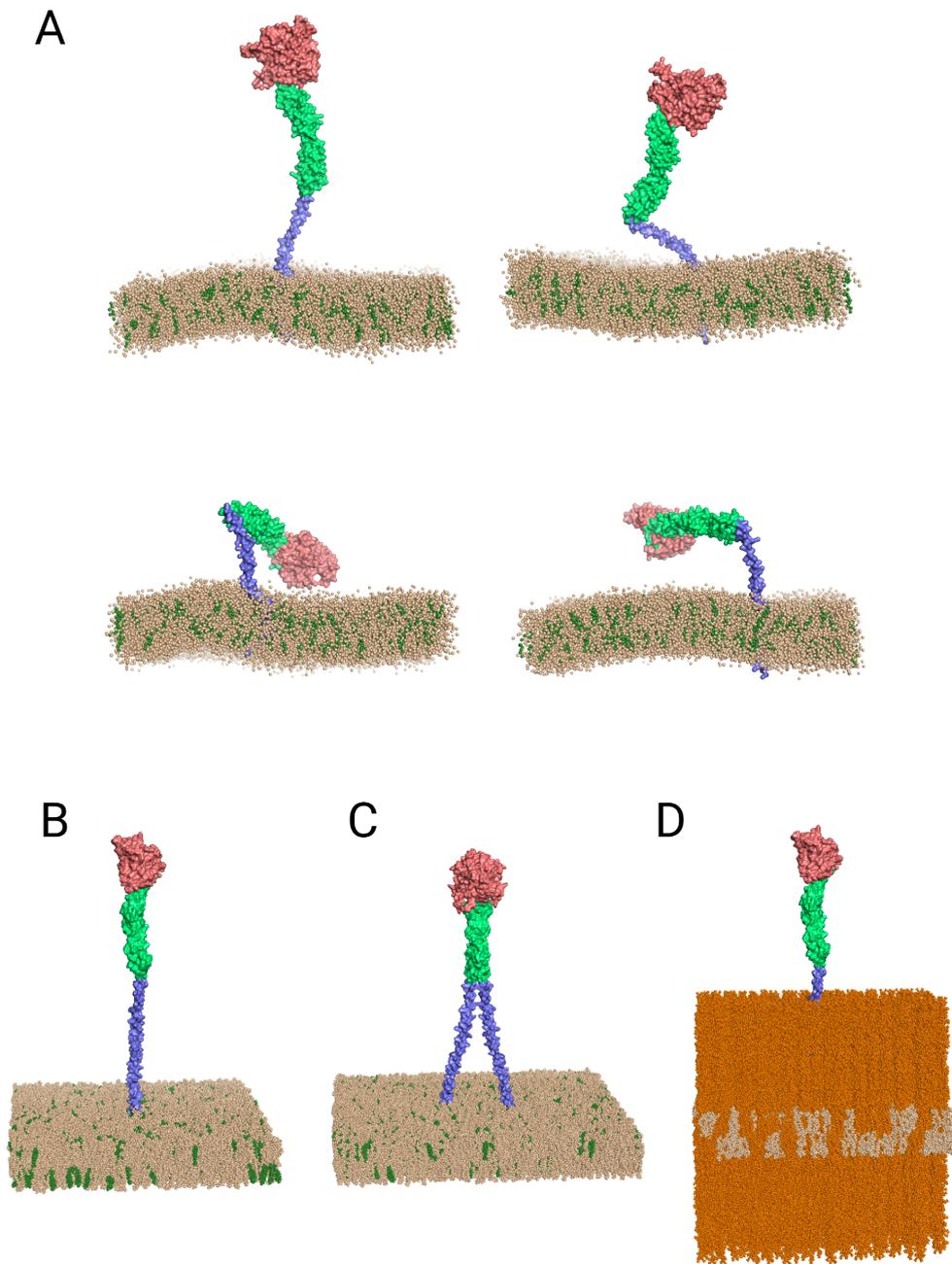
Due to time limitations, this project is not finished, and the initial proof of concept should be supported by future all-atom pulling simulations of GBPs, perhaps combining the two pull vectors.

## 4.3 Membrane

We know from Loschwitz et al. [89] that GBP monomers already bind the membrane, but were intrigued by the possibility of a pincushion-model, as suggested by Cui et al. and Zhu et al. [189], [206]. In this, the same membrane-anchoring domain (isoprenyl anchor or CT tail) is used, but the open conformation is assumed by the GBP.

An initial  $1\ \mu\text{s}$  coarse-grained trial run of mGBP5, with folded-out  $\alpha 12$ , in a DOPC/c-cholesterol membrane shows great promise in curving the membrane. However, helix  $\alpha 12$  is still not as stable as expected, and forms a kink where it emerges from the membrane.

We therefore plan to observe hGBP5 in all-atom DOPC/Chol membrane, as monomer, dimer, and with added sugars. From literature, we know that *Shigella flexneri* and *Salmonella enterica* are coated with LPS [207], while *T.gondii* is coated in glycolipids [208]–[211], all of which have been shown to interact with GBPs. Furthermore, parameters for these lipids are available via CHARMM-GUI. Especially the glycolipid system looks promising, due to its potential stabilizing effect on the  $\alpha 12$  helix (Fig. 4.4).



**Fig. 4.4.:** Different setups of GBPs interacting with DOPC/Chol membranes. A: mGBP5 directly in the membrane, selected snapshots from a  $1\ \mu\text{s}$  trajectory. B: setup of hGBP5 with farnesyl-anchor, C: hGBP5 dimer with farnesyl-anchor, D: hGBP5 with farnesyl-anchor in a glycosylated membrane

## Conclusion

When making the inevitable comparison between BDLP and GBPs, at first glance, it seems that no hinge1 exists in GBPs, only hinge2. Something akin to a hinge1 motion can be achieved by kinking the M/E-domain of GBPs in half, as extensively shown by Loschwitz et al. for hGBP1, mGBP2, mGBP7 and mGBP9 [89]. But we can now take it one step further and correlate this motion with the predicted loosening of the  $\alpha 4'$ - $\alpha 12$  interaction, since our pulling simulation shows they coincide, maybe even cause each other. This opens up the conformational space for GBPs considerably, as previously suggested by several authors [204], [206], [207], [212]. The motion of  $\alpha 12$  brought forward by Voepel et al. has the added benefit of similarity to the shear motion we observed for BDLP.

Why has this conformational change been so hard to produce for GBPs? It was not for lack of trying, since  $\alpha 12$  has been shown to be unstable on its own [88]. Taking a step back, we now suggest that either dimerization (Fig. 4.4), or interaction with other molecules (possibly glycolipids, possibly SRS29C or ISG15, see Sec.5) is needed to stabilize this open form.

Meanwhile, SAXS results from Jens Reiners and Sander Smits add a missing puzzle piece, a half-open structure, was revealed in mGBP9, which could be an intermediate before the fully extended structure. This would also fit with our proposed cycle for BDLP (Fig. 3.4).

Though the exact cycle of GBP action still has many degrees of freedom left, we can turn to dimers as the next relevant unit. The abundance of GBPs (with 11 homologs in mice and 7 in humans [213]) could be a product of a finely tuned process, where different multimer arrangements are triggered by the nucleotides, proteins and lipids present. The small sequence variations between isoforms could lead to heterooligomers where one unit acts as chain starter, some as elongators, and some as chain breakers. For example, mGBP2 attaches to the PVM first, followed by mGBP7, and mGBP7 deletion is more lethal than mGBP2 deletion [71]. Perhaps mGBP7's various polymerization options allow a larger carpet of GBPs to form, which can then more effectively remodel membranes, or recruit autophagosomes through signal amplification.

Additionally, membranes could be shaped using a scaffolding-effect or a bilayer-couple mechanism, they could be tethered and fused, or tightened and fissured, as discussed by us in [197]. But DSPs are not limited to immune reactions against *T.gondii*, but are effective against *Chlamydia*, influenza and *Salmonella* as well [35], [38], [214]. The different isoforms could be adapted to different organisms as well, explaining their variety.



## Interaction partners

Working in tandem with the group of Klaus Pfeffer and Daniel Degrandi, an opportunity presented itself to study interaction partners of GBPs, as identified by co-immunoprecipitation [67]. Herein, proteins are isolated by binding to specific antibodies. If a stable protein complex exists, all members of the complex can be precipitated together, and then identified by mass spectrometry, SDS-PAGE or Western Blot [215]. Among the identified proteins interacting with mGBP7, were interferon stimulated gene 15 (ISG15), SAG related sequence 29C (SRS29C) and target of Myb1 (TOM1), which will be the subject of the following sections.

### 5.1 Methods

#### Initial simulations

After obtaining the structures of the interaction partners from the PDB or via homology modeling, ISG15, SRS29C and TOM1 were all solvated alone in dodecahedral boxes with at least 1 nm distance to the edges, minimized and equilibrated. No additional constraints were used for the production run. In the case of ISG15 and SRS29C, the full sequence was simulated for 100 ns using AMBER99SB\*ILDNP as force field, with TIP3P water, then post-translational processing was applied, (followed by a shorter 50 ns simulation), while TOM1 was simulated for 100 ns directly. A typical mdp file can be found in the appendix in Sec. C.2.3. These simulations were performed to obtain equilibrated structures of the mGBP7 interaction partners, which could then be used for docking.

#### Docking

In the choice of docking programs, the CAPRI competition and a review of docking programs by Porter et al. were considered [216], with the goal to apply the most promising one. For an initial rigid docking, the servers PatchDock [167], Zdock

[169], Firedock and Swarmdock and ClusPro [174] were used, from which the most frequent contacts were compiled into restraint sets, which were then evaluated with HADDOCK [175]. Furthermore, using several servers enables comparison of the results to ensure consistency. Since these experiments were conceived in 2020, more recent progress made by AlphaFold-Multimer docking is not included yet.

## 5.2 ISG15

ISG15 is a ubiquitin-like protein, consisting of two subunits, similar to a fused ubiquitin dimer. After maturation, the 165 amino acid sequence ends C-terminally in an LRLRGG motif, which becomes covalently linked to a lysine of the target protein, a process termed ISGylation [217]–[219]. Both ubiquitination and ISGylation mark proteins for degradation, and ISG plays a role in several immune responses [220]–[222].

The ISG15 model was obtained from the PDB (5TLA) [223], originating from the mouse. Only two amino acids had to be added in PyMol to complete the C-terminal LRLRGG-sequence, the demethionation at the N-terminus accurately depicts the posttranslational modifications of the protein.

For the docking, the CT-tail of mGBP7 was removed, since we expect it to be embedded in the membrane. Several servers (listed above) plausibly predicted that ISG15 would face mGBP7 via its GG motif. Overall, the docking servers localize ISG15 in a disc around the G-domain. The most frequent binding areas could be easily mapped to lysines present on mGBP7, without that information having been given to the docking servers. From these lysines, 8 restraint sets for HADDOCK were generated (see Tab. 5.1).

**Tab. 5.1.:** ISG15 HADDOCK restraint sets

lysine residue	description	best HADDOCK score
557,565,580,586	close to CT tail	-55
381,524,533,539	middle of stalk	-59
452,508,510	distal end of stalk	-58
479,486	stalk tip	-43
26,88,90,329	G-domain, base of stalk	-53
266,277	G-domain, opposing base of stalk	-69
241,243	G-domain, top	-63
188,216	G-domain, opposite of stalk tip	-81

Of these sets, the best-performing one was the G-domain area, and the resulting two best poses were simulated for 100 ns, showing no dissociation (see Fig. 5.2). The docking of an, in reality, covalent attachment, is an artificial setup, but the previous knowledge of the lysine-directed modification serves to benchmark our docking method. Since ubiquitination has been proven to be possible on multiple sites for the same protein [224], several of the ISGylation sites could be correct, which would explain their similar scores.

## 5.3 SRS29C

SRS29C is a *T.gondii* surface protein, downregulating virulence to improve persistence in the host [225].

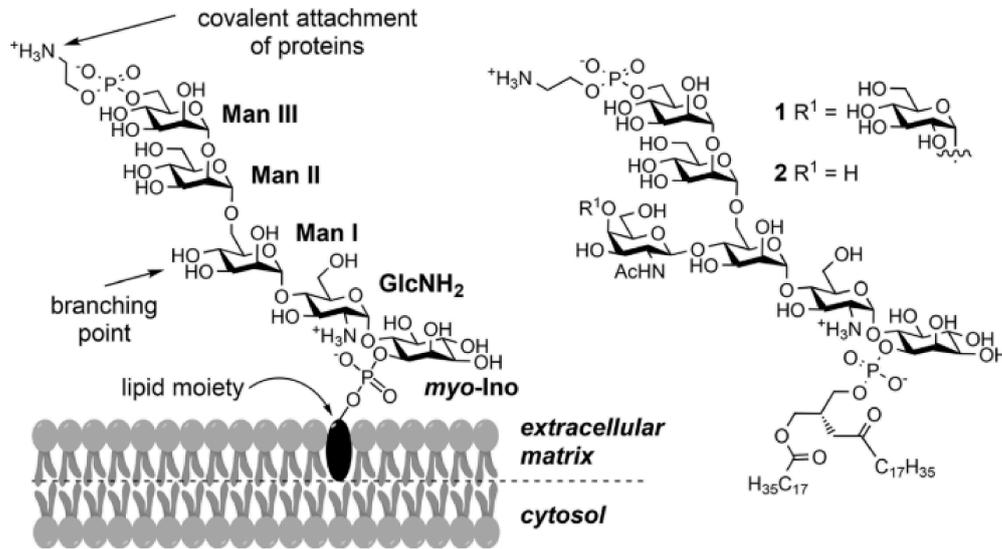
SRS29C undergoes cleavage of the first 57 amino acids (confirmed by SignalP5.0 [226]), constituting a Sec signal peptide, and cleavage of the last 24 amino acids, to reveal S348 which is conjugated to a GPI-anchor to mediate membrane binding. The 24 amino acids are likely to function as a membrane-anchor before their replacement, according to ProtScale, Phobius and PredictProtein ([227]–[229]).

The GPI-anchor is linked to the protein via an ethanolamine-phosphate, followed by a branched chain of sugars, and ending in phosphatidylinositol, which is then linked to a diacylglycerol (Fig. 5.1), most likely containing C16/C18 fatty acids [209]–[211].

The protein part of SRS29C could be modeled with Phyre2 [230] and SWISSMODEL [231] at 100% confidence, albeit only for residues 57-310. The model generated by I-TASSER for residues 1-375 only achieves a C-score of -2.34, using the template 1KZQ, the surface antigen 1 (SAG1) of *T.gondii*. Despite the rather bad score (+5 would be ideal), the protein was remarkably stable during subsequent simulation, probably due to high conservation among the SAG related sequences (SRSSRSs). SRS29C showed a clear preference for the "underside" of mGBP7, which corresponds to helices  $\alpha_{12/13}$ . We know from [190] that  $\alpha_{13}$  mediates membrane interaction, which should create spatial proximity to membrane-bound SRS29C. As illustrated in 5.2, the two best docking poses offer themselves for a location atop the membrane. Additionally, the docking was performed with and without the CT-tail present in this case, and its inclusion leads to much better scores (see Fig. 5.4).

mGBP7 very likely inserts its CT tail into the membrane as an anchor [190], which would no longer be possible with SRS29C's length exceeding that of the CT tail in model 2. While Wasmuth et al. [232] propose a "lollipop" model for the SAG-related

proteins, Banerjee et al. [233] show that a "flop-down" is entirely possible, allowing both models to coexist, and solving the insertion problems. During 100 ns simulation in solution, the two complexes remain stable.



**Fig. 5.1.:** GPI-anchor structure. Conserved pseudopentasaccharide core structure of GPIs and *T.gondii*-specific glycolipids 1 and 2. GlcNH<sub>2</sub> = D-glucosamine, Ino=inositol, Man=D-mannose. Figure taken from Götze et al. [211], reproduced with permission.

**Tab. 5.2.:** SRS29C HADDOCK restraint sets

restraint	mGBP7	SRS29C	descr. mGBP7	descr. SRS29C
1	335-339	201,202	stalk	interdomain
2	561-565	61-72	CT tail	Nter
3	157,588,589	254,258, 259,275	CT tail	domain 2
4	207-215	61-72,141	GD side	domain 1
5	68-70,245-246	61-72,141	GD, opposite side	domain 1
6	26,27	201,202	GD	interdomain
7	101-108	333-339	GD, top	Cter
8	207-215,265	349,352, 213,333-339	GD	domain 2, Cter
9	311,157,521,374, 375,588,589	61-72,141	CT tail	domain1

## 5.4 TOM1

TOM1 is an agent in membrane trafficking. It contains a VHS domain – a composite name after proteins Vps27 (vacuolar protein sorting 27), Hrs (hepatocyte growth factor-regulated tyrosine kinase substrate) and STAM (signal transducing adapter molecule) [234]– which further contains a ubiquitin interacting motifs (UIMs), to sort ubiquitinated proteins into endosomes, which later become lysosomes [235]. A basic patch in the VHS domain is thought to mediate membrane interaction [236], probably localized between amino acids 30 and 80 in TOM1.

The template used for homology modeling was 1ELK, the human TOM1. The second isoform of murine TOM1 could be modeled with 100% confidence for the residues 2-152 (in Phyre2 [230]), lacking 364 residues at the C-terminus. These were modeled with I-TASSER [237], [238], achieving a C-score of -0.38 (where +5 would be an optimal value). While the score is not optimal, it should be noted that in the case of TOM1 the first 150 amino acids were an exact match to the template and the last 360 amino acids accordingly showed great flexibility in a 100 ns test MD simulation. Here, the secondary structure was mostly stable, but the ordering of the helices changed from an ordered crescent shape to a criss-cross-shape (see Fig. 5.2). In parallel to ISG15 and SRS29C, the multiple docking results were reduced to restraint sets, which were then evaluated with HADDOCK.

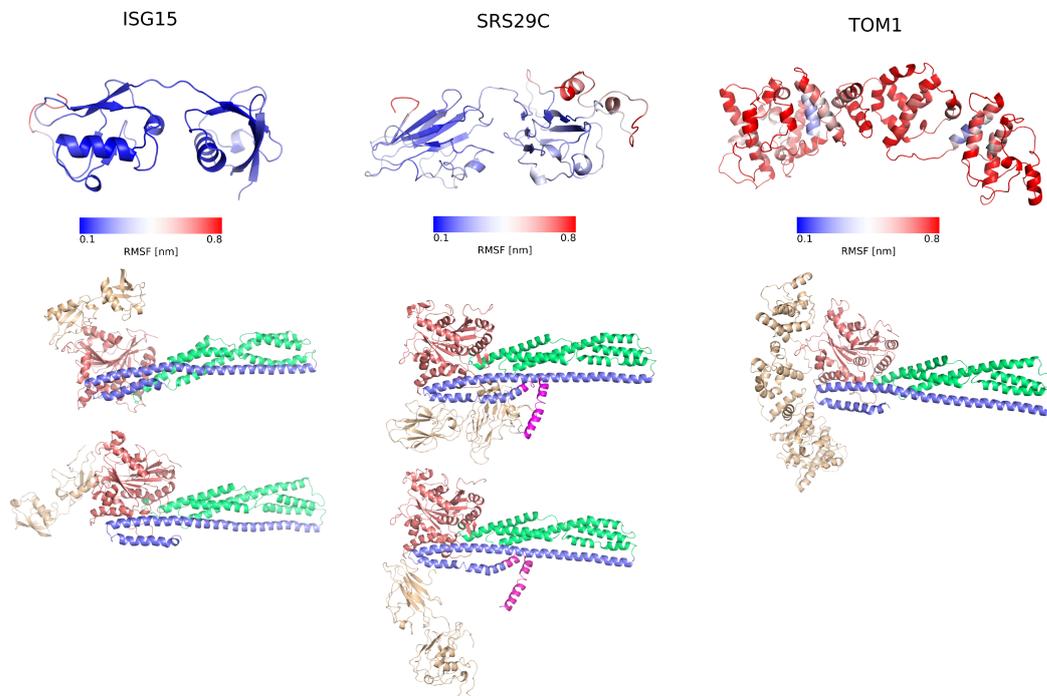
**Tab. 5.3.:** TOM1 HADDOCK restraint sets

restraint	mGBP7	TOM1	descr.
1	131,185,186	90,92,93,96,129,131,130	GD
2	131,185,186	305,266,267	GD
3	457,458,461,468,500	162,129,31,211	stalk tip
4	201,208,209, 210,211,212	199,202,203,200,201, 211,212,49,57	GD
5	346,347	152,153,154,155,156,158,159	stalk
6	346,347	199,202,203,200,201,211,212	stalk
7	304	162,129,31	GD
8	457,458,461,468,500	152,153,154,155, 156,158,159	stalk tip

The docking results indicate that TOM1 binds preferably to the G-domain of mGBP7, and this pose was stable during a 100 ns simulation. As with ISG15, the CT tail was excised from mGBP7, since we do not expect TOM1 to bind closely to the PVM, but rather to the opposing membrane, for fusion with the endosome.

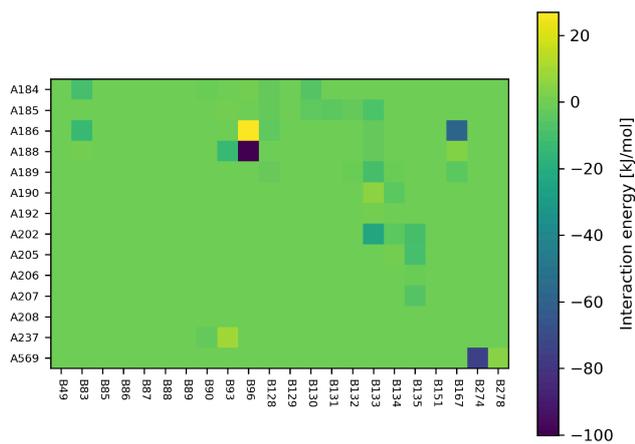
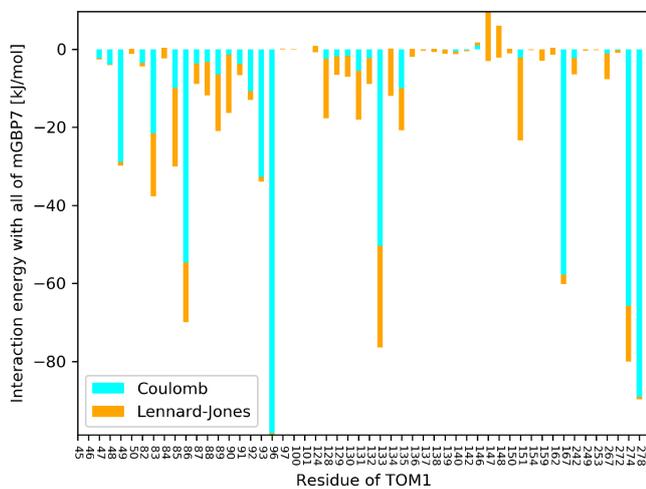
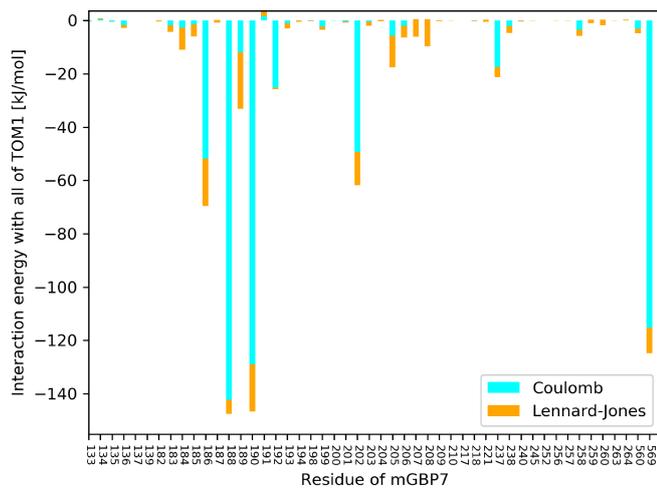
Subsequently, the best pose was simulated with included energy decomposition (via

GROMACS energy groups), to get exact contributions of all amino acids close to the interface. This required three separate runs, because GROMACS allows a maximum of 64 energy groups.



**Fig. 5.2.:** Flexibility of the mGBP7 interaction partners (in terms of RMSF, see color scale) during simulations of the interaction partners alone (top) and resulting complex with mGBP9 (bottom).

The contributing residues were identified by a mix of contact frequency, contact distance, and energy decomposition. In the first simulation, the residues of mGBP7 were split up and their interaction energy (both Coulomb and Lennard-Jones) with all of TOM1 was calculated (Fig. 5.4). In the second simulation, this was reversed. From this, the residues contributing the most were chosen for a residue-residue matrix energetic analysis. While the matrix confirms several strong residues as pairwise interactors, it is possible that pairs with asymmetric strength of interaction were filtered out at this step, and can only be inferred from the bar graph together with structural analysis.

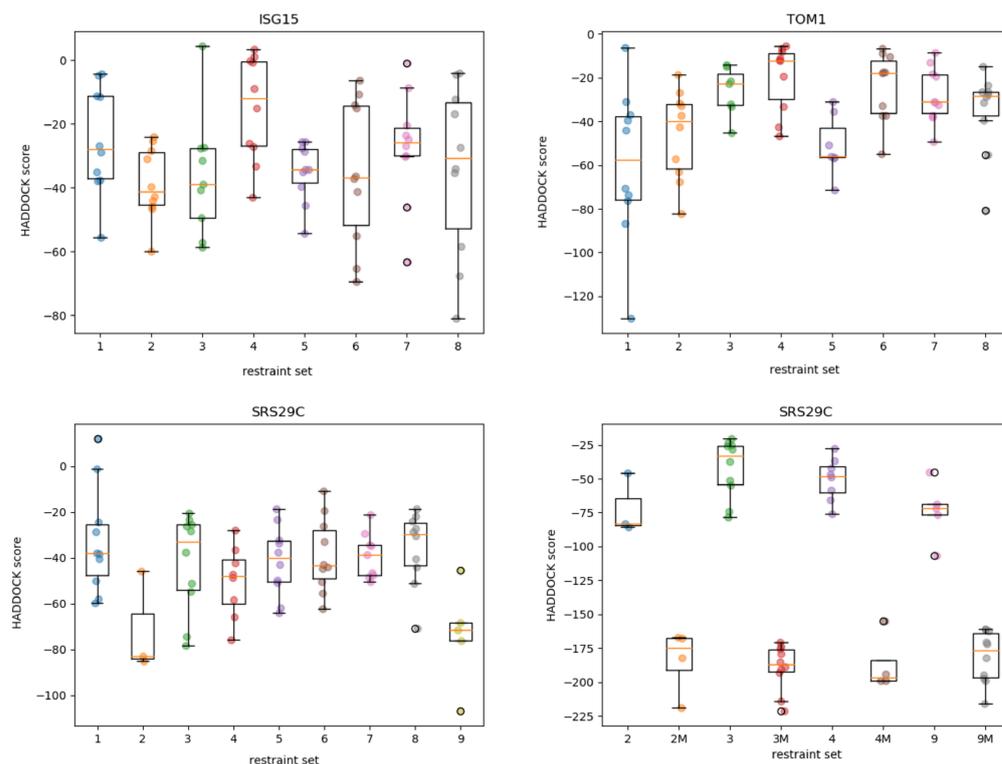


**Fig. 5.3.:** Interaction energies of mGBP7 (chain A) and TOM1 (chain B) in their complex. A selection of the residues contributing the most was used for the matrix, in which interaction energy is the sum of Coulomb and Lennard-Jones energy.

After considering all these aspects, the residues most involved in the binding process are listed in the table below:

**Tab. 5.4.:** Residues contributing the most towards mGBP7-TOM1 interaction

Protein	Residues
mGBP7	F183 V184 L185 Q186 L187 K188 L189 Q190 R238 P239 T240 L257
TOM1	K48 D49 R84 F85 H86 V87 L88 F128 R129 S130 S131 P132 D133 T135



**Fig. 5.4.:** Distribution of HADDOCK scores for different systems of restraints. The corresponding restraint systems are explained in Tabs. 5.1,5.2, 5.3

## 5.5 Conclusion

In this chapter, we have shown that both ISG15 and TOM1 are likely to interact with the GD of mGBP7, which is fitting, for they are also expected to interact with each other. Looking back at Fig. 4.4, this might be the only part of GBPs accessible after binding to a glycosylated membrane. As seen in Fig. 4.1, this decoration of GBPs with ISG15 is incompatible with GD-dimerization, perhaps explaining why hGBP1

is not ISGylated [217], while mGBP7 dimers leave the GD free for modification. Ubiquitination of the host-defending GBPs has been shown as a strategy of the invading pathogen [239], resulting in degradation of the GBPs by the proteasome. ISGylation of GBPs, however, happens already in absence of *T.gondii* infections, although it increases by a factor of five after infection [67]. Especially K373 (at the junction of helices  $\alpha 8$  and  $\alpha 9$ ) ISGylation increases by a factor three. We do not know if this impairs or facilitates mGBP7 hinge motion. Free ISG has important immune functions [221], [222], so its binding to mGBP7 could just as well sequester it away from those functions, as a pathogen strategy.

It is known that TOM1 interacts with quite a few proteins of the post-Golgi trafficking or vesicle reshaping machinery (Arf1, COPA, Gate16, Clathrin, Tollip, Endofin, Myosin 6), suggesting that Tom1 recruits autophagosomes to endosomes (the endosome in this case could contain *T.gondii*) [67]. We hope to have our TOM1 interaction sites verified or falsified by experimental results from our collaborators Sophia Kasbrink, Daniel Degrandi and Klaus Pfeffer this year.

As for SRS29C, the question can be raised: Could an envelope of GPI-anchors and the attached SAGs stabilize GBPs in the pincushion-conformation? We do not know yet if GBPs bind only the host-derived PV membrane, or also its inlying *T.gondii* membrane, and if both these binding modes would be identical. Moreover, thus far the two membranes could not be experimentally separated yet, so that their exact composition is still unknown. In conclusion, the complex decorations of glycolipids and lipopolysaccharides to biological membranes is becoming more important for future work. Parameters for such constructs are available in CHARMM-GUI [115] and MARTINI [233], [240], enabling a continuation of this project in this direction.



## Conclusion

In this work, we have resolved a dynamin-superfamily hinge motion for the first time, using ca.  $180\mu s$  of USMD simulations, supported by explorative HREMD and coarse-grained simulations (App. A). The hinge motion is enabled by GTP-binding-induced changes, which are relayed to the stalk domain, and also influence the membrane-binding paddle domain. Membrane-binding and dimerization together then stabilize the open form of BDLP. Importantly, the hinge motion was found to be a shearing motion, which distributes the amount of energy needed over a greater time span. The interplay of transmembrane helix insertion, hinge motion/constriction and membrane thinning could cause the destabilization necessary for membrane fusion or fission. Knowing that GTP hydrolysis only occurs when one full helical turn of BDLP around the invaginated membrane is completed, leading to polymer dissociation, raises the question: how does the dynamin hinge motion happen? We observe the hinge motion barrier as similar in height to the energy released by GTP hydrolysis, but the latter must only occur after the conformational change already took place.

When transferring our observations from BDLP to GBPs, some preliminary results suggest that dimerization and membrane binding also stabilize the previously unstable long helix  $\alpha 12$ , leading to the extended open form proposed in the literature [204]. Our pulling simulations further suggest that the kinking we observed in the middle of the M/E domain is correlated with this extension. Again, small changes in the GTPase domain motifs could be responsible for larger motions. In the case of mGBP9, we saw that GTP binding stabilizes the dimer by causing an induced fit via *trans* interaction with switch II. GBP dimerization could play a part in the loosening of saltbridges, which also contributes to an opening. Different dimer models and modes of membrane deformation have been proposed in our published work App. B. In the future, GBP micelles and pincushion models, or multimers on the membrane should be pursued.

We additionally studied the interaction of GBPs with other proteins and predicted docking sites for experimental validation. SRS29C likely functions as one of the antigens recruiting GBPs to *T.gondii*. ISG15, ubiquitin and TOM1 point towards the subsequent recruiting of endosomes, leading to autophagy of the invading pathogen.

The conformation GBPs assume as multimeric coatomers on the PVM could influence which parts of the protein are accessible for such interactions. Bridging the gap between atomistic level details and macroscopic effects remains an interesting challenge, best addressed in a collaborative effort.

## References

- [1]D. Casares, P. V. Escribá, and C. A. Rosselló, “Membrane lipid composition: Effect on membrane and organelle structure, function and compartmentalization and therapeutic avenues”, *International Journal of Molecular Sciences*, vol. 20, no. 9, 2019 (cit. on p. 1).
- [2]A. A. Spector and M. A. Yorek, “Membrane lipid composition and cellular function”, *Journal of Lipid Research*, vol. 26, no. 9, pp. 1015–1035, 1985 (cit. on p. 1).
- [3]G. Van Meer, D. R. Voelker, and G. W. Feigenson, “Membrane lipids: Where they are and how they behave”, *Nature Reviews Molecular Cell Biology*, vol. 9, no. 2, pp. 112–124, 2008 (cit. on p. 1).
- [4]B. Fadeel and D. Xue, “The ins and outs of phospholipid asymmetry in the plasma membrane: roles in health and disease”, *Critical reviews in biochemistry and molecular biology*, vol. 44, no. 5, p. 264, 2009 (cit. on p. 1).
- [5]T. Harayama and H. Riezman, “Understanding the diversity of membrane lipid composition”, *Nature Reviews Molecular Cell Biology*, vol. 19, no. 5, pp. 281–296, 2018 (cit. on p. 2).
- [6]D. of Education Open Textbook Pilot Project, the UC Davis Office of the Provost, and t. C. S. U. A. L. S. P. the UC Davis Library, *13.3: Membranes and Membrane Lipids - Chemwiki* (cit. on p. 2).
- [7]S. J. Singer and G. L. Nicolson, “The Fluid Mosaic Model of the Structure of Cell Membranes”, *Science*, vol. 175, no. 4023, pp. 720–731, Feb. 1972 (cit. on p. 3).
- [8]D. Lingwood and K. Simons, “Lipid rafts as a membrane-organizing principle”, *Science*, vol. 327, no. 5961, pp. 46–50, 2010 (cit. on p. 3).
- [9]E. Sezgin, I. Levental, S. Mayor, and C. Eggeling, “The mystery of membrane organization: Composition, regulation and roles of lipid rafts”, *Nature Reviews Molecular Cell Biology*, vol. 18, no. 6, pp. 361–374, 2017 (cit. on p. 3).
- [10]H. J. Risselada and S. J. Marrink, “The molecular face of lipid rafts in model membranes”, *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 45, pp. 17 367–17 372, 2008 (cit. on p. 3).
- [11]J. P. Overington, B. Al-Lazikani, and A. L. Hopkins, “How many drug targets are there?”, *Nature reviews. Drug discovery*, vol. 5, no. 12, pp. 993–996, Dec. 2006 (cit. on p. 3).
- [12]H. T. McMahon and J. L. Gallop, *Membrane curvature and mechanisms of dynamic cell membrane remodelling*, Dec. 2005 (cit. on pp. 3, 44).

- [13]J. Zimmerberg and M. M. Kozlov, “How proteins produce cellular membrane curvature”, *Nature Reviews Molecular Cell Biology*, vol. 7, no. 1, pp. 9–19, Jan. 2006 (cit. on pp. 3, 44).
- [14]O. Daumke, A. Roux, and V. Haucke, “BAR domain scaffolds in dynamin-mediated membrane fission”, vol. 156, no. 5, pp. 882–892, 2014 (cit. on pp. 3, 4, 44).
- [15]F. Campelo, H. T. McMahon, and M. M. Kozlov, “The Hydrophobic Insertion Mechanism of Membrane Curvature Generation by Proteins”, *Biophysical Journal*, vol. 95, no. 5, pp. 2325–2339, Sep. 2008 (cit. on p. 3).
- [16]K. M. Davies, C. Anselmi, I. Wittig, J. D. Faraldo-Gómez, and W. Kühlbrandt, “Structure of the yeast F<sub>1</sub>F<sub>0</sub>-ATP synthase dimer and its role in shaping the mitochondrial cristae”, *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, no. 34, pp. 13 602–13 607, Aug. 2012 (cit. on p. 3).
- [17]M. H. Stowell, B. Marks, P. Wigge, and H. T. McMahon, “Nucleotide-dependent conformational changes in dynamin: evidence for a mechanochemical molecular spring”, *Nature cell biology*, vol. 1, no. 1, pp. 27–32, 1999 (cit. on pp. 3, 4).
- [18]Y. J. Chen, P. Zhang, E. H. Egelman, and J. E. Hinshaw, “The stalk region of dynamin drives the constriction of dynamin tubes”, *Nature structural & molecular biology*, vol. 11, no. 6, pp. 574–575, 2004 (cit. on p. 3).
- [19]A. Roux, K. Uyhazi, A. Frost, and P. De Camilli, “GTP-dependent twisting of dynamin implicates constriction and tension in membrane fission”, *Nature*, vol. 441, no. 7092, pp. 528–531, 2006 (cit. on p. 3).
- [20]M. M. Kozlov, F. Campelo, N. Liska, *et al.*, “Mechanisms shaping cell membranes”, *Current Opinion in Cell Biology*, vol. 29, no. 1, pp. 53–60, Aug. 2014 (cit. on pp. 3, 4).
- [21]S. Morlot, V. Galli, M. Klein, *et al.*, “Membrane Shape at the Edge of the Dynamin Helix Sets Location and Duration of the Fission Reaction”, *Cell*, vol. 151, no. 3, p. 619, 2012 (cit. on p. 4).
- [22]M. G. J. Ford and J. S. Chappie, “The structural biology of the dynamin-related proteins: New insights into a diverse, multitasking family”, *Traffic*, vol. 20, no. 10, pp. 717–740, 2019 (cit. on p. 4).
- [23]G. J. K. Praefcke and H. T. McMahon, “The dynamin superfamily: universal membrane tubulation and fission molecules?”, *Nature Reviews Molecular Cell Biology*, vol. 5, no. 2, pp. 133–147, 2004 (cit. on p. 4).
- [24]K. Melen, T. Ronni, B. Broni, *et al.*, “Interferon-induced Mx proteins form oligomers and contain a putative leucine zipper.”, *Journal of Biological Chemistry*, vol. 267, no. 36, pp. 25 898–25 907, 1992 (cit. on p. 4).
- [25]J. R. Jimah and J. E. Hinshaw, “Structural Insights into the Mechanism of Dynamin Superfamily Proteins”, vol. 29, no. 3, pp. 257–273, 2019 (cit. on pp. 5, 6, 49).
- [26]R. Kalia and A. Frost, “Open and cut: allosteric motion and membrane fission by dynamin superfamily proteins”, *Molecular Biology of the Cell*, vol. 30, no. 17, D. G. Drubin, Ed., pp. 2097–2104, Aug. 2019 (cit. on p. 5).

- [27]B. Antonny, C. Burd, P. De Camilli, *et al.*, “Membrane fission by dynamin: what we know and what we need to know”, *The EMBO Journal*, vol. 35, no. 21, pp. 2270–2284, Nov. 2016 (cit. on p. 6).
- [28]O. Daumke and G. J. K. Praefcke, “Invited review: Mechanisms of GTP hydrolysis and conformational transitions in the dynamin superfamily”, *Biopolymers*, vol. 105, no. 8, pp. 580–593, Aug. 2016 (cit. on pp. 6, 43).
- [29]K. A. Michie, A. Boysen, H. H. Low, J. Møller-Jensen, and J. Löwe, “LeoA, B and C from Enterotoxigenic Escherichia coli (ETEC) Are Bacterial Dynamins”, *PLOS ONE*, vol. 9, no. 9, e107211, Sep. 2014 (cit. on p. 6).
- [30]K. Faelber, S. Gao, M. Held, *et al.*, “Oligomerization of Dynamin Superfamily Proteins in Health and Disease”, *Progress in Molecular Biology and Translational Science*, vol. 117, pp. 411–443, Jan. 2013 (cit. on p. 6).
- [31]J. Oró, “Mechanism of Synthesis of Adenine from Hydrogen Cyanide under Possible Primitive Earth Conditions”, *Nature*, vol. 191, no. 4794, pp. 1193–1194, Sep. 1961 (cit. on p. 6).
- [32]D. Zala, U. Schlattner, T. Desvignes, *et al.*, “The advantage of channeling nucleotides for very processive functions”, *F1000Research*, vol. 6, p. 724, Jul. 2017 (cit. on p. 6).
- [33]D. G. Nicholls and S. J. ( J. Ferguson, *Bioenergetics*. Academic Press, 2002, p. 297 (cit. on p. 7).
- [34]J. M. Berg, J. L. Tymoczko, G. J. Gatto, and L. Stryer, “Stryer Biochemie”, *Stryer Biochemie*, 2018 (cit. on p. 7).
- [35]A. Xavier, M. A. Al-Zeer, T. F. Meyer, and O. Daumke, “hGBP1 Coordinates Chlamydia Restriction and Inflammasome Activation through Sequential GTP Hydrolysis”, *Cell Reports*, vol. 31, no. 7, p. 107667, May 2020 (cit. on pp. 7, 53).
- [36]M. P. Wandel, C. Pathe, E. I. Werner, *et al.*, “GBPs Inhibit Motility of Shigella flexneri but Are Targeted for Degradation by the Bacterial Ubiquitin Ligase IpaH9.8”, *Cell Host & Microbe*, vol. 22, no. 4, 507–518.e5, Oct. 2017 (cit. on p. 7).
- [37]E. Braun, D. Hotter, L. Koepke, *et al.*, “Guanylate-Binding Proteins 2 and 5 Exert Broad Antiviral Activity by Inhibiting Furin-Mediated Processing of Viral Envelope Proteins”, *Cell Reports*, vol. 27, no. 7, 2092–2104.e10, May 2019 (cit. on p. 7).
- [38]Z. Zhu, Z. Shi, W. Yan, *et al.*, “Nonstructural protein 1 of influenza A virus interacts with human guanylate-binding protein 1 to antagonize antiviral activity”, *PloS one*, vol. 8, no. 2, Feb. 2013 (cit. on pp. 7, 53).
- [39]E. Guenzi, K. Töpolt, E. Cornali, *et al.*, “The helical domain of GBP-1 mediates the inhibition of endothelial cell proliferation by inflammatory cytokines”, *The EMBO Journal*, vol. 20, no. 20, p. 5568, Oct. 2001 (cit. on p. 7).
- [40]E. Guenzi, K. Töpolt, C. Lubeseder-Martellato, *et al.*, “The guanylate binding protein-1 GTPase controls the invasive and angiogenic capability of endothelial cells through inhibition of MMP-1 expression”, *The EMBO Journal*, vol. 22, no. 15, p. 3772, Aug. 2003 (cit. on p. 7).

- [41]N. Britzen-Laurent, C. Herrmann, E. Naschberger, R. S. Croner, and M. Stürzl, “Pathophysiological role of guanylate-binding proteins in gastrointestinal diseases”, *World Journal of Gastroenterology*, vol. 22, no. 28, p. 6434, Jul. 2016 (cit. on p. 7).
- [42]Y. Itsui, N. Sakamoto, S. Kakinuma, *et al.*, “Antiviral effects of the interferon-induced protein guanylate binding protein 1 and its interaction with the hepatitis C virus NS5B protein.”, *Hepatology (Baltimore, Md.)*, vol. 50, no. 6, pp. 1727–1737, Dec. 2009 (cit. on p. 7).
- [43]D. Hotter, D. Sauter, and F. Kirchhoff, “Guanylate binding protein 5: Impairing virion infectivity by targeting retroviral envelope glycoproteins”, *Small GTPases*, vol. 8, no. 1, p. 31, Jan. 2017 (cit. on p. 7).
- [44]C. Krapp, D. Hotter, A. Gawanbacht, *et al.*, “Guanylate Binding Protein (GBP) 5 Is an Interferon-Inducible Inhibitor of HIV-1 Infectivity”, *Cell Host & Microbe*, vol. 19, no. 4, pp. 504–514, Apr. 2016 (cit. on p. 7).
- [45]A. R. Bass and S. Shin, “Human GBP1 promotes pathogen vacuole rupture and inflammasome activation during *Legionella pneumophila* infection”, *bioRxiv*, p. 2020.05.27.120477, Jun. 2020 (cit. on p. 7).
- [46]“Guanylate-Binding Proteins Are Critical for Effective Control of *Francisella tularensis* Strains in a Mouse Co-Culture System of Adaptive Immunity”, *Frontiers in Cellular and Infection Microbiology*, vol. 10, Dec. 2020 (cit. on p. 7).
- [47]M. Lipoldová and Y. Sohrabi, “Role of interferon-induced GTPases in leishmaniasis”, *PLOS Neglected Tropical Diseases*, vol. 16, no. 1, D. S. Zamboni, Ed., e0010093, Jan. 2022 (cit. on p. 7).
- [48]“Interferon inducible GBPs restrict *Burkholderia thailandensis* motility induced cell-cell fusion”, *PLOS Pathogens*, vol. 16, no. 3, D. J. Philpott, Ed., e1008364, Mar. 2020 (cit. on p. 7).
- [49]M. W. Black and J. C. Boothroyd, “Lytic Cycle of *Toxoplasma gondii*”, *Microbiology and Molecular Biology Reviews*, vol. 64, no. 3, p. 607, Sep. 2000 (cit. on p. 7).
- [50]J. P. Dubey, N. L. Miller, and J. K. Frenkel, “*Toxoplasma gondii* life cycle in cats.”, *Journal of the American Veterinary Medical Association*, vol. 157, no. 11, pp. 1767–1770, Dec. 1970 (cit. on p. 7).
- [51]J. P. Dubey, *Toxoplasmosis of Animals and Humans*. CRC Press, Apr. 2016 (cit. on p. 7).
- [52]F. Robert-Gangneux and M. L. Dardé, “Epidemiology of and diagnostic strategies for toxoplasmosis”, *Clinical microbiology reviews*, vol. 25, no. 2, pp. 264–296, Apr. 2012 (cit. on p. 8).
- [53]E. J. Melo, M. Attias, and W. De Souza, “The single mitochondrion of tachyzoites of *Toxoplasma gondii*”, *Journal of structural biology*, vol. 130, no. 1, pp. 27–33, 2000 (cit. on p. 8).
- [54]K. Hu, J. Johnson, L. Florens, *et al.*, “Cytoskeletal Components of an Invasion MachineThe Apical Complex of *Toxoplasma gondii*”, *PLOS Pathogens*, vol. 2, no. 2, e13, Feb. 2006 (cit. on p. 8).

- [55]K. A. Joiner, S. A. Fuhrman, H. M. Miettinen, L. H. Kasper, and I. Mellman, “Toxoplasma gondii: Fusion Competence of Parasitophorous Vacuoles in Fc Receptor-Transfected Fibroblasts”, *Science*, vol. 249, no. 4969, pp. 641–646, 1990 (cit. on p. 9).
- [56]E. Maréchal and M. F. Cesbron-Delauw, “The apicoplast: A new member of the plastid family”, *Trends in Plant Science*, vol. 6, no. 5, pp. 200–205, May 2001 (cit. on p. 9).
- [57]J. P. Dubey, D. S. Lindsay, and C. A. Speer, “Structures of *Toxoplasma gondii* tachyzoites, bradyzoites, and sporozoites and biology and development of tissue cysts”, *Clinical Microbiology Reviews*, vol. 11, no. 2, pp. 267–299, 1998 (cit. on p. 9).
- [58]J. Flegr, J. Prandota, M. Soviková, and Z. H. Israili, “Toxoplasmosis A Global Threat. Correlation of Latent Toxoplasmosis with Specific Disease Burden in a Set of 88 Countries”, *PLoS ONE*, vol. 9, no. 3, Mar. 2014 (cit. on p. 9).
- [59]P. R. Torgerson and P. Mastroiacovo, “The global burden of congenital toxoplasmosis: a systematic review”, *Bulletin of the World Health Organization*, vol. 91, no. 7, p. 501, Jul. 2013 (cit. on p. 9).
- [60]G. M. Bhopale, “Pathogenesis of toxoplasmosis”, *Comparative Immunology, Microbiology and Infectious Diseases*, vol. 26, no. 4, pp. 213–222, 2003 (cit. on p. 9).
- [61]A. M. Tenter, A. R. Heckeroth, and L. M. Weiss, “*Toxoplasma gondii*: from animals to humans”, *International journal for parasitology*, vol. 30, no. 12-13, p. 1217, 2000 (cit. on p. 9).
- [62]B. Alberts, A. Johnson, J. Lewis, *et al.*, *Molecular Biology of the Cell*. Garland Science, 2017 (cit. on pp. 9, 10).
- [63]D. Fisch, H. Bando, B. Clough, *et al.*, “Human GBP 1 is a microbespecific gatekeeper of macrophage apoptosis and pyroptosis”, *The EMBO Journal*, vol. 38, no. 13, Jul. 2019 (cit. on p. 10).
- [64]R. T. Gazzinelli, S. Hieny, T. A. Wynn, S. Wolf, and A. Sher, “Interleukin 12 is required for the T-lymphocyte-independent induction of interferon gamma by an intracellular parasite and induces resistance in T-cell-deficient hosts”, *Proceedings of the National Academy of Sciences of the United States of America*, vol. 90, no. 13, pp. 6115–6119, 1993 (cit. on p. 10).
- [65]C. A. Hunter, C. S. Subauste, V. H. Van Cleave, and J. S. Remington, “Production of Gamma Interferon by Natural Killer Cells from *Toxoplasma gondii*-Infected SCID Mice: Regulation by Interleukin-10, Interleukin-12, and Tumor Necrosis Factor Alpha”, *INFECTION AND IMMUNITY*, pp. 2818–2824, 1994 (cit. on p. 10).
- [66]L. Legewie, “Murines GBP7 und interagierende Proteine in der Wirtsabwehr”, PhD thesis, Heinrich-Heine-Universität Düsseldorf, 2020 (cit. on p. 10).
- [67]——, “Funktions- und Strukturanalysen zur Charakterisierung des murinen GBP9 Proteins bei Infektionen mit intrazellulären Erregern”, PhD thesis, Heinrich-Heine-Universität Düsseldorf, 2020 (cit. on pp. 10, 55, 63).
- [68]A. K. Haldar, H. A. Saka, A. S. Piro, *et al.*, “IRG and GBP Host Resistance Factors Target Aberrant, Non-self Vacuoles Characterized by the Missing of Self IRGM Proteins”, *PLoS Pathogens*, vol. 9, no. 6, e1003414, Jun. 2013 (cit. on p. 10).

- [69]M. Yamamoto, M. Okuyama, J. S. Ma, *et al.*, “A cluster of interferon- $\gamma$ -inducible p65 gtpases plays a critical role in host defense against toxoplasma gondii”, *Immunity*, vol. 37, no. 2, pp. 302–313, Aug. 2012 (cit. on p. 10).
- [70]E. M. Selleck, S. J. Fentress, W. L. Beatty, *et al.*, “Guanylate-binding Protein 1 (Gbp1) Contributes to Cell-autonomous Immunity against Toxoplasma gondii”, *PLoS Pathogens*, vol. 9, no. 4, p. 1 003 320, Apr. 2013 (cit. on p. 10).
- [71]N. Steffens, C. Beuter-Gunia, E. Kravets, *et al.*, “Essential Role of mGBP7 for Survival of Toxoplasma gondii Infection”, *mBio*, vol. 11, no. 1, L. David Sibley, Ed., Jan. 2020 (cit. on pp. 10, 53).
- [72]D. Degrandi, C. Konermann, C. Beuter-Gunia, *et al.*, “Extensive Characterization of IFN-Induced GTPases mGBP1 to mGBP10 Involved in Host Defense”, *The Journal of Immunology*, vol. 179, no. 11, pp. 7729–7740, Dec. 2007 (cit. on p. 10).
- [73]D. Degrandi, E. Kravets, C. Konermann, *et al.*, “Murine Guanylate Binding Protein 2 (mGBP2) controls Toxoplasma gondii replication”, *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, no. 1, pp. 294–299, Jan. 2013 (cit. on pp. 10, 45).
- [74]E. Kravets, D. Degrandi, Q. Ma, *et al.*, “Guanylate binding proteins directly attack Toxoplasma gondii via supramolecular complexes”, *eLife*, vol. 5, Jan. 2016 (cit. on pp. 10, 43).
- [75]S. Park, J. Choi, S. B. Biering, *et al.*, “Targeting by Autophagy proteins (TAG): Targeting of IFNG-inducible GTPases to membranes by the LC3 conjugation system of autophagy”, *Autophagy*, vol. 12, no. 7, p. 1153, Jul. 2016 (cit. on p. 11).
- [76]A. K. Haldar, A. S. Piro, D. M. Pilla, M. Yamamoto, and J. Coers, “The E2-Like Conjugation Enzyme Atg3 Promotes Binding of IRG and Gbp Proteins to Chlamydia- and Toxoplasma-Containing Vacuoles and Host Resistance”, *PLOS ONE*, vol. 9, no. 1, e86684, Jan. 2014 (cit. on p. 11).
- [77]J. Choi, S. Park, S. B. Biering, *et al.*, “The parasitophorous vacuole membrane of Toxoplasma gondii is targeted for disruption by ubiquitin-like conjugation systems of autophagy”, *Immunity*, vol. 40, no. 6, pp. 924–935, Jun. 2014 (cit. on p. 11).
- [78]Z. Zhao, B. Fux, M. Goodwin, *et al.*, “Autophagosome-independent essential function for the autophagy protein Atg5 in cellular immunity to intracellular pathogens”, *Cell host & microbe*, vol. 4, no. 5, pp. 458–469, Nov. 2008 (cit. on p. 11).
- [79]A. Khaminets, J. P. Hunn, S. Könen-Waisman, *et al.*, “Coordinated loading of IRG resistance GTPases on to the Toxoplasma gondii parasitophorous vacuole”, *Cellular Microbiology*, vol. 12, no. 7, p. 939, Jul. 2010 (cit. on p. 11).
- [80]B. Clough and E. M. Frickel, “The Toxoplasma Parasitophorous Vacuole: An Evolving Host-Parasite Frontier”, *Trends in Parasitology*, vol. 33, no. 6, pp. 473–488, 2017 (cit. on p. 11).
- [81]S. Besteiro, “Toxoplasma control of host apoptosis: The art of not biting too hard the hand that feeds you”, *Microbial Cell*, vol. 2, no. 6, p. 178, Jun. 2015 (cit. on p. 11).

- [82]M. Müller, K. Katsov, and M. Schick, “A new mechanism of model membrane fusion determined from Monte Carlo simulation”, *Biophysical Journal*, vol. 85, no. 3, pp. 1611–1623, 2003 (cit. on p. 11).
- [83]J. Mattila, A. Shnyrova, A. Sundborger, *et al.*, “A hemi-fission intermediate links two mechanistically distinct stages of membrane fission”, *Nature*, vol. 524, no. 7563, pp. 109–113, 2015 (cit. on p. 11).
- [84]M. Fuhrmans and M. Müller, “Coarse-grained simulation of dynamin-mediated fission”, *Soft Matter*, vol. 11, no. 8, pp. 1464–1480, 2015 (cit. on p. 11).
- [85]M. Pannuzzo, Z. McDargh, and M. Deserno, “The role of scaffold reshaping and disassembly in dynamin driven membrane fission”, *eLife*, vol. 7, 2018 (cit. on p. 11).
- [86]J. Noel, F. Noé, O. Daumke, and A. Mikhailov, “Polymer-like Model to Study the Dynamics of Dynamin Filaments on Deformable Membrane Tubes”, *Biophysical Journal*, vol. 117, no. 10, pp. 1870–1891, 2019 (cit. on p. 11).
- [87]A. Kadosh, A. Colom, B. Yellin, A. Roux, and T. Shemesh, “The tilted helix model of dynamin oligomers”, *Proceedings of the National Academy of Sciences of the United States of America*, vol. 116, no. 26, pp. 2845–12 850, 2019 (cit. on p. 11).
- [88]B. Barz, J. Loschwitz, and B. Strodel, “Large-scale, dynamin-like motions of the human guanylate binding protein 1 revealed by multi-resolution simulations”, *PLOS Computational Biology*, vol. 15, no. 10, P. M. Kasson, Ed., e1007193, 2019 (cit. on pp. 11, 50, 53).
- [89]J. Loschwitz, N. Steffens, X. Wang, *et al.*, “Domain motions, dimerization, and membrane interactions of the murine guanylate binding protein<sup>2</sup>”, *Scientific Reports*, vol. 13, no. 1, pp. 1–20, Jan. 2023 (cit. on pp. 11, 50, 51, 53).
- [90]Y. Chen, L. Zhang, L. Graf, *et al.*, “Conformational dynamics of dynamin-like MxA revealed by single-molecule FRET”, *Nature Communications*, vol. 8, no. 1, p. 15 744, 2017 (cit. on p. 11).
- [91]T. Vöpel, C. S. Hengstenberg, T. O. Peulen, *et al.*, “Triphosphate induced dimerization of human guanylate binding protein 1 involves association of the C-terminal helices: A joint double electron-electron resonance and FRET Study”, *Biochemistry*, vol. 53, no. 28, pp. 4590–4600, 2014 (cit. on p. 11).
- [92]J. A. McCammon, B. R. Gelin, and M. Karplus, “Dynamics of folded proteins”, *Nature*, vol. 267, no. 5612, pp. 585–590, 1977 (cit. on p. 13).
- [93]A. Hospital, J. R. Goñi, M. Orozco, and J. L. Gelpí, “Molecular dynamics simulations: advances and applications”, *Advances and Applications in Bioinformatics and Chemistry : AABC*, vol. 8, p. 37, 2015 (cit. on p. 13).
- [94]S. J. Marrink, V. Corradi, P. C. Souza, *et al.*, “Computational Modeling of Realistic Cell Membranes”, *Chemical Reviews*, vol. 119, no. 9, pp. 6184–6226, May 2019 (cit. on p. 13).
- [95]K. Murata and M. Wolf, “Cryo-electron microscopy for structural analysis of dynamic biological macromolecules”, *Biochimica et Biophysica Acta (BBA) - General Subjects*, vol. 1862, no. 2, pp. 324–334, Feb. 2018 (cit. on p. 14).

- [96]G. M. Clore and A. M. Gronenborn, “NMR structure determination of proteins and protein complexes larger than 20 kDa”, *Current Opinion in Structural Biology*, no. 2, pp. 564–570, 1998 (cit. on p. 14).
- [97]R. O. Dror, M. Jensen, D. W. Borhani, and D. E. Shaw, “Perspectives on: Molecular dynamics and computational methods: Exploring atomic resolution physiology on a femtosecond to millisecond timescale using molecular dynamics simulations”, *The Journal of General Physiology*, vol. 135, no. 6, p. 555, Jun. 2010 (cit. on p. 14).
- [98]H. Berendsen, D. van der Spoel, and R. van Drunen, “GROMACS: A message-passing parallel molecular dynamics implementation”, *Computer Physics Communications*, vol. 91, no. 1-3, pp. 43–56, Sep. 1995 (cit. on p. 14).
- [99]D. van der Spoel, E. Lindahl, and B. Hess, “GROMACS 3.0: a package for molecular simulation and trajectory analysis”, *Journal of Molecular Modeling*, vol. 7, pp. 306–317, 2001 (cit. on p. 14).
- [100]D. Van Der Spoel, E. Lindahl, B. Hess, *et al.*, “GROMACS: Fast, flexible, and free”, *Journal of Computational Chemistry*, vol. 26, no. 16, pp. 1701–1718, Dec. 2005 (cit. on pp. 14, 23).
- [101]B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl, “GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation”, *Journal of Chemical Theory and Computation*, vol. 4, no. 3, pp. 435–447, Mar. 2008 (cit. on p. 14).
- [102]S. Pronk, S. Páll, R. Schulz, *et al.*, “GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit”, *Bioinformatics*, vol. 29, no. 7, pp. 845–854, Apr. 2013 (cit. on p. 14).
- [103]P. Szilárd, M. J. Abraham, C. Kutzner, B. Hess, and E. Lindahl, “Tackling exascale software challenges in molecular dynamics simulations with gromacs”, in *Solving Software Challenges for Exascale*, S. Markidis and E. Laure, Eds., Cham: Springer International Publishing, 2015, pp. 3–27 (cit. on p. 14).
- [104]M. J. Abraham, T. Murtola, R. Schulz, *et al.*, “GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers”, *SoftwareX*, vol. 1-2, pp. 19–25, 2015 (cit. on pp. 14, 46).
- [105]S. Jo, T. Kim, and W. Im, “Automated builder and database of protein/membrane complexes for molecular dynamics simulations”, *PLoS ONE*, vol. 2, no. 9, Sep. 2007 (cit. on p. 14).
- [106]S. Jo, T. Kim, V. G. Iyer, and W. Im, “CHARMM-GUI: A web-based graphical user interface for CHARMM”, *Journal of Computational Chemistry*, vol. 29, no. 11, pp. 1859–1865, Aug. 2008 (cit. on p. 14).
- [107]S. Jo, J. B. Lim, J. B. Klauda, and W. Im, “CHARMM-GUI membrane builder for mixed bilayers and its application to yeast membranes”, *Biophysical Journal*, vol. 97, no. 1, pp. 50–58, 2009 (cit. on p. 14).

- [108]B. R. Brooks, C. L. Brooks, A. D. Mackerell, *et al.*, “CHARMM: The biomolecular simulation program”, *Journal of Computational Chemistry*, vol. 30, no. 10, pp. 1545–1614, Jul. 2009 (cit. on p. 14).
- [109]S. Jo, X. Cheng, S. M. Islam, *et al.*, “CHARMM-GUI PDB manipulator for advanced modeling and simulations of proteins containing nonstandard residues”, *Advances in Protein Chemistry and Structural Biology*, vol. 96, pp. 235–265, 2014 (cit. on p. 14).
- [110]S. Jo, K. C. Song, H. Desaire, A. D. MacKerell, and W. Im, “Glycan reader: Automated sugar identification and simulation preparation for carbohydrates and glycoproteins”, *Journal of Computational Chemistry*, vol. 32, no. 14, pp. 3135–3141, Nov. 2011 (cit. on p. 14).
- [111]E. L. Wu, X. Cheng, S. Jo, *et al.*, “CHARMM-GUI membrane builder toward realistic biological membrane simulations”, *Journal of Computational Chemistry*, vol. 35, no. 27, pp. 1997–2004, Oct. 2014 (cit. on p. 14).
- [112]Y. Qi, H. I. Ingólfsson, X. Cheng, *et al.*, “CHARMM-GUI Martini Maker for Coarse-Grained Simulations with the Martini Force Field”, *Journal of Chemical Theory and Computation*, vol. 11, no. 9, pp. 4486–4494, Sep. 2015 (cit. on p. 14, 27).
- [113]J. Lee, X. Cheng, J. M. Swails, *et al.*, “CHARMM-GUI Input Generator for NAMD, GRO-MACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field”, *Journal of Chemical Theory and Computation*, vol. 12, no. 1, pp. 405–413, Jan. 2016 (cit. on p. 14).
- [114]S. J. Park, J. Lee, D. S. Patel, *et al.*, “Glycan Reader is improved to recognize most sugar types and chemical modifications in the Protein Data Bank”, *Bioinformatics*, vol. 33, no. 19, pp. 3051–3057, Oct. 2017 (cit. on p. 14).
- [115]P. C. Hsu, B. M. Bruininks, D. Jefferies, *et al.*, “Charmm-gui martini maker for modeling and simulation of complex bacterial membranes with lipopolysaccharides”, *Journal of Computational Chemistry*, vol. 38, no. 27, pp. 2354–2363, 2017 (cit. on pp. 14, 24, 63).
- [116]S. J. Park, J. Lee, Y. Qi, *et al.*, “CHARMM-GUI Glycan Modeler for modeling and simulation of carbohydrates and glycoconjugates”, *Glycobiology*, vol. 29, no. 4, pp. 320–331, Feb. 2019 (cit. on p. 14).
- [117]J. Lee, D. S. Patel, J. Ståhle, *et al.*, “CHARMM-GUI Membrane Builder for Complex Biological Membrane Simulations with Glycolipids and Lipoglycans”, *Journal of Chemical Theory and Computation*, vol. 15, no. 1, pp. 775–786, Jan. 2019 (cit. on p. 14).
- [118]W. Schumann, “A molecular dynamics study of the early steps of functional csga amyloid formation”, Master’s thesis, Heinrich-Heine-Universität Düsseldorf, Düsseldorf, Nov. 2019 (cit. on p. 14).
- [119]U. Essmann, L. Perera, and M. L. Berkowitz, “A Smooth PARTICLE Mesh Ewald Method”, *J. Chem. Phys.*, vol. 103, no. 19, pp. 8577–8593, 1995 (cit. on pp. 16, 21).
- [120]S. J. Weiner, P. A. Kollman, D. T. Nguyen, and D. A. Case, “An all atom force field for simulations of proteins and nucleic acids”, *Journal of Computational Chemistry*, vol. 7, no. 2, pp. 230–252, Apr. 1986 (cit. on p. 17).

- [121]J. A. Lemkul, “Additive and Polarizable Force Fields for Proteins”, *Journal of Chemical Information and Modeling*, vol. 53, no. 9, pp. 1689–1699, 2019. eprint: arXiv:1011.1669v3 (cit. on p. 17).
- [122]J. W. Ponder and D. A. Case, “Force fields for protein simulations”, *Advances in Protein Chemistry*, vol. 66, pp. 27–85, 2003 (cit. on p. 17).
- [123]R. B. Best, D. de Sancho, and J. Mittal, “Residue-Specific  $\alpha$ -Helix Propensities from Molecular Simulation”, *Biophysical Journal*, vol. 102, no. 6, pp. 1462–1467, Mar. 2012 (cit. on p. 17).
- [124]D. A. Case, T. E. Cheatham, T. Darden, *et al.*, “The Amber biomolecular simulation programs”, *Journal of Computational Chemistry*, vol. 26, no. 16, pp. 1668–1688, Dec. 2005 (cit. on p. 17).
- [125]K. N. Kirschner, A. B. Yongye, S. M. Tschampel, *et al.*, “GLYCAM06: A generalizable biomolecular force field. Carbohydrates”, *Journal of Computational Chemistry*, vol. 29, no. 4, pp. 622–655, Mar. 2008 (cit. on p. 17).
- [126]C. J. Dickson, B. D. Madej, Å. A. Skjevik, *et al.*, “Lipid14: The amber lipid force field”, *Journal of Chemical Theory and Computation*, vol. 10, no. 2, pp. 865–879, Feb. 2014 (cit. on p. 17).
- [127]C. Tian, K. Kasavajhala, K. A. Belfon, *et al.*, “Ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution”, *Journal of Chemical Theory and Computation*, vol. 16, no. 1, pp. 528–552, Jan. 2020 (cit. on p. 18).
- [128]M. Abraham, D. van der Spoel, E. Lindahl, B. Hess, and the GROMACS development Team, *GROMACS User Manual 2018.6*, R Foundation for Statistical Computing, Groningen, Netherlands, 2019 (cit. on p. 19).
- [129]S. Nosé, “A molecular dynamics method for simulations in the canonical ensemble”, *Molecular Physics*, vol. 52, no. 2, pp. 255–268, 1984 (cit. on p. 20).
- [130]W. G. Hoover, “Canonical dynamics: Equilibrium phase-space distributions”, *Physical Review A*, vol. 31, no. 3, p. 1695, 1985 (cit. on p. 20).
- [131]G. J. Martyna, M. L. Klein, and M. Tuckerman, “NoséHoover chains: The canonical ensemble via continuous dynamics”, *The Journal of Chemical Physics*, vol. 97, no. 4, p. 2635, Jun. 1998 (cit. on p. 20).
- [132]H. J. C. Berendsen, “Transport Properties Computed by Linear Response through Weak Coupling to a Bath”, in *Computer Simulation in Materials Science*, Springer, Dordrecht, 1991, pp. 139–155 (cit. on p. 20).
- [133]M. Parrinello and A. Rahman, “Polymorphic transitions in single crystals: A new molecular dynamics method”, *Journal of Applied Physics*, vol. 52, no. 12, p. 7182, 1981 (cit. on p. 21).
- [134]S. Nosé and M. L. Klein, “Constant pressure molecular dynamics for molecular systems”, *Molecular Physics*, vol. 50, no. 5, pp. 1055–1076, 1983 (cit. on p. 21).

- [135]D. Y. T. Darden and L. Pedersen, “PARTICLE Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems”, *J. Chem. Phys.*, vol. 98, no. 12, pp. 10 089–10 092, 1993 (cit. on p. 21).
- [136]W. F. van Gunsteren and H. J. C. Berendsen, “Moleküldynamik-Computersimulationen; Methodik, Anwendungen und Perspektiven in der Chemie”, *Angewandte Chemie*, vol. 102, no. 9, pp. 1020–1055, Sep. 1990 (cit. on pp. 21, 22).
- [137]T. Schlick, *Molecular Modeling and Simulation: An Interdisciplinary Guide*, 4. 2010, vol. 82, p. 757. eprint: 1406.6401 (cit. on p. 22).
- [138]W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, “Comparison of simple potential functions for simulating liquid water”, *The Journal of Chemical Physics*, vol. 79, no. 2, pp. 926–935, Jul. 1983 (cit. on pp. 23, 46).
- [139]W. L. Jorgensen, “Transferable Intermolecular Potential Functions for Water, Alcohols, and Ethers. Application to Liquid Water”, *Journal of the American Chemical Society*, vol. 103, no. 2, pp. 335–340, 1981 (cit. on p. 23).
- [140]A. V. Onufriev and S. Izadi, “Water models for biomolecular simulations”, *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 8, no. 2, e1347, Mar. 2018 (cit. on p. 23).
- [141]S. Miyamoto and P. A. Kollman, “Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models”, *Journal of Computational Chemistry*, vol. 13, no. 8, pp. 952–962, Oct. 1992 (cit. on p. 23).
- [142]B. Hess, H. Bekker, H. J. C. Berendsen, and J. Fraaije, “LINCS: A linear constraint solver for molecular simulations”, *J. Comput. Chem.*, vol. 18, no. 12, pp. 1463–1472, 1997 (cit. on p. 23).
- [143]J. P. Ryckaert, G. Ciccotti, and H. J. Berendsen, “Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes”, *Journal of Computational Physics*, vol. 23, no. 3, pp. 327–341, Mar. 1977 (cit. on p. 23).
- [144]S. J. Marrink, A. H. De Vries, and A. E. Mark, “Coarse Grained Model for Semiquantitative Lipid Simulations”, *Journal of Physical Chemistry B*, vol. 108, no. 2, pp. 750–760, Jan. 2004 (cit. on p. 24).
- [145]S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, and A. H. De Vries, “The MARTINI force field: coarse grained model for biomolecular simulations”, *The journal of physical chemistry. B*, vol. 111, no. 27, pp. 7812–7824, Jul. 2007 (cit. on p. 24).
- [146]J. J. Uusitalo, H. I. Ingólfsson, P. Akhshi, D. P. Tieleman, and S. J. Marrink, “Martini Coarse-Grained Force Field: Extension to DNA”, *Journal of Chemical Theory and Computation*, vol. 11, no. 8, pp. 3932–3945, Aug. 2015 (cit. on p. 24).
- [147]P. C. Souza, R. Alessandri, J. Barnoud, *et al.*, “Martini 3: a general purpose force field for coarse-grained molecular dynamics”, *Nature Methods* 2021 18:4, vol. 18, no. 4, pp. 382–388, Mar. 2021 (cit. on p. 24).

- [148]S. Kmiecik, D. Gront, M. Kolinski, *et al.*, “Coarse-Grained Protein Models and Their Applications”, *Chemical Reviews*, vol. 116, no. 14, pp. 7898–7936, Jul. 2016 (cit. on p. 24).
- [149]X. Periole, M. Cavalli, S. J. Marrink, and M. A. Ceruso, “Combining an elastic network with a coarse-grained molecular force field: Structure, dynamics, and intermolecular recognition”, *Journal of Chemical Theory and Computation*, vol. 5, no. 9, pp. 2531–2543, Sep. 2009 (cit. on p. 25).
- [150]I. Siuda and L. Thøgersen, “Conformational flexibility of the leucine binding protein examined by protein domain coarse-grained molecular dynamics”, *Journal of Molecular Modeling*, vol. 19, no. 11, pp. 4931–4945, Nov. 2013 (cit. on pp. 26, 39, 49).
- [151]T. A. Wassenaar, K. Pluhackova, R. A. Böckmann, S. J. Marrink, and D. P. Tieleman, “Going Backward: A Flexible Geometric Approach to Reverse Transformation from Coarse Grained to Atomistic Models”, vol. 10, no. 2, pp. 676–690, Feb. 2014 (cit. on p. 27).
- [152]Q. Liao, “Enhanced sampling and free energy calculations for protein simulations”, *Progress in molecular biology and translational science*, vol. 170, pp. 177–213, Jan. 2020 (cit. on pp. 28–30).
- [153]Y. Sugita and Y. Okamoto, “Replica-exchange multicanonical algorithm and multi-canonical replica-exchange method for simulating systems with rough energy landscape”, *Chemical Physics Letters*, vol. 329, no. 3-4, pp. 261–270, 2000 (cit. on p. 28).
- [154]T. Okabe, M. Kawata, Y. Okamoto, and M. Mikami, “Replica-exchange Monte Carlo method for the isobaric-isothermal ensemble”, *Chemical Physics Letters*, vol. 335, no. 5-6, pp. 435–439, Mar. 2001 (cit. on p. 28).
- [155]G. Bussi, “Hamiltonian replica exchange in GROMACS: A flexible implementation”, *Mol. Phys.*, vol. 112, no. 3-4, pp. 379–384, 2014 (cit. on p. 28).
- [156]J. Kästner, “Umbrella sampling”, *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 1, no. 6, pp. 932–942, Nov. 2011 (cit. on p. 29).
- [157]S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman, “The weighted histogram analysis method for freeenergy calculations on biomolecules”, *Journal of Computational Chemistry*, vol. 13, no. 8, pp. 1011–1021, Oct. 1992 (cit. on p. 29).
- [158]J. S. Hub, B. L. de Groot, and D. van der Spoel, “g wham - A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates”, *Journal of Chemical Theory and Computation*, vol. 6, no. 12, pp. 3713–3720, 2010 (cit. on p. 31).
- [159]M. Bonomi, D. Branduardi, G. Bussi, *et al.*, “PLUMED: A portable plugin for free-energy calculations with molecular dynamics”, *Computer Physics Communications*, vol. 180, no. 10, pp. 1961–1972, Oct. 2009. eprint: 0902.0874 (cit. on pp. 31, 49).
- [160]G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, and G. Bussi, “PLUMED 2: New feathers for an old bird”, *Comp. Phys. Commun.*, vol. 185, no. 2, pp. 604–613, 2014 (cit. on pp. 31, 49).

- [161]G. Bussi and G. A. Tribello, “Analyzing and biasing simulations with PLUMED”, *Methods in Molecular Biology*, vol. 2022, pp. 529–578, Dec. 2018. eprint: 1812.08213 (cit. on pp. 31, 49).
- [162]T. Vreven, I. H. Moal, A. Vangone, *et al.*, “Updates to the Integrated ProteinProtein Interaction Benchmarks: Docking Benchmark Version 5 and Affinity Benchmark Version 2”, *Journal of Molecular Biology*, vol. 427, no. 19, pp. 3031–3041, Sep. 2015 (cit. on p. 31).
- [163]J. Jumper, R. Evans, A. Pritzel, *et al.*, “Highly accurate protein structure prediction with AlphaFold”, *Nature* 2021 596:7873, vol. 596, no. 7873, pp. 583–589, Jul. 2021 (cit. on p. 31).
- [164]G. M. Morris, D. S. Goodsell, R. S. Halliday, *et al.*, “Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function”, *Journal of Computational Chemistry*, vol. 19, no. 14, pp. 1639–1662, Nov. 1998 (cit. on p. 31).
- [165]O. Trott and A. J. Olson, “AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading”, *Journal of Computational Chemistry*, vol. 31, no. 2, 2009 (cit. on p. 31).
- [166]J. Eberhardt, D. Santos-Martins, A. F. Tillack, and S. Forli, “AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings”, *Journal of Chemical Information and Modeling*, vol. 61, no. 8, pp. 3891–3898, Jun. 2021 (cit. on p. 31).
- [167]D. Schneidman-Duhovny, Y. Inbar, R. Nussinov, and H. J. Wolfson, “PatchDock and SymmDock: Servers for rigid and symmetric docking”, *Nucleic Acids Research*, vol. 33, no. SUPPL. 2, Jul. 2005 (cit. on pp. 33, 55).
- [168]R. Chen and Z. Weng, “Docking unbound proteins using shape complementarity, desolvation, and electrostatics”, *Proteins*, vol. 47, no. 3, pp. 281–294, May 2002 (cit. on p. 33).
- [169]B. G. Pierce, K. Wiehe, H. Hwang, *et al.*, “ZDOCK server: Interactive docking prediction of protein-protein complexes and symmetric multimers”, *Bioinformatics*, vol. 30, no. 12, pp. 1771–1773, Jun. 2014 (cit. on pp. 33, 56).
- [170]X. Li, I. H. Moal, and P. A. Bates, “Detection and refinement of encounter complexes for protein-protein docking: Taking account of macromolecular crowding”, *Proteins: Structure, Function and Bioinformatics*, vol. 78, no. 15, pp. 3189–3196, Nov. 2010 (cit. on p. 33).
- [171]I. H. Moal and P. A. Bates, “SwarmDock and the Use of Normal Modes in Protein-Protein Docking”, *International Journal of Molecular Sciences*, vol. 11, no. 10, p. 3623, Oct. 2010 (cit. on pp. 33, 34).
- [172]M. Torchala, I. H. Moal, R. A. Chaleil, J. Fernandez-Recio, and P. A. Bates, “SwarmDock: A server for flexible protein-protein docking”, *Bioinformatics*, vol. 29, no. 6, pp. 807–809, Mar. 2013 (cit. on p. 33).
- [173]E. Mashiach, D. Schneidman-Duhovny, N. Andrusier, R. Nussinov, and H. J. Wolfson, “FireDock: a web server for fast interaction refinement in molecular docking.”, *Nucleic acids research*, vol. 36, no. Web Server issue, W229, 2008 (cit. on p. 33).

- [174]D. Kozakov, D. R. Hall, B. Xia, *et al.*, “The ClusPro web server for protein-protein docking”, *Nature Protocols*, vol. 12, no. 2, pp. 255–278, Feb. 2017 (cit. on pp. 33, 56).
- [175]C. Dominguez, R. Boelens, and A. M. Bonvin, “HADDOCK: A protein-protein docking approach based on biochemical or biophysical information”, *Journal of the American Chemical Society*, vol. 125, no. 7, pp. 1731–1737, Feb. 2003 (cit. on pp. 33, 56).
- [176]S. J. De Vries, A. D. Van Dijk, M. Krzeminski, *et al.*, “HADDOCK versus HADDOCK: New features and performance of HADDOCK2.0 on the CAPRI targets”, in *Proteins: Structure, Function and Genetics*, vol. 69, John Wiley & Sons, Ltd, Dec. 2007, pp. 726–733 (cit. on p. 33).
- [177]G. C. Van Zundert, J. P. Rodrigues, M. Trellet, *et al.*, “The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes”, *Journal of Molecular Biology*, vol. 428, no. 4, pp. 720–725, Feb. 2016 (cit. on p. 33).
- [178]L. Yang, G. Song, and R. L. Jernigan, “How well can we understand large-scale protein motions using normal modes of elastic network models?”, *Biophysical Journal*, vol. 93, no. 3, pp. 920–929, Aug. 2007 (cit. on p. 34).
- [179]F. Tama and Y. H. Sanejouand, “Conformational change of proteins arising from normal mode calculations”, *Protein Engineering*, vol. 14, no. 1, pp. 1–6, 2001 (cit. on p. 34).
- [180]D. K. Brown, D. L. Penkler, O. S. Amamuddy, *et al.*, “MD-TASK: a software suite for analyzing molecular dynamics trajectories”, *Bioinformatics*, vol. 33, no. 17, p. 2768, Sep. 2017 (cit. on p. 35).
- [181]H. H. Low and J. Löwe, “A bacterial dynamin-like protein”, *Nature*, vol. 444, no. 7120, pp. 766–769, 2006 (cit. on pp. 37, 39).
- [182]H. H. Low, C. Sachse, L. A. Amos, and J. Löwe, “Structure of a Bacterial Dynamin-like Protein Lipid Tube Provides a Mechanism For Assembly and Membrane Curving”, *Cell*, vol. 139, no. 7, pp. 1342–1352, 2009 (cit. on p. 37).
- [183]M. L. Rennie, S. A. McKelvie, E. M. Bulloch, and R. L. Kingston, “Transient dimerization of human MxA promotes GTP hydrolysis, resulting in a mechanical power stroke”, *Structure*, vol. 22, no. 10, pp. 1433–1445, Oct. 2014 (cit. on pp. 42, 43).
- [184]S. Shydlovskiy, A. Y. Zienert, S. Ince, *et al.*, “Nucleotide-dependent farnesyl switch orchestrates polymerization and membrane binding of human guanylate-binding protein”, *Proceedings of the National Academy of Sciences of the United States of America*, vol. 114, no. 28, E5559–E5568, Jul. 2017 (cit. on p. 43).
- [185]K. Faelber, M. Held, S. Gao, *et al.*, “Structural insights into dynamin-mediated membrane fission”, vol. 20, no. 10, pp. 1621–1628, Oct. 2012 (cit. on p. 43).
- [186]N. Britzen-Laurent, M. Bauer, V. Berton, *et al.*, “Intracellular Trafficking of Guanylate-Binding Proteins Is Regulated by Heterodimerization in a Hierarchical Manner”, *PLoS ONE*, vol. 5, no. 12, A. M. Delprato, Ed., e14246, Dec. 2010 (cit. on p. 43).
- [187]S. Ince, M. Kutsch, S. Shydlovskiy, and C. Herrmann, “The human guanylate-binding proteins hGBP-1 and hGBP-5 cycle between monomers and dimers only”, *The FEBS Journal*, vol. 284, no. 14, pp. 2284–2301, Jul. 2017 (cit. on p. 43).

- [188]A. Ghosh, G. J. K. Praefcke, L. Renault, A. Wittinghofer, and C. Herrmann, “How guanylate-binding proteins achieve assembly-stimulated processive cleavage of GTP to GMP”, *Nature*, vol. 440, no. 7080, pp. 101–104, 2006 (cit. on pp. 43, 47).
- [189]W. Cui, E. Braun, W. Wang, *et al.*, “Structural basis for GTP-induced dimerization and antiviral function of guanylate-binding proteins”, *Proceedings of the National Academy of Sciences*, vol. 118, no. 15, e2022269118, Apr. 2021 (cit. on pp. 43, 51).
- [190]L. Legewie, J. Loschwitz, N. Steffens, *et al.*, “Biochemical and structural characterization of murine GBP7, a guanylate binding protein with an elongated C-terminal tail”, *Biochemical Journal*, Oct. 2019 (cit. on pp. 44, 49, 57).
- [191]D. R. Fattal and A. Ben-Shaul, “A molecular model for lipid-protein interaction in membranes: the role of hydrophobic mismatch”, *Biophysical Journal*, vol. 65, no. 5, pp. 1795–1809, Nov. 1993 (cit. on p. 44).
- [192]V. Lindenberg, K. Mölleken, E. Kravets, *et al.*, “Broad recruitment of mGBP family members to Chlamydia trachomatis inclusions”, *PLoS ONE*, vol. 12, no. 9, Sep. 2017 (cit. on p. 45).
- [193]A. E. Aliev, M. Kulke, H. S. Khaneja, *et al.*, “Motional timescale predictions by molecular dynamics simulations: Case study using proline and hydroxyproline sidechain dynamics”, *Proteins*, vol. 82, no. 2, p. 195, 2014 (cit. on p. 46).
- [194]R. B. Best and G. Hummer, “Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides”, *The journal of physical chemistry. B*, vol. 113, no. 26, p. 9004, 2009 (cit. on p. 46).
- [195]K. Lindorff-Larsen, S. Piana, K. Palmo, *et al.*, “Improved side-chain torsion potentials for the Amber ff99SB protein force field”, *Proteins*, vol. 78, no. 8, pp. 1950–1958, 2010 (cit. on p. 46).
- [196]X. Daura, K. Gademann, B. Jaun, *et al.*, “Peptide Folding: When Simulation Meets Experiment”, *Angew. Chem. Int. Ed.*, vol. 38, no. 1-2, pp. 236–240, 1999 (cit. on p. 46).
- [197]W. Schumann, J. Loschwitz, D. Degrandi, *et al.*, “Dynamics and dimerization of different mGBPs in comparison via computational and experimental methods”, *bioRxiv*, 2022 (cit. on pp. 46, 50, 53).
- [198]M. Wehner, S. Kunzelmann, and C. Herrmann, “The guanine cap of human guanylate-binding protein 1 is responsible for dimerization and self-activation of GTP hydrolysis”, *FEBS Journal*, vol. 279, no. 2, pp. 203–210, Jan. 2012 (cit. on p. 47).
- [199]S. Ince, P. Zhang, M. Kutsch, *et al.*, “Catalytic activity of human GuanylateBinding Protein 1 coupled to the release of structural restraints imposed by the Cterminal domain”, *The FEBS Journal*, febs.15348, Apr. 2020 (cit. on p. 48).
- [200]J. Lichte, “Funktions- und Strukturanalysen zur Charakterisierung des murinen GBP9 Proteins bei Infektionen mit intrazellulären Erregern”, 2022 (cit. on p. 48).
- [201]R. Gasper, S. Meyer, K. Gotthardt, M. Sirajuddin, and A. Wittinghofer, “It takes two to tango: regulation of G proteins by dimerization”, *Nature reviews. Molecular cell biology*, vol. 10, no. 6, pp. 423–429, Jun. 2009 (cit. on p. 49).

- [202]R. Narayanan, M. Leonard, B. D. Song, S. L. Schmid, and M. Ramaswami, “An internal GAP domain negatively regulates presynaptic dynamin in vivo: a two-step model for dynamin function”, *The Journal of Cell Biology*, vol. 169, no. 1, p. 117, 2005 (cit. on p. 49).
- [203]S. M. Ferguson and P. De Camilli, “Dynamin, a membrane-remodelling GTPase”, *Nature Reviews Molecular Cell Biology*, vol. 13, no. 2, pp. 75–88, Feb. 2012 (cit. on p. 49).
- [204]L. Sistemich, M. Kutsch, B. Hämisch, *et al.*, “The Molecular Mechanism of Polymer Formation of Farnesylated Human Guanylate-binding Protein 1”, *Journal of Molecular Biology*, vol. 432, no. 7, pp. 2164–2185, Mar. 2020 (cit. on pp. 50, 53, 65).
- [205]L. Sistemich, L. D. Stanchev, M. Kutsch, *et al.*, “Structural requirements for membrane binding of human guanylatebinding protein 1”, *The FEBS Journal*, febs.15703, Jan. 2021 (cit. on p. 50).
- [206]S. Zhu, C. J. Bradfield, A. Maminska, *et al.*, “Cryo-ET of a human GBP coatomer governing cell-autonomous innate immunity to infection”, *bioRxiv*, pp. 1–65, Aug. 2021 (cit. on pp. 51, 53).
- [207]M. Kutsch, L. Sistemich, C. F. Lesser, *et al.*, “Direct binding of polymeric GBP1 to LPS disrupts bacterial cell envelope functions”, *The EMBO Journal*, vol. 39, no. 13, e104926, Jul. 2020 (cit. on pp. 51, 53).
- [208]B. Striepen, S. Tomavo, J. F. Dubremetz, and R. T. Schwarz, “Identification and characterisation of glycosyl-inositolphospholipids in *Toxoplasma gondii*”, *Biochemical Society Transactions*, vol. 20, no. 3, 1992 (cit. on p. 51).
- [209]B. Striepen, C. F. Zinecker, J. B. Damm, *et al.*, “Molecular structure of the low molecular weight antigen of *Toxoplasma gondii*: a glucose  $\alpha$ 1-4 N-acetylgalactosamine makes free glycosyl-phosphatidylinositols highly immunogenic”, *Journal of Molecular Biology*, vol. 266, no. 4, pp. 797–813, Mar. 1997 (cit. on pp. 51, 57).
- [210]N. Azzouz, H. Shams-Eldin, S. Niehus, *et al.*, “*Toxoplasma gondii* grown in human cells uses GalNAc-containing glycosylphosphatidylinositol precursors to anchor surface antigens while the immunogenic GlcGalNAc-containing precursors remain free at the parasite cell surface”, *The International Journal of Biochemistry & Cell Biology*, vol. 38, no. 11, pp. 1914–1925, Jan. 2006 (cit. on pp. 51, 57).
- [211]S. Götze, N. Azzouz, Y.-H. Tsai, *et al.*, “Diagnosis of Toxoplasmosis Using a Synthetic Glycosylphosphatidylinositol Glycan”, *Angewandte Chemie International Edition*, vol. 53, no. 50, pp. 13 701–13 705, Dec. 2014 (cit. on pp. 51, 57, 58).
- [212]T.-O. O. Peulen, C. S. Hengstenberg, R. Biehl, *et al.*, “Integrative dynamic structural biology unveils conformers essential for the oligomerization of a large GTPase Integrative dynamic structural biology unveils conformers essential for the oligomerization of a large GTPase”, *arxiv*, no. April, Apr. 2020. arXiv: 2004.04229 (cit. on p. 53).
- [213]E. Kravets, D. Degrandi, S. Weidtkamp-Peters, *et al.*, “The GTPase Activity of Murine Guanylate-binding Protein 2 (mGBP2) Controls the Intracellular Localization and Recruitment to the Parasitophorous Vacuole of *Toxoplasma gondii*”, *Journal of Biological Chemistry*, vol. 287, no. 33, pp. 27 452–27 466, Aug. 2012 (cit. on p. 53).

- [214]J. C. Santos and P. Broz, “Sensing of invading pathogens by GBPs: At the crossroads between cell-autonomous and innate immunity”, *Journal of Leukocyte Biology*, vol. 104, no. 4, pp. 729–735, Oct. 2018 (cit. on p. 53).
- [215]I. M. Rosenberg, “Protein analysis and purification: Benchtop techniques: Second edition”, *Protein Analysis and Purification: Benchtop Techniques: Second Edition*, pp. 1–520, 2005 (cit. on p. 55).
- [216]K. A. Porter, I. Desta, D. Kozakov, and S. Vajda, “What method to use for proteinprotein docking?”, *Current Opinion in Structural Biology*, vol. 55, pp. 1–7, 2019 (cit. on p. 55).
- [217]“Covalent protein modification with ISG15 via a conserved cysteine in the hinge region”, *PLoS ONE*, vol. 7, no. 6, Jun. 2012 (cit. on pp. 56, 63).
- [218]E. Knight, D. Fahey, B. Cordova, *et al.*, “A 15-kDa interferon-induced protein is derived by COOH-terminal processing of a 17-kDa precursor.”, *The Journal of biological chemistry*, vol. 263, no. 10, pp. 4520–2, Apr. 1988 (cit. on p. 56).
- [219]D. Zhang and D.-E. Zhang, “Interferon-stimulated gene 15 and the protein ISGylation system.”, *Journal of interferon & cytokine research*, vol. 31, no. 1, pp. 119–30, Jan. 2011 (cit. on p. 56).
- [220]M. R. Rodriguez, K. Monte, L. B. Thackray, and D. J. Lenschow, “ISG15 Functions as an Interferon-Mediated Antiviral Effector Early in the Murine Norovirus Life Cycle”, *Journal of Virology*, vol. 88, no. 16, pp. 9277–9286, Aug. 2014 (cit. on p. 56).
- [221]A. Napolitano, A. G. van der Veen, M. Bunyan, *et al.*, “Cysteine-Reactive Free ISG15 Generates IL-1 $\beta$ -Producing CD8 $\alpha$ + Dendritic Cells at the Site of Infection”, *Journal of immunology (Baltimore, Md. : 1950)*, vol. 201, no. 2, pp. 604–614, Jul. 2018 (cit. on pp. 56, 63).
- [222]C. Zhao, T. Y. Hsiang, R. L. Kuo, and R. M. Krug, “ISG15 conjugation system targets the viral NS1 protein in influenza A virusinfected cells”, *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, no. 5, p. 2253, Feb. 2010 (cit. on pp. 56, 63).
- [223]C. M. Daczkowski, J. V. Dzimianski, J. R. Clasman, *et al.*, “Structural Insights into the Interaction of Coronavirus Papain-Like Proteases and Interferon-Stimulated Gene Product 15 from Different Species”, *Journal of Molecular Biology*, vol. 429, no. 11, pp. 1661–1683, Jun. 2017 (cit. on p. 56).
- [224]D. Komander and M. Rape, “The ubiquitin code”, *Annual review of biochemistry*, vol. 81, pp. 203–229, Jul. 2012 (cit. on p. 57).
- [225]Y. Wang and H. Yin, “Research progress on surface antigen 1 (SAG1) of *Toxoplasma gondii*”, vol. 7, no. 1, Apr. 2014 (cit. on p. 57).
- [226]J. J. Almagro Armenteros, K. D. Tsirigos, C. K. Sønderby, *et al.*, “SignalP 5.0 improves signal peptide predictions using deep neural networks”, *Nature Biotechnology* 2019 37:4, vol. 37, no. 4, pp. 420–423, Feb. 2019 (cit. on p. 57).
- [227]E. Gasteiger, C. Hoogland, A. Gattiker, *et al.*, “Protein Identification and Analysis Tools on the ExpASY Server”, in *The Proteomics Protocols Handbook*, Humana Press, 2005, pp. 571–607 (cit. on p. 57).

- [228]L. Käll, A. Krogh, and E. L. Sonnhammer, “Advantages of combined transmembrane topology and signal peptide prediction-the Phobius web server”, *Nucleic Acids Research*, vol. 35, no. SUPPL.2, Jul. 2007 (cit. on p. 57).
- [229]B. Rost, G. Yachdav, and J. Liu, “The PredictProtein server”, *Nucleic Acids Research*, vol. 32, no. WEB SERVER ISS. Jul. 2004 (cit. on p. 57).
- [230]L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass, and M. J. Sternberg, “The Phyre2 web portal for protein modeling, prediction and analysis”, *Nature Protocols*, vol. 10, no. 6, pp. 845–858, Jun. 2015 (cit. on pp. 57, 59).
- [231]N. Guex, M. C. Peitsch, and T. Schwede, “Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective”, *Electrophoresis*, vol. 30, Jun. 2009 (cit. on p. 57).
- [232]J. D. Wasmuth, V. Pszenny, S. Haile, *et al.*, “Integrated Bioinformatic and Targeted Deletion Analyses of the SRS Gene Superfamily Identify SRS29C as a Negative Regulator of Toxoplasma Virulence”, *mBio*, vol. 3, no. 6, L. M. Weiss, Ed., Nov. 2012 (cit. on p. 57).
- [233]P. Banerjee, M. Wehle, R. Lipowsky, and M. Santer, “A molecular dynamics model for glycosylphosphatidyl-inositol anchors: "flop down" or "lollipop"?", *Physical Chemistry Chemical Physics*, vol. 20, no. 46, pp. 29 314–29 324, Nov. 2018 (cit. on pp. 58, 63).
- [234]J. Schultz, F. Milpetz, P. Bork, and C. P. Ponting, “SMART, a simple modular architecture research tool: Identification of signaling domains”, *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, no. 11, pp. 5857–5864, May 1998 (cit. on p. 59).
- [235]T. Wang, N. S. Liu, L.-F. Seet, and W. Hong, “The Emerging Role of VHS Domain-Containing Tom1, Tom1L1 and Tom1L2 in Membrane Trafficking”, *Traffic*, vol. 11, no. 9, pp. 1119–1128, Sep. 2010 (cit. on p. 59).
- [236]S. Misra, B. Beach, and J. Hurley, “Structure of the VHS domain of human Tom1 (target of myb 1): insights into interactions with proteins and membranes”, *Biochemistry*, vol. 39, pp. 11 282–11 290, 2000 (cit. on p. 59).
- [237]Y. Zhang, “I-TASSER server for protein 3D structure prediction”, *BMC Bioinformatics*, vol. 9, no. 1, p. 40, Dec. 2008 (cit. on p. 59).
- [238]A. Roy, A. Kucukural, and Y. Zhang, “I-TASSER: A unified platform for automated protein structure and function prediction”, *Nature Protocols*, vol. 5, no. 4, pp. 725–738, 2010 (cit. on p. 59).
- [239]P. Li, W. Jiang, Q. Yu, *et al.*, “Ubiquitination and degradation of GBPs by a Shigella effector to suppress host defence”, *Nature*, vol. 551, no. 7680, pp. 378–383, Nov. 2017 (cit. on p. 63).
- [240]P. Banerjee, R. Lipowsky, and M. Santer, “Coarse-Grained Molecular Model for the Glycosylphosphatidylinositol Anchor with and without Protein”, *Journal of Chemical Theory and Computation*, vol. 16, no. 6, pp. 3889–3903, Jun. 2020 (cit. on p. 63).

## List of Abbreviations

- AA** all-atom
- AIDS** acquired immune deficiency syndrome
- AMBER** assisted model building with energy refinement
- ATG** autophagy protein
- ATP** adenosine triphosphate
- BAR** Bin/Amphiphysin/Rvs
- BDLP** bacterial dynamin-like protein
- BFGS** Broyden-Fletcher-Goldfarb-Shanno
- BIN** bridging integrator
- BSE** bundle signalling element
- CAPRI** critical assessment of predicted interactions
- CASP** critical assessment of techniques for protein structure prediction
- CG** coarse grained
- CHARMM** chemistry at Harvard macromolecular mechanics
- CMAP** 2D dihedral energy grid correction map
- CT tail** C terminal tail
- CV** collective variable
- DOPC** dipalmitoyl-phosphatidylcholine
- Drp1** dynamin related protein 1
- DSP** dynamin superfamily protein
- E domain** effector domain
- ELNEDIN** elastic network dynamics
- EM** electron microscopy
- ENM** elastic network models

**FES** free energy surface

**Fzo1p** fuzzy onion protein 1

**G1** phosphate loop

**G2** switch 1

**G3** switch 2

**G5** guanine cap

**GAP** GTPase activating protein

**GBP** guanylate binding protein

**GC** guanine cap

**GD** GTPase domain

**GDP** guanosine diphosphate

**GEF** guanine nucleotide exchange factor

**GMP** guanosine monophosphate

**G protein** guanine nucleotide-binding protein

**GPI** glycosyl-phosphatidyl-inositol

**GPL** glycerophospholipid

**GROMACS** Groningen machine for chemical simulations

**GTP** guanosine triphosphate

**GUI** graphical user interface

**HIV** human immune deficiency virus

**HREMD** Hamiltonian replica exchange molecular dynamics simulation

**Hrs** hepatocyte growth factor-regulated tyrosine kinase substrate

**IFN** interferon

**IL** interleukin

**IRG** immunity related GTPase

**ISG15** interferon stimulated gene 15

**L1** loop 1

**L2** loop 2

**Leo** labile enterotoxin output

**LINCS** linear constraint solver

**LPS** lipopolysaccharides

**MD** molecular dynamics

**M domain** middle domain

**mdp** molecular dynamics parameters

**mGBP2** murine guanylate binding protein 2

**mGBP7** murine guanylate binding protein 7

**mGBP9** murine guanylate binding protein 9

**Mgm1p** mitochondrial genome maintenance protein 1

**MPI** message passing interface

**MxA** myxovirus resistance protein A

**NK cells** natural killer cell

**NMR** nuclear magnetic resonance

**NpT** isothermal-isobaric ensemble

**NVT** canonical ensemble

**OP** order parameter

**OPA1** optic atrophy 1

**OpenMP** Open Multi-Processing

**P-loop** phosphate loop

**PAMP** pathogen associated molecular pattern

**PBC** periodic boundary conditions

**PC** phosphatidylcholine

**PCA** principal component analysis

**PDB** Protein Data Bank

**Pi** inorganic phosphate

**PL** phospholipid

**PMF** potential of mean force

**POPC** palmitoyl-oleoyl-phosphatidylcholine

**PPi** inorganic pyrophosphate

**PRR** pattern recognition receptor

**PVM** parasitophorous vacuole membrane

**QM** quantum mechanics

**QM/MM** quantum mechanics/molecular mechanics

**REMD** replica exchange molecular dynamics simulation

**RG** radius of gyration

**RMSD** root mean square deviation

**RMSF** root mean square fluctuation

**SAG** surface antigen

**SAG1** surface antigen 1

**SAXS** small angle X-ray scattering

**SRS** SAG related sequence

**SRS29C** SAG related sequence 29C

**STAM** signal transducing adapter molecule

**SW1** switch 1

**SW2** switch 2

**TIP3P** three-site transferrable intermolecular potential

**TOM1** target of Myb1

**UIM** ubiquitin interacting motif

**USMD** umbrella sampling molecular dynamics

**VHS** Vps27/Hrs/STAM

**Vps27** vacuolar protein sorting 27

**Vpsp1** vacuole sorting protein 1

**WHAM** weighted histogram analysis method

**XL-MS** crosslinking mass spectrometry



# List of Figures

1.1. Biological lipids and membranes . . . . .	2
1.2. Membrane shaping mechanisms . . . . .	4
1.3. Domain structure of dynamin superfamily proteins . . . . .	5
1.4. <i>T. gondii</i> life cycle . . . . .	8
2.1. Scope of Membrane Simulations over Time . . . . .	13
2.2. Biophysical methods . . . . .	14
2.3. Periodic boundary conditions . . . . .	22
2.4. MARTINI coarse graining principle . . . . .	24
2.5. ELNEDIN vs. domELNEDIN . . . . .	26
2.6. Replica exchange molecular dynamics principle . . . . .	28
2.7. Umbrella Sampling . . . . .	30
3.1. PMF of BDLP USMD . . . . .	38
3.2. 1 $\mu$ s CG simulation of BDLP systems . . . . .	40
3.3. 1 $\mu$ s CG simulation of BDLP systems . . . . .	41
3.4. BDLP GTPase cycle . . . . .	42
4.1. Dimerization modes of GBPs . . . . .	45
4.2. mGBP9 comparison across monomer/dimer and apo/holo . . . . .	47
4.3. Pulling of hGBP1 . . . . .	50
4.4. GBPs on DOPC/Chol membranes . . . . .	52
5.1. GPI-anchor structure . . . . .	58
5.2. Interaction Parters . . . . .	60
5.3. Interaction Energies of mGBP7 and TOM1 . . . . .	61
5.4. HADDOCK scores . . . . .	62

## List of Tables

5.1. ISG15 HADDOCK restraint sets . . . . .	56
5.2. SRS29C HADDOCK restraint sets . . . . .	58
5.3. TOM1 HADDOCK restraint sets . . . . .	59
5.4. Interacting residues . . . . .	62
C.1. NVT Equilibration Parameter File . . . . .	218
C.2. NpT Equilibration Parameter File . . . . .	220
C.3. Production Run Parameter File . . . . .	222
C.4. HREMD Production Run Parameter File . . . . .	224
C.5. USMD Production Run Parameter File . . . . .	226
C.6. MARTINI Production Run Parameter File . . . . .	228

## List of Listings

C.1. ELNEDIN file for BDLP . . . . .	229
C.2. Hinge Motion with tcl . . . . .	229
C.3. plumed pulling . . . . .	232

## List of Algorithms

1. Pseudocode of GROMACS . . . . .	19
2. Pseudocode of Minimization . . . . .	32

# Acknowledgements

I would like express my sincere thanks to Birgit Strodel for being not just my supervisor, full of enthusiasm for the topic, but also an understanding mentor in challenging times. I also appreciate Laura Hartmann, who was officially my mentor, but also took interest in my progress. Daniel Degrandi, Klaus Pfeffer, Sander Smits and Kai Stühler organized exciting collaborative projects, and Sander further agreed to be my second reviewer.

Jan Meisner, who adopted us into his offices, enlivened the workplace with kind and witty exchange, and sociable group events. Cordula Kruse, Gudrun Brauwers and Dorothea Stobbe helped me with more than one administrative process and also supplied kind conversation.

I am grateful for the financing from Jürgen Manchot Stiftung and the scientific exchange provided by SFB1208, as well as computing time granted by the Leibniz supercomputing center in Munich. Technical support from the HPC teams of JURECA, HILBERT, SuperMUC, and personally from Klaus Eifert, Volker Söhnitz and Friedrich Haubensak was also very welcome.

With Sophia Kasbrink, Gowshiga Sivalingam, Gereon Porschmann and Jens Reiners I had fruitful collaboration. I am especially grateful to Jennifer Loschwitz for her help with beautiful figures. Larissa Legewie and Nora Steffens laid the groundwork for my work, and their theses and defenses inspired me. All members of the Strodel group managed to keep up long-distance group spirit, interspersed with outings, dinners and events. I would like to especially thank Anna Jäckering, Moritz Schäffler, Hiba Fatafta, Mohammed Khaled and Lara Scharbert for agreeing to read a draft of this thesis.

I am grateful to Marie Erdorf-De Vilder for support in the hardest of times.

My parents and brothers, Ulrike, Jürgen, Torben and Birger Schumann gave me their unconditional love and support and helped shape the person I am.

Finally, I want to express my deepest thanks to Thomas Wagner for his patience, love, unwavering support and much more.



Publication 1

A

# **Allosteric communication induced by GTP binding sets off a closed-to-open transition in a bacterial dynamin-like protein**

Wibke Schumann<sup>†,‡</sup> and Birgit Strodel<sup>\*,†,‡</sup>

*†Institute of Theoretical and Computational Chemistry, Heinrich Heine University  
Düsseldorf, 40225 Düsseldorf, Germany*

*‡Institute of Biological Information Processing: Structural Biochemistry, Forschungszentrum  
Jülich, 52428 Jülich, Germany*

E-mail: [b.strodel@fz-juelich.de](mailto:b.strodel@fz-juelich.de)

## Abstract

Dynamin superfamily proteins are mechanochemical GTPases that operate in highly oligomeric and highly cooperative superstructures to deform lipid membranes. It is known from the structures of a bacterial dynamin-like protein (BDLP) that binding of GTP and association of BDLP with lipids causes a transition from closed to open hinge 1 that affects oligomerization. We trace this radical, large-scale conformational change at the atomic level with unbiased, replica exchange and umbrella sampling molecular dynamics simulations. We decipher how GTP loading from the GTPase domain to the distal stalk end is mediated by an allosteric network of salt bridges that act in response to GTP binding and subsequent conformational changes in GTPase domain motifs. Two previously undiscovered motifs have been identified whose movements free the paddle from the GTPase domain, allowing large-scale domain rearrangements. In addition, a novel wide-open state of BDLP reminiscent of human dynamin 1 is discovered. Our results explain several aspects of the BDLP cycle and have broad implications for other members of the dynamin family.

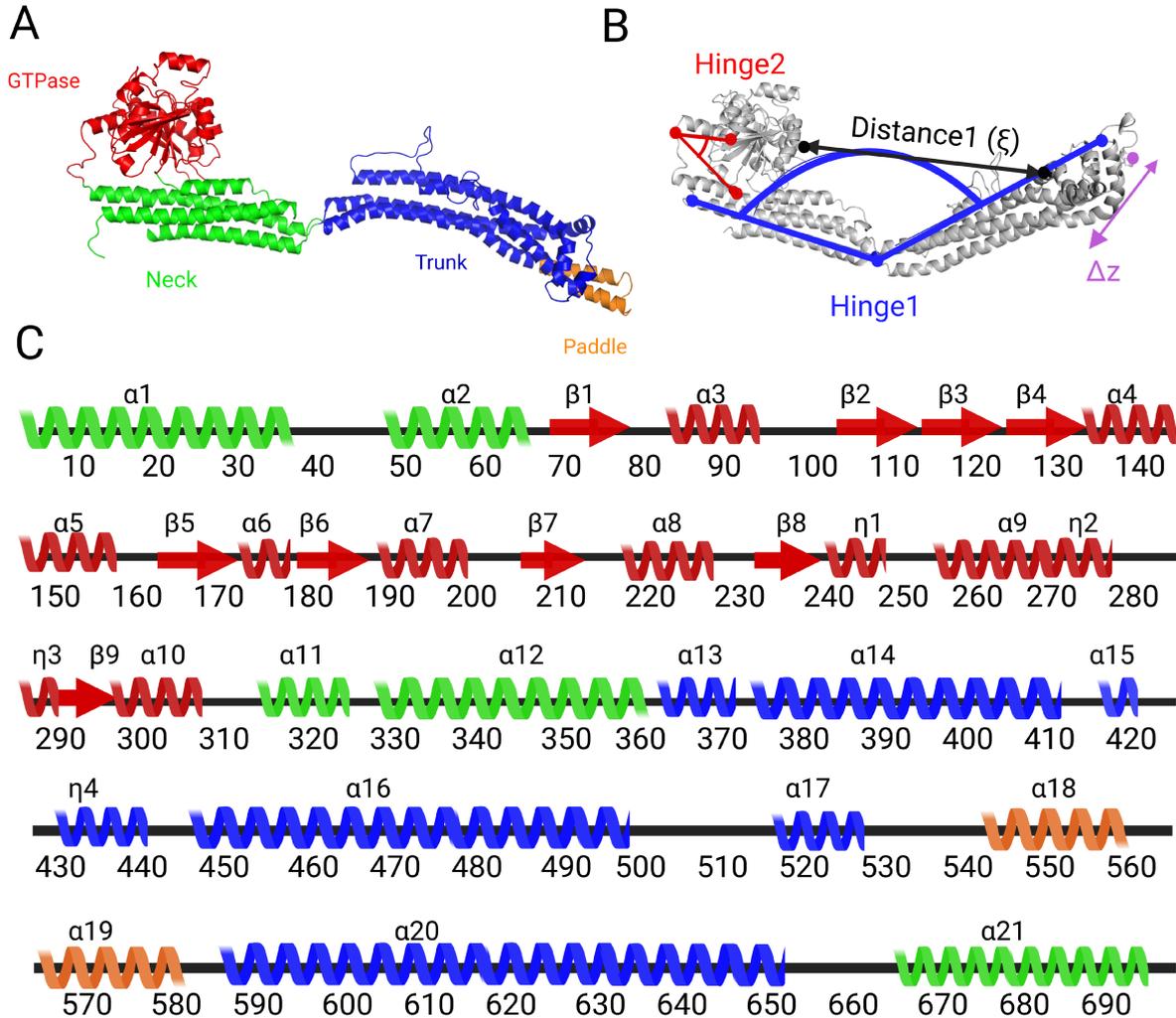
**Keywords:** dynamin-like proteins, large GTPases, molecular dynamics, umbrella sampling, large-scale motion

# 1 Introduction

Dynamin-superfamily proteins (DSPs) are mechanochemical enzymes, involved in critical cellular functions like endocytosis, cell division and immune response.<sup>1-4</sup> Next to flagellar proteins, they generate some of the highest torques known for proteins, in the range of a thousand piconewton-nanometers (or a few attojoule).<sup>1,5</sup> DSPs differ from the smaller Ras-like GTPases in their lack of accessory proteins, lower substrate affinity, and higher basal hydrolysis rate, which is highly stimulated in oligomers.<sup>6-9</sup> The oligomers form ordered lattices, rings, or helices, which tubulate membranes.<sup>2</sup> The conserved GTPase domain of DSPs spans about 300 residues and features an internal GTPase-activating domain to replace the external activation factors.<sup>9-11</sup> GTP binding leads to oligomerization and stronger membrane association, while GTP hydrolysis leads to fission and oligomer dissociation.<sup>3,12</sup> GTPase activity needs to be highly concerted, local and fast, but not simultaneous, rather propagating along the supramolecular helix to avoid weak points.<sup>13,14</sup> In polymerized DSPs, 2-4 GTPs can be hydrolyzed per second, and the open and closed states have lifetimes in the range of a few seconds.<sup>15,16</sup> The GTP dissociation rate lies in the range of 10-100/s and thus, dimer lifetime of DSPs is short (in the range of hundreds of milliseconds).<sup>17</sup> Assembly of a microscopically visible dynamin sheath on a membrane can take up to an hour,<sup>17</sup> while membrane fission can occur as soon as 10-20 subunits are assembled, with the final fission happening in seconds to minutes.<sup>17</sup> Several mechanisms for dynamin action have been proposed, among them the poppase, pinchase, and twistase mechanisms.<sup>18-20</sup> The poppase action works by extending the membrane stalk, thinning the bilayer, which then ruptures. In the pinchase theory, it is a conformational change that leads to a reduction of the DSP helix diameter and thus to membrane constriction. The twistase mechanism achieves the same effect by forming supramolecular winches, which decrease the number of proteins per turn. Permanent membrane binding ability is necessary for fusion DSPs, while several cycles of constriction are needed for fission.

Bacterial dynamin-like proteins have been associated with several membrane-related

functions, including membrane vesicle formation and membrane fusion. The current model is that bacterial dynamin-like proteins are recruited to sites where homotypic membrane fusion is required. The bacterial dynamin-like protein from *Nostoc punctiforme* (called BDLP henceforth) is supposed to be involved in fusion rather than fission. BDLP has a canonical G domain separated by hinge2 from the neck and trunk region, which are connected via hinge1 Figure 1. The membrane interaction is mediated by a paddle domain. BDLP localizes to the outer leaflet of the inner plasma membrane and could be the cyanobacterial ancestor of the thylakoid-reshaping fuzzy-onion-like protein in higher plants. Membrane binding of BDLP is connected to the dimerization of its G domain, self-assembly of the C-terminal GTPase effector domain, and the paddle region contacting the lipids promoting membrane curvature. This can only be achieved in the open or extended state of BDLP, whose structure was resolved to 9 Å with cryogenic electron microscopy as a dimer loaded with a GTP analogue (GMPPNP, guanosine-5'-[( $\beta$ , $\gamma$ )-imido]triphosphate) and with the paddles inserted into a lipid membrane in a pincushion-fashion.<sup>21</sup> This requires a large-scale conformational change from the closed state of BDLP, whose GDP-bound structure was resolved with X-ray diffraction to a resolution of 3 Å. It features an acute triangular arrangement of the G domain, neck and trunk<sup>22</sup> and is less likely to bind to membranes.<sup>23</sup> The BDLP is the only DSP for which high-quality structures exist in both the open and closed conformation.



**Figure 1: Structural characteristics of BDLP.** **A** The open form of BDLP is shown as cartoon and colored red for the G domain, green for the neck, blue for the trunk, and the membrane-binding paddle is shown in orange. **B** The open and closed states and their interconversion are characterized by several order parameters. Hinge1 is described by the angle  $\alpha$  between the  $C_{\alpha}$  atoms of residues 4, 359, and 587 (blue), with  $\alpha \approx 175^{\circ}$  representing the open BDLP conformation and  $\alpha \approx 40^{\circ}$  the closed one. Hinge2, which was previously proposed to be an important switch for the GTP loading state, is characterized by the angle  $\beta$  between the  $C_{\alpha}$  atoms of residues 291, 303 and 323 (red). The amount of opening or closing is measured by the distance  $\xi$  between the  $C_{\alpha}$  atoms of residues 224 and 453 (black arrow), and the lateral stalk trunk motions are defined by changes along the  $z$  coordinate (magenta arrow). **C** The secondary structure of BDLP along with the labeling of its helices and  $\beta$ -strands is shown.

The aim of the current work is to unravel the molecular details of the closed-to-open transition and provide a rationale for the structural prerequisites and implications of that

radical conformational change. To reach that goal, we perform all-atom molecular dynamics (MD) simulations, which has become an accepted method to fill the gaps left by experimental methods as it provides a higher spatial and temporal resolution than the experiments.<sup>24-26</sup> Thus far, no all-atom MD simulations on DSPs have been reported yet; only coarse-grained simulations with non-atomic resolution of DSPs are published to date.<sup>27-32</sup> In our lab, we simulated different guanylate-binding proteins (GBPs), which belong to the dynamin-related superfamily, at atomic resolution, which revealed a large-scale hinge movement that may correspond to hinge1 in the DSPs.<sup>33,34</sup> Support for such a hinge movement is provided by experimental observation.<sup>15,35</sup> Nonetheless, the hinge movement sampled in our simulations of GBPs did not involve a complete open-to-closed transition, as suggested by the BDLP structures. We therefore set out to close this gap, using the open and closed BDLP structures as input, to characterize the complete hinge movement for the first time. We employ standard MD simulations, Hamiltonian replica exchange and umbrella sampling MD simulations (HREMD and USMD respectively), and explore the allosteric effects triggered by GTP binding.

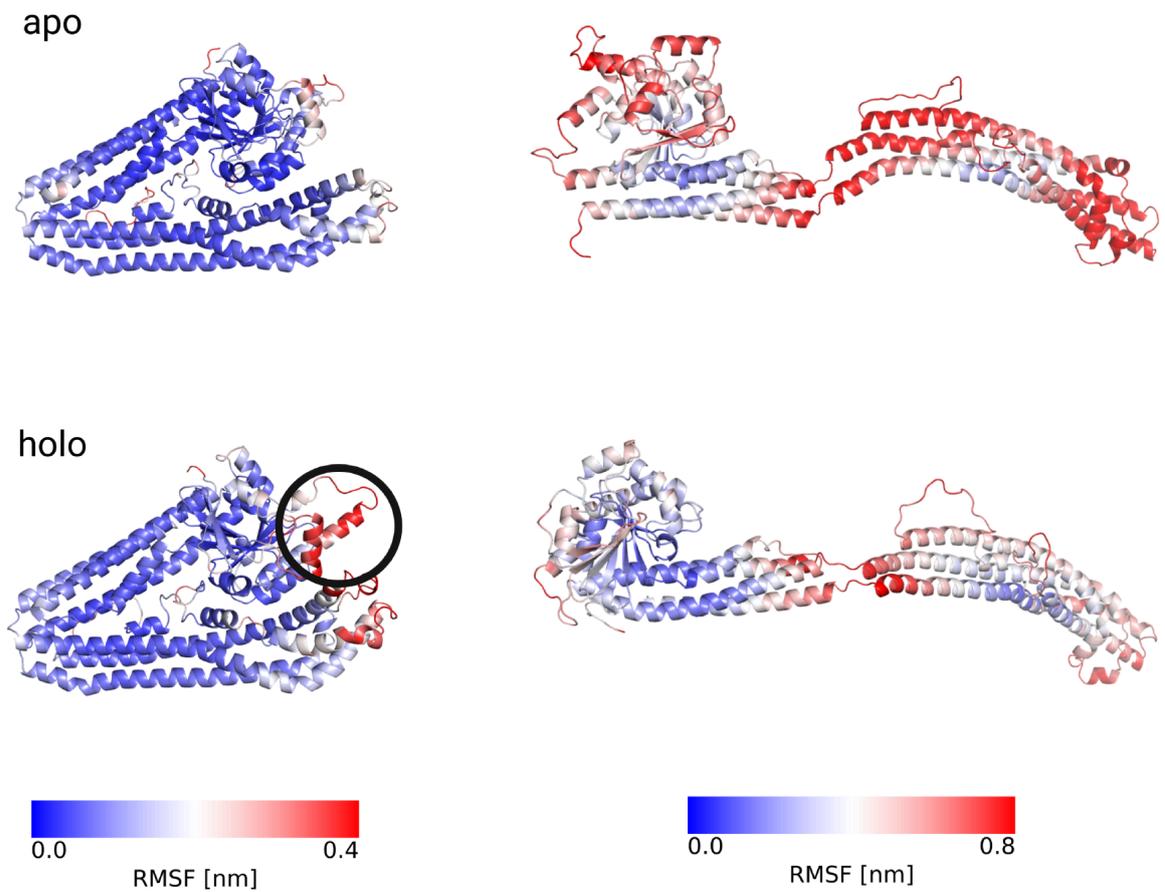
## 2 Results

### 2.1 Open BDLP is highly dynamic and can adopt a semi-closed conformation

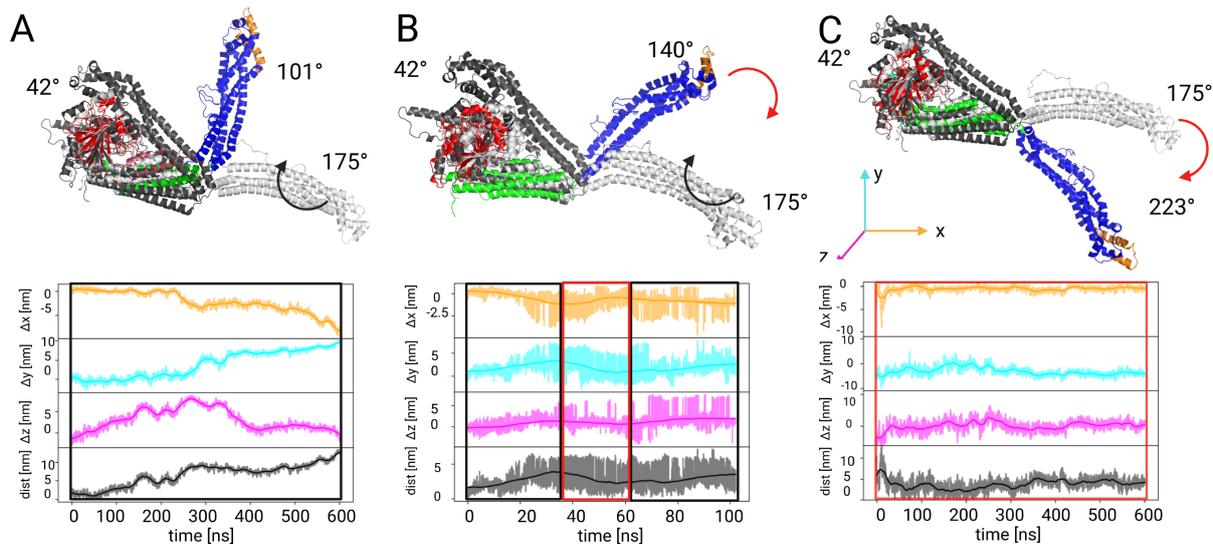
We started the study by exploring the overall stabilities and local flexibilities of apo- and holo-BDLP in the open and closed states in unbiased MD simulations. The analysis of protein flexibilities, as measured by the root mean square fluctuations of the  $C_\alpha$  atoms (RMSF), revealed that on the simulated timescale of 600 ns, both apo- and holo-BDLP in the closed conformation are stable, while the open forms displayed more flexibilities, yet more so in the case of apo-BDLP (Figure 2). Indeed, the open conformation of apo-BDLP started to spontaneously close (Figure 3A), reaching a semi-closed state with  $\alpha \approx 100^\circ$ . In the closed

holo-BDLP state, GTP binding induced flexibilities, yet it remained closed. Nonetheless, the different flexibilities in closed apo- and holo-BDLP indicate that GTP loading may have the potential to drive the protein towards the open conformation, while apo-BDLP prefers closed conformations.

To further explore possible conformational transitions, an HREMD simulation of apo-BDLP starting from the open conformation was performed. Here, the switching between the open and semi-closed state (here with  $\alpha \approx 140^\circ$ ) was sampled more often and reversibly (Figure 3B). We then simulated both the open and closed form of holo-BDLP using HREMD. The setup for the closed state confirmed the observation from the MD simulations that there are structural instabilities at the interface between the G domain and the paddle (Figure 2). They result from motions in the G domain set off by GTP binding, especially involving  $\alpha 5$  (residues 143–156) and  $\alpha 9$  plus the preceding loop (residues 251–271), which we denote flap1 and flap2 because of their swinging motions further discussed below. The open holo-BDLP, on the other hand, quickly assumed  $\alpha$  angles over  $180^\circ$ , hereafter referred to as wide-open conformation (Figure 3C), which is highly reminiscent of human dynamin 1 and has not been observed for BDLP before. Another interesting observation is that the open-to-closed motion in BDLP involves movements of the trunk into all three spatial directions, as resolved by the analysis of the stalk tip motions ( $\Delta x$ ,  $\Delta y$ , and  $\Delta z$  in Figure 3), so that in fact hinge1 is rather a ball joint, at least a restricted ball joint, than just a hinge as its name suggests.



**Figure 2:** Root mean square fluctuations of the  $C_{\alpha}$  atoms observed during 600 ns MD simulations of BDLP in different states: apo/closed, apo/open, holo/closed, holo/open. The RMSF values are project onto the cartoon presentations of BDLP, using colors ranging from blue for low RMSF to red for high RMSF. The exact color scales are given below the less mobile closed states (left) and the more flexible open states (right). The simulations revealed the existence of two important structural elements in the G domain that interact with the paddle in the closed form and respond to GTP binding by conformational changes. Because of their flapping motion, we called them flap1 and flap2. Their location is highlighted by a black circle in the representation of the closed holo-BDLP.



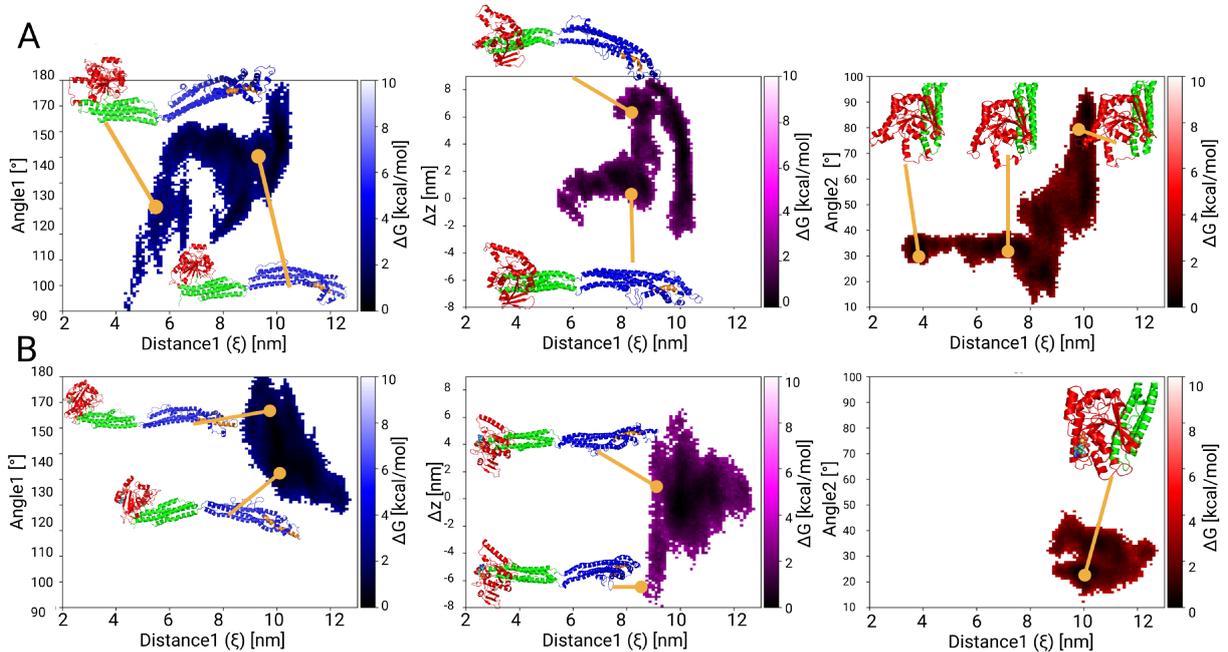
**Figure 3: Hinge movement of BDLP illustrated by structures (top) and deviations of the stalk-tip position from its initial location (bottom).** The movements are quantified by  $\Delta x$ ,  $\Delta y$ ,  $\Delta z$ , and overall distance, where the opening and closing directions are indicated by black and red boxes, respectively in the bottom panels. **A** In the MD simulation that started from the open conformation (white) of apo-BDLP, the protein assumed a semi-closed structure with  $\alpha \approx 100^\circ$ . For comparison, the fully closed state is shown (black), which remained closed in an MD simulation that started from this conformation. **B** In the HREMD simulation of apo-BDLP that started from the open conformation, both closing and opening movements were observed, with a minimal  $\alpha \approx 140^\circ$  that was reached. **C** In the HREMD simulation of holo-BDLP, starting from the open conformation, the conformation not only remained in the open state, but even adopted a wide-open conformation with  $\alpha$  reaching up to  $\approx 225^\circ$ .

## 2.2 Hinge1 is not just a hinge but a restricted ball joint

A more general overview of the conformational landscape of BDLP is provided by free energy surfaces. The free energy ( $\Delta G$ ) is projected onto different combinations of the order parameters  $\alpha$  and  $\Delta z$ , which were already discussed above, as well as the distance  $\xi$  to measure the amount of opening and the angle  $\beta$  to assess hinge2 movements. The results in Figure 4A and B are shown for the HREMD simulations of the open forms of apo- and holo-BDLP, respectively. The  $\Delta G(\xi, \alpha)$  values for apo-BDLP confirm that it closes spontaneously to reach the semi-closed state, which leads to reductions in both  $\xi$  and  $\alpha$  values. However, for angles between  $110$  and  $150^\circ$ , different distances between the stalk tip and the G domain can

be assumed (around 6 nm and 8–10 nm). This degeneracy indicates that motions in another direction must occur in parallel, and this third dimension is the lateral motion  $\Delta z$ . The projection  $\Delta G(\xi, \Delta z)$  shows that the trunk of BDLP upon closing first moves into positive  $\Delta z$  direction (see Figure 3 for the definition of the coordinate system) and then starts to move into the opposite direction at  $\xi \approx 9$  nm. This confirms the conclusion made above that hinge1 is more than a hinge, it can rather be considered a restricted ball joint. The hinge1 movements are correlated with hinge2 movements, as revealed by snapshots in Figure S1. The monomeric BDLP in open conformation was constructed from a dimer structure, where the G domain is rotated against the neck precisely at hinge2, and in part this difference is conserved during simulation. In the closed conformation, the hinge2 helical region was resolved, while the rotated open conformation had to be modeled with a disordered loop region. However, the disordered hinge2 region becomes more helical with GTP being bound.

The corresponding  $\Delta G$  plots for holo-BDLP in Figure 4B confirm the observation made above that GTP binding stabilizes the open conformation and even induces a wide-open conformation with  $\alpha > 180^\circ$  and  $\xi > 10$  nm. The movements toward the wide-open structure also involve lateral motions of the trunk, yet into negative  $z$  direction. Thus, using the BDLP presentation as shown here, with the G domain on the left, one can state that apo-BDLP closes with the trunk moving towards the reader, while holo-BDLP wildly opens away from the reader. Since no major closing in holo-BDLP took place, hinge2 did not change much. However, the hinge2 region displays a more structured head-neck conformation, which in apo-BDLP was only reached upon hinge1 closing. Especially at  $\xi \approx 8$  nm a coupling between the stalk and the G domain in apo-BDLP must occur, as at that distance not only the lateral movement of the stalk tip abruptly reverts back to  $\Delta z \approx 0$ , but also  $\beta$  converts to smaller angles here (Figure S1). The lateral motion restriction could be due to the interaction of a loop in the trunk region (490–510) with the neck, guiding the swivel motion into a more linear approach towards the G domain.

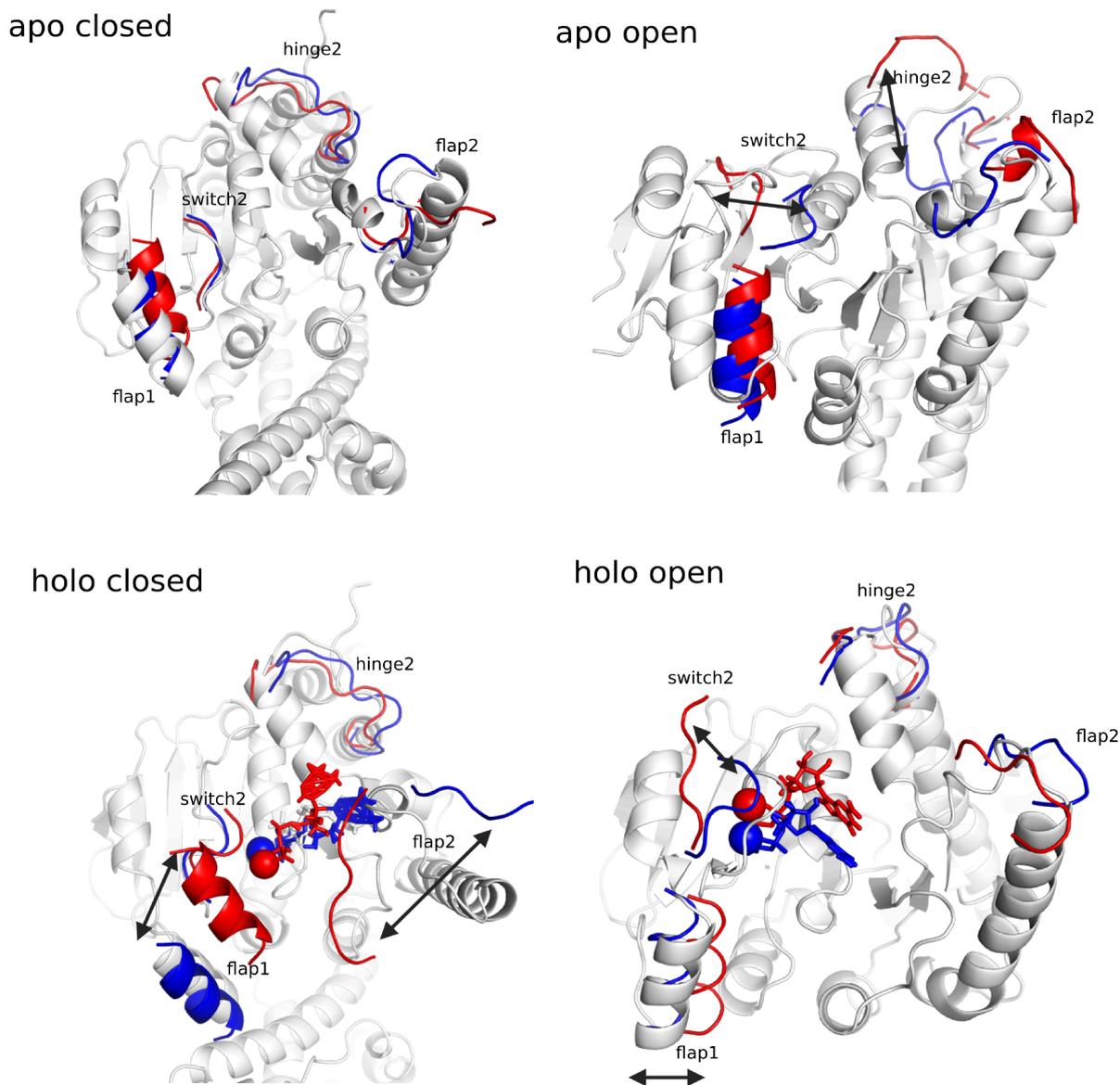


**Figure 4: Free energy surfaces as a function of selected order parameters.** The free energy is shown as 2D projections  $\Delta G(\xi, \alpha)$  (left),  $\Delta G(\xi, \Delta z)$  (middle), and  $\Delta G(\xi, \beta)$  (right), along with representative structures for **A** apo-BDLP and **B** holo-BDLP.  $\Delta G$  values are color-coded according to the scales on the right of the plots. Low  $\Delta G$  values indicate stable structures.

### 2.3 GTP binding sets off long-distance communication

In order to understand the coupling between the G domain and stalk, we continue by analyzing the structural flexibilities of the G domain and how they are affected by GTP binding. To this end, a principal component analysis of the most mobile regions of the G domain was performed, which revealed the swinging motions of the previously mentioned flap1 and flap2 (Figure 5). Especially the closed conformation of holo-BDLP displays pronounced flap1 and flap2 movements to cover GTP. G domains are typically characterized by the presence of certain motifs that are directly or indirectly involved in GTP binding. The only canonical GTPase motif that can be clearly identified in BDLP is the P-loop at residues 76–84. The other motifs deviate, but can be mapped to G2 (or switch I, 102–103), G3 (or switch II, 180–184), G4 (238–241), and G5 (245–268). Moreover, hinge2 also belongs to the G domain (Figure 5) and its conformational change is of relevance for the coupling between the domains.

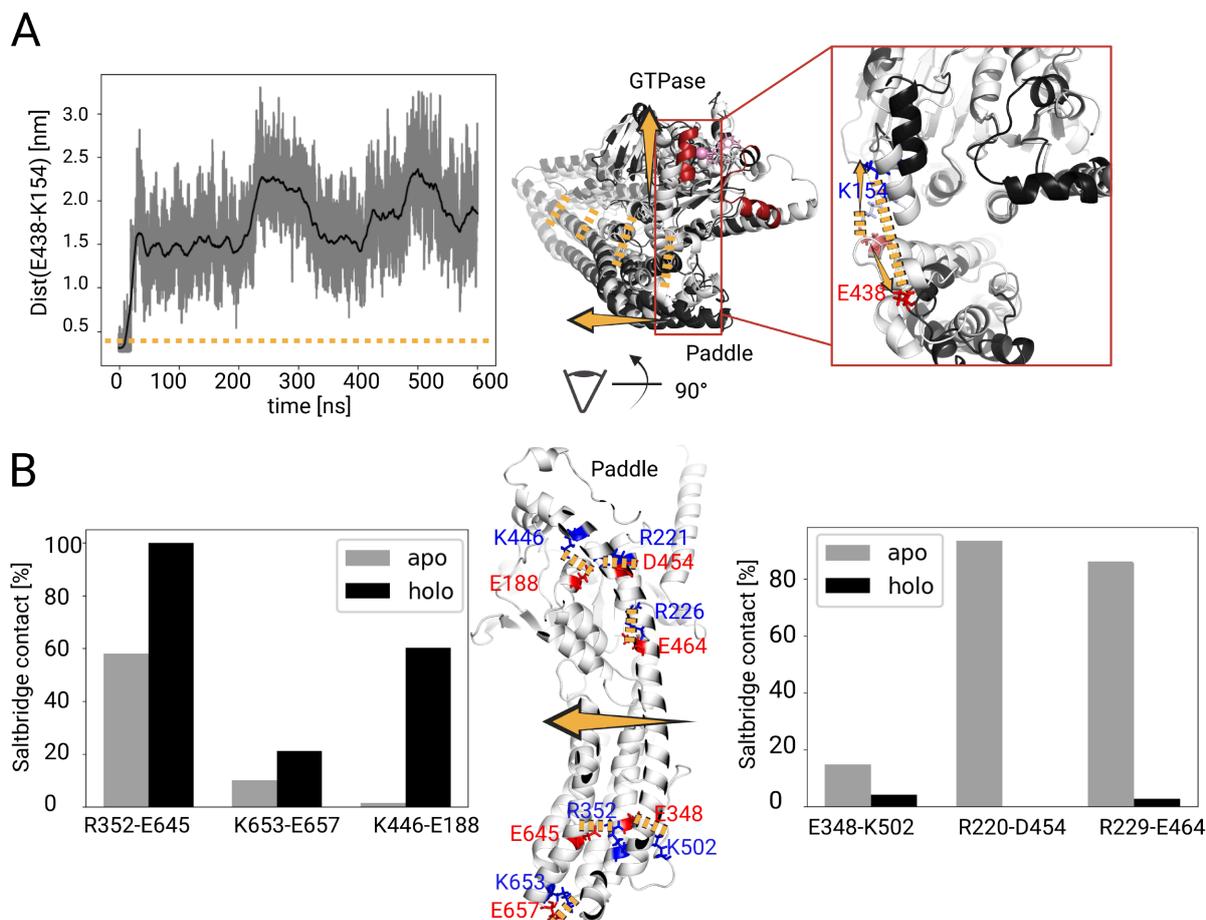
As previously observed in our studies of murine guanylate-binding protein 2,<sup>34</sup> GTP binding has the capability to stiffen the G domain. This also occurs in BDLP when it is in the open conformation. Conformational clustering of all G domain motifs produced 103 rather ordered clusters for holo-BDLP, while this number sextuples for apo-BDLP. Here, 624 clusters are found and a high variance among the cluster conformations of hinge2 and flap2 is present. However, the situation reverses for the closed form of BDLP. In that case, in the GTP-bound state even more clusters are found than for apo-BDLP: 178 clusters for holo-BDLP with a high variance in hinge2 and flap2 and only 61 clusters with a high degree of structural order for apo-BDLP. This suggests that the open conformation can better accommodate GTP, while GTP binding induces structural instabilities in the closed state that might induce further structural changes in the stalk region.



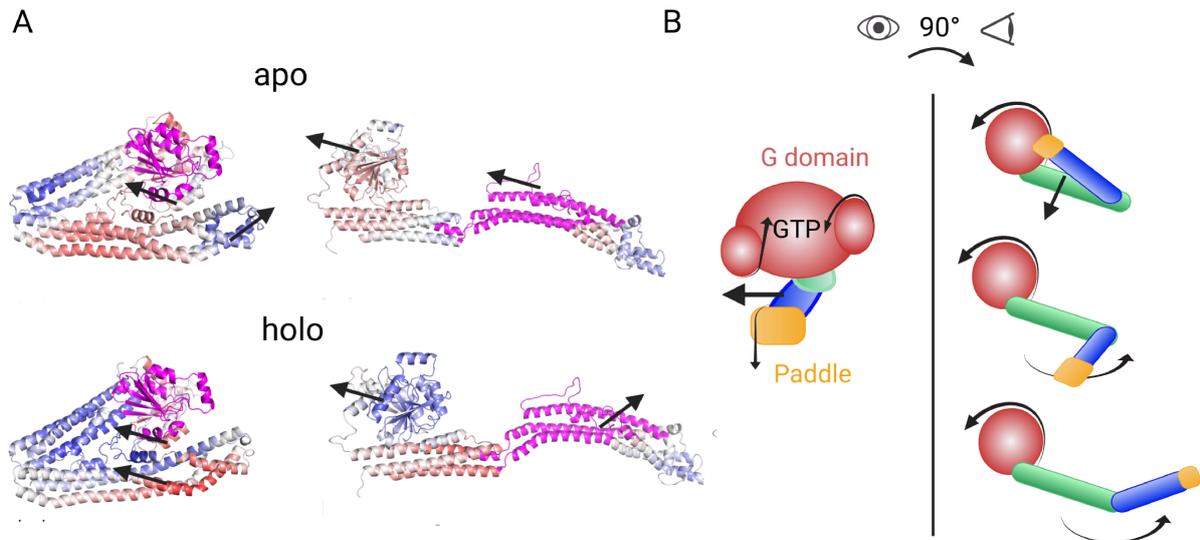
**Figure 5: Principal motions of important structural elements of the G domain in apo-BDLP (top) and holo-BDLP (bottom) in the closed (left) and open (right) conformations.** Red and blue mark the maximum displacement in either direction of the motions of flap1, flap2, hinge2, and switch2. In the case of holo-BDLP, also the relocations of  $Mg^{2+}$  and GTP are shown. The arrows indicate the respective movement.

With the aim to identify communication pathways between the G domain and the stalk, we identified salt bridges that change their occupancies upon GTP binding in the MD simulations of closed BDLP. The most obvious candidate that might function as an allosteric

switch is a salt bridge connecting flap1 with the trunk, K154–E438, which is dissolved by a flap1 relocation. In the GTP-bound state, flap1 and thus K154 moved towards GTP and thus away from E438 (Figure 6A), located adjacent to the paddle domain. Two further salt bridges between the G domain and the trunk broke: R221–D454 and R226–E464. Of note is also the salt bridge E348–K502 connecting the neck with the trunk, whose stability also decreased followed GTP binding. Interestingly, these residue pairs are all on the same side of BDLP (which is the right side in the protein presentation in Figure 6B). Contrariwise, on the other side of the protein, there are three salt bridges that gained in strength upon GTP binding: E188–K446 connecting the G domain with the trunk, R352–E645 between the neck and the trunk, and K653–E657 associating the loop region between the trunk and neck with the paddle domain. This weakening and strengthening of salt bridges across the protein strikingly reveals how GTP binding gives rise to conformational changes in the closed BDLP that are expected to facilitate the closed-to-open transition. The salt bridges thus represent important allosteric switches that enable the information flow from the GTP-binding site in the G domain to the different parts of the stalk region. Moreover, the two-sided distribution of weakened/dissolved and strengthened/newly formed salt bridges further confirms the observation made above that movements of hinge1 are more than a swinging motion but also involve lateral (or shearing) movements. This is further supported by a normal mode analysis of the closed conformations (Figure 7), which shows that in holo-BDLP the two longitudinal sides of the stalk move in opposite directions. This is not the case in the closed apo-BDLP.



**Figure 6: Changes in salt bridges following GTP binding in the closed BDLP.** **A** The time series of the K154–E438 distance in holo-BDLP is shown (left). A salt bridge is considered to be present when the distance between the N atom of the Lys side chain and either O atom of the Glu side chain is below 0.45 nm (dotted yellow line). Breaking of this salt bridge releases the trunk from the G domain in that area (middle). The motions of the G domain and the stalk are indicated by yellow arrows, and flap1 and flap2 being highlighted in red. The initial and final protein conformations of the MD simulation are shown as light-gray and black cartoon, respectively. Zooming into the K154–E438 region shows how flap1 and the trunk moved away from each other, breaking that salt bridge (right). **B** Occupancies of salt bridges that strengthen (left) or weaken (right) as a result of GTP binding. These strengthened and weakened salt bridges are located at different longitudinal sides of the trunk, giving rise to a shearing motion.



**Figure 7: Normal mode analysis of BDLP initial conformations.** **A** The normal mode correlation matrix projected onto the structure, where blue indicates anticorrelated and red correlated motions. The regions used as reference are colored in magenta. A reversal of correlations is observed for two cases, when using the flap1-half of the G-domain and the trunk as reference. Arrows are added to indicate the direction of selected motions. More data from that analysis can be found in Figure S2. **B** The principal motions are summarized in a cartoon.

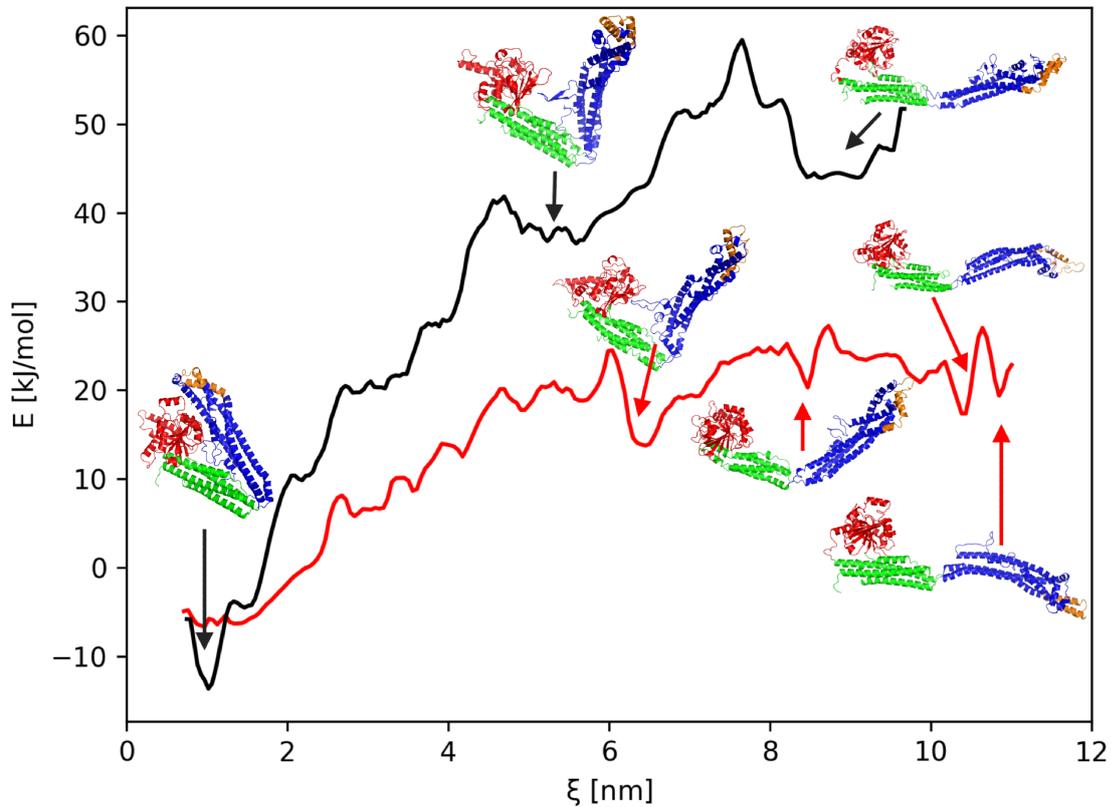
## 2.4 BDLP opening takes less than GTP hydrolysis

Since neither in the standard MD nor in the HREMD simulations the open-to-closed (or closed-to-open) pathway was fully sampled but only led to the semi-closed state, we resorted to USMD to enforce that transition in both apo- and holo-BDLP and calculate the corresponding free energy (also called potential of mean force here<sup>36</sup>). To create initial structures for the USMD windows, we employed a pulling simulation for the closed-to-open transition of apo-BDLP (see Figure S3A for the pulling force applied). That pulling simulation once again confirmed that the hinge1 opening is accompanied by lateral motions of the trunk. Moreover, during the pulling simulation of apo-BDLP, hinge1 opening also coincides with an opening of the GTP binding pocket, an effect which is already visible in Figure 5. For the USMD windows we preferentially used structures sampled in the unbiased simulations and only for the missing parts of the complete pathway, conformations from the pulling simulation

were taken. We obtained a range of ca. 100 intermediate initial conformations between  $\xi = 0.8$  nm, which is the neck–trunk distance of the closed conformation, and  $\xi = 11$  nm for the open state. These windows were restrained with harmonic potentials (see Figure S3B for the distribution of the windows) and simulated for 100 ns. The potential of mean force (PMF) obtained from applying the weighted histogram analysis method (WHAM) to the USMD simulations data of apo-BDLP is shown in Figure 8. We checked for convergence of these simulations by extending each window to 300 ns. The resulting changes in the free energy profile are marginal, and therefore 100 ns is used here (Figure S3C). The complete opening of apo-BDLP requires more than 60 kJ/mol. The initial steep rise of the energy profile between  $\xi = 1$  and 4.5 nm can be attributed to the breaking of several salt bridges between the neck and trunk. At  $\xi \approx 5$  nm, a metastable conformation corresponding to the semi-closed state already sampled in the MD and HREMD simulation is encountered. This structure is stabilized by a loop of the trunk (490–510) becoming a  $\beta$ -hairpin that forms a contact with the G domain. Once this contact is broken, the opening proceeds until an energy valley corresponding to (almost) fully opened apo-BDLP conformations is hit. We repeated the USMD simulations after having docked GTP into the GTP binding site of the starting conformation of each window and adding  $\text{Mg}^{2+}$  to it. The PMF after 100 ns per US window is shown in Figure 8 (for the error estimate, see Figure S3D). The overall energetic threshold for the closed-to-open transition is lowered to about 30 kJ/mol. Thus, the binding of GTP renders the open conformation more favorable.

As discussed above, the addition of GTP sets off a number of structural changes in the G domain, which in turn break key salt bridges between the G domain and the trunk, enabling the opening at hinge1. However, also hinge2 is affected by GTP binding, as it causes the G domain to roll away from the neck, which lowers the energy barrier for dissolving the interaction between the loop of the trunk and the G domain that is present in the intermediate at  $\xi \approx 6.5$  nm. It should be noted, however, that both 60 kJ/mol and 30 kJ/mol are in principle accessible by the hydrolysis of a single GTP molecule,<sup>37,38</sup> which releases 60 kJ/mol.

On the other hand, one also needs to consider the time scales involved for the different energetic barriers. Using the Arrhenius equation,  $\tau = \tau_0 \exp(\Delta G^\# / k_B T)$  with  $\Delta G^\#$  as the overall free energy barrier,  $T = 310$  K,  $k_B$  being the Boltzmann constant, and  $\tau_0 \approx 10^{-12}$  s at 310 K, we obtain that, on average, more than 7 ms are needed to for the closed-to-open transition of apo-BDLP, while this reduces to 100–200 ns for holo-BDLP. The conformational change of BDLP can thus take place more quickly than GTP hydrolysis and before the dimer dissociates, and is well below the seconds to minutes reported for membrane fission itself.<sup>17</sup> A valid question though is why we did not sample that transition in the 600 ns MD or HREMD simulation of holo-BDLP. The most likely answer is that with USMD one concentrates on a narrow part of the energy landscape between two end states, while in unbiased simulations the protein has more possibilities to explore its conformational space and does not necessarily strike out to the end state we expect it to develop to. The same observation we made in a study of a much smaller movement, of a loop in triosephosphate isomerase that switches between an open and closed conformation depending on the substrate-loading state.<sup>39</sup>



**Figure 8: The free energy of the close-to-open transition of BDLP.** These results are obtained from umbrella sampling MD simulations apo-BDLP (black) and holo-BDLP (red). The energies smoothed with a uniform filter. Representative structures of local minima are show.

### 3 Discussion

Combining the analysis of the BDLP conformations and that of salt bridges with the knowledge gained about the forces needed for the closed-to-open transition, we deduce that the energetic barrier for BDLP opening mostly consists of the breaking of six salt bridges between the neck and trunk, after which less force is needed to complete the opening. The closed-to-open transition not only involves a hinge1 movement, but involves a shearing motion in lateral direction to minimize the electrostatic force needed. The open form of BDLP is accessible by the energy released from the hydrolysis of a single GTP molecule, and just the binding

of GTP lowers the energy barrier for the opening. This allosteric action of GTP binding is relayed from the GTP binding site via the motions of two flaps in the G domain, closing to cover the GTP-loaded site, which breaks salt bridges between the G domain and the trunk and initiates the detachment of the paddle from the G domain. Salt bridges on the other longitudinal side of the trunk are formed, which guides the lateral motion of the stalk region during opening, while the release of salt bridges between stalk and G domain cause the latter to roll away at hinge2 (Figure S1). However, further hinge1 opening coincides with an opening of the GTP binding pocket (Figure 5), which explains why GTP hydrolysis by dynamins usually requires the dynamins to be polymerized where the G domains would stabilize each other to provide the structural stability needed for the hydrolysis reaction to take place. The mechanism of binding GTP and hinge1 opening must thus be connected in both directions, since the open form is more likely to be assumed with GTP (and not GDP) bound, in polymeric, membrane-associated BDLP.<sup>22</sup> Vice versa, the polymeric form must communicate its status to the GTPase domain, in order to stimulate GTPase activity, and possibly to facilitate GDP release afterwards. We thus agree with<sup>21</sup> that GTP binding causes the hinge1 opening and provide the mechanistic behind that transition. The PMF of the open-to-closed transition in Figure 8 further shows that the open state of the BDLP monomer in solution is less stable than than the closed state, which clarifies why Low et al. only observed the open holo-BDLP form when being membrane-bound and polymeric. Importantly, GTP hydrolysis cannot be the cause for the opening, since membrane tubulation also happens with non-hydrolyzable GTP analogues. Moreover, GDP is thought to cause depolymerization and membrane dissociation.<sup>22</sup> The insights gained about BDLP's mechanism of action may be transferable to other DLPs, contributing to our understanding of fundamental cellular processes, such as endocytosis, cell division, and immune response.

## 4 Methods

The open and closed structures of BDLP were taken from the Protein Data Bank (2J68 and 2W6D)<sup>22,23</sup> and completed using RCD+, GalaxyLoop, DaReUs-Loop and ModLoop.<sup>40–43</sup> In the all simulations, AMBERff99SB\*ILDNP<sup>44</sup> and the TIP3P water model were used for modeling the protein and its surrounding. After energy minimization, MD simulations for equilibration in the NVT and NpT ensemble were performed, followed by production MD runs, HREMD or USMD. The protein was sampled in its apo form (4 fs timestep) and with GTP and Mg<sup>2+</sup> bound (2 fs timestep), called apo- and holo-BDLP henceforth. Depending on the protein conformation, the system sizes were between 300,000 and 1,000,000 atoms. All simulations, including the HREMD and USMD simulations, were performed with GROMACS (version 2016.4).<sup>26</sup> For the HREMD simulations, 30 replicas with 100 ns per replica were used, resulting in an average acceptance rate of 30% for exchanges between the replicas. For the USMD simulations, 74 and 67 windows with 100 ns per window were used for apo-BDLP and holo-BDLP, respectively. To test for convergence, the USMD windows for apo-BDLP were extended to 300 ns. Initial conformations for the windows were generated in preceding pulling simulations, and the resulting conformations were restrained with a harmonic potential in the USMD simulations. The free energy profile from the USMD simulations was obtained from a WHAM analysis.<sup>36,45</sup> All simulations performed in this study are summarized in Table S1; they yielded an accumulated simulation time of 19.5  $\mu$ s. Different GROMACS tools were used for analysis, among them the calculation of the RMSF, various distances and angles, spatial distributions, cluster analysis, and principal component analysis (PCA). The RMSF values between 0.1 and 4 Å were projected onto the structures, with red areas denoting flexible residues and blue areas rigid ones (RMSF > 2 Å and  $\leq$  2 Å, respectively). The distances and angles used to characterize the hinge movements are defined in Figure 1. To calculate the displacement of the stalk tip compared to the initial position, denoted as  $\Delta x$ ,  $\Delta y$ , and  $\Delta z$ , a custom Tcl script was used in VMD<sup>46</sup> text mode. Here,  $\Delta z$  corresponds to lateral motions of the trunk. The VMD-integrated tool ProDy<sup>47</sup> was used to study normal modes of the initial

structures. Free energy surfaces were created using a custom Python script. In the analysis of the HREMD simulations, only the corresponding target replica was used. Figures were generated with PyMOL, Inkscape, and BioRender.

More details about the simulation and analysis methods are provided in the Supplementary Information.

## 5 Acknowledgments

B.S. gratefully acknowledges funding from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) (project A07 of the CRC 1208). W.S. is grateful for a PhD fellowship from Jürgen Manchot Stiftung. The project was made possible thanks to computing time granted through the Gauss Centre for Supercomputing e.V. ([www.gauss-centre.eu](http://www.gauss-centre.eu)) on the GCS Supercomputer SuperMUC-NG at Leibniz Supercomputing Centre ([www.lrz.de](http://www.lrz.de)) (project pn98zo).

## 6 Author contributions

B.S. designed the research. W.S. performed the simulations and data analysis. W.S. and B.S. wrote and reviewed the manuscript. B.S. supervised the project.

## 7 Additional information

Supplementary information is available.

## References

- (1) Morlot, S.; Galli, V.; Klein, M.; Chiaruttini, N.; Manzi, J.; Humbert, F.; Dinis, L.; Lenz, M.; Cappello, G.; Roux, A. Membrane Shape at the Edge of the Dynamically Helix

- Sets Location and Duration of the Fission Reaction. *Cell* **2012**, *151*, 619.
- (2) Ford, M. G. J.; Chappie, J. S. The structural biology of the dynamin-related proteins: New insights into a diverse, multitasking family. *Traffic* **2019**, *20*, 717–740.
  - (3) Praefcke, G. J. K.; McMahon, H. T. The dynamin superfamily: universal membrane tubulation and fission molecules? *Nature Reviews Molecular Cell Biology* **2004**, *5*, 133–147.
  - (4) Melen, K.; Ronni, T.; Broni, B.; Krug, R. M.; Von Bonsdorff, C. H.; Julkunen, I. Interferon-induced Mx proteins form oligomers and contain a putative leucine zipper. *Journal of Biological Chemistry* **1992**, *267*, 25898–25907.
  - (5) Daumke, O.; Roux, A.; Haucke, V. BAR domain scaffolds in dynamin-mediated membrane fission. 2014.
  - (6) Tuma, P. L.; Collins, C. A. Activation of dynamin GTPase is a result of positive cooperativity. *Journal of Biological Chemistry* **1994**, *269*, 30842–30847.
  - (7) Warnock, D. E.; Hinshaw, J. E.; Schmid, S. L. Dynamin Self-assembly Stimulates Its GTPase Activity. *Journal of Biological Chemistry* **1996**, *271*, 22310–22314.
  - (8) Ghosh, A.; Praefcke, G. J. K.; Renault, L.; Wittinghofer, A.; Herrmann, C. How guanylate-binding proteins achieve assembly-stimulated processive cleavage of GTP to GMP. *Nature* **2006**, *440*, 101–104.
  - (9) Narayanan, R.; Leonard, M.; Song, B. D.; Schmid, S. L.; Ramaswami, M. An internal GAP domain negatively regulates presynaptic dynamin in vivo: a two-step model for dynamin function. *The Journal of Cell Biology* **2005**, *169*, 117.
  - (10) Jimah, J. R.; Hinshaw, J. E. Structural Insights into the Mechanism of Dynamin Superfamily Proteins. 2019; <https://www.sciencedirect.com/science/article/pii/S0962892418301892>.

- (11) Ferguson, S. M.; De Camilli, P. Dynamin, a membrane-remodelling GTPase. *Nature Reviews Molecular Cell Biology* **2012**, *13*, 75–88.
- (12) Sever, S.; Muhlberg, A. B.; Schmid, S. L. Impairment of dynamin’s GAP domain stimulates receptor-mediated endocytosis. *Nature* *1999* *398:6727* **1999**, *398*, 481–486.
- (13) Bashkirov, P. V.; Akimov, S. A.; Evseev, A. I.; Schmid, S. L.; Zimmerberg, J.; Frolov, V. A. GTPase Cycle of Dynamin Is Coupled to Membrane Squeeze and Release, Leading to Spontaneous Fission. *Cell* **2008**, *135*, 1276–1286.
- (14) Colom, A.; Redondo-Morata, L.; Chiaruttini, N.; Roux, A.; Scheuring, S. Dynamic remodeling of the dynamin helix during membrane constriction. *Proceedings of the National Academy of Sciences of the United States of America* **2017**, *114*, 5449–5454.
- (15) Chen, Y.; Zhang, L.; Graf, L.; Yu, B.; Liu, Y.; Kochs, G.; Zhao, Y.; Gao, S. Conformational dynamics of dynamin-like MxA revealed by single-molecule FRET. *Nature Communications* **2017**, *8*, 15744.
- (16) Cocucci, E.; Gaudin, R.; Kirchhausen, T. Dynamin recruitment and membrane scission at the neck of a clathrin-coated pit. *Molecular Biology of the Cell* **2014**, *25*, 3595–3609.
- (17) Antony, B. et al. Membrane fission by dynamin: What we know and what we need to know. *EMBO J.* **2016**, *35*, 2270–2284.
- (18) Stowell, M. H.; Marks, B.; Wigge, P.; McMahon, H. T. Nucleotide-dependent conformational changes in dynamin: evidence for a mechanochemical molecular spring. *Nature cell biology* **1999**, *1*, 27–32.
- (19) Chen, Y. J.; Zhang, P.; Egelman, E. H.; Hinshaw, J. E. The stalk region of dynamin drives the constriction of dynamin tubes. *Nature structural & molecular biology* **2004**, *11*, 574–575.

- (20) Roux, A.; Uyhazi, K.; Frost, A.; De Camilli, P. GTP-dependent twisting of dynamin implicates constriction and tension in membrane fission. *Nature* **2006**, *441*, 528–531.
- (21) Low, H. H.; Sachse, C.; Amos, L. A.; Löwe, J. Structure of a Bacterial Dynamin-like Protein Lipid Tube Provides a Mechanism For Assembly and Membrane Curving. *Cell* **2009**, *139*, 1342–1352.
- (22) Low, H. H.; Löwe, J. A bacterial dynamin-like protein. *Nature* **2006**, *444*, 766–769.
- (23) Low, H. H.; Löwe, J. Dynamin architecture — from monomer to polymer. *Current Opinion in Structural Biology* **2010**, *20*, 791–798.
- (24) Ciccotti, G.; Dellago, C.; Ferrario, M.; Hernández, E. R.; Tuckerman, M. E. Molecular simulations: past, present, and future (a Topical Issue in EPJB). *The European Physical Journal B* **2022**, *95*, 1–12.
- (25) Strodel, B.; Barz, B. In *Progress in Molecular Biology and Translational Science*; Strodel, B., Barz, B., Eds.; Academic Press: London, UK, 2020.
- (26) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1-2*, 19–25.
- (27) Müller, M.; Katsov, K.; Schick, M. A new mechanism of model membrane fusion determined from Monte Carlo simulation. *Biophysical Journal* **2003**, *85*, 1611–1623.
- (28) Mattila, J.; Shnyrova, A.; Sundborger, A.; Hortelano, E.; Fuhrmans, M.; Neumann, S.; Müller, M.; Hinshaw, J.; Schmid, S.; Frolov, V. A hemi-fission intermediate links two mechanistically distinct stages of membrane fission. *Nature* **2015**, *524*, 109–113.
- (29) Fuhrmans, M.; Müller, M. Coarse-grained simulation of dynamin-mediated fission. *Soft Matter* **2015**, *11*, 1464–1480.

- (30) Pannuzzo, M.; McDargh, Z.; Deserno, M. The role of scaffold reshaping and disassembly in dynamin driven membrane fission. *eLife* **2018**, *7*.
- (31) Noel, J.; Noé, F.; Daumke, O.; Mikhailov, A. Polymer-like Model to Study the Dynamics of Dynamin Filaments on Deformable Membrane Tubes. *Biophysical Journal* **2019**, *117*, 1870–1891.
- (32) Kadosh, A.; Colom, A.; Yellin, B.; Roux, A.; Shemesh, T. The tilted helix model of dynamin oligomers. *Proceedings of the National Academy of Sciences of the United States of America* **2019**, *116*, 2845–12850.
- (33) Barz, B.; Loschwitz, J.; Strodel, B. Large-scale, dynamin-like motions of the human guanylate binding protein 1 revealed by multi-resolution simulations. *PLOS Computational Biology* **2019**, *15*, e1007193.
- (34) Loschwitz, J.; Steffens, N.; Wang, X.; Schäffler, M.; Pfeffer, K.; Degrandi, D.; Strodel, B. Domain motions, dimerization, and membrane interactions of the murine guanylate binding protein 2. *Sci. Rep.* **2023**, *13*, 679.
- (35) Vöpel, T.; Hengstenberg, C. S.; Peulen, T. O.; Ajaj, Y.; Seidel, C. A.; Herrmann, C.; Klare, J. P. Triphosphate induced dimerization of human guanylate binding protein 1 involves association of the C-terminal helices: A joint double electron-electron resonance and FRET Study. *Biochemistry* **2014**, *53*, 4590–4600.
- (36) Kästner, J. Umbrella sampling. *WIREs Computational Molecular Science* **2011**, *1*, 932–942.
- (37) Berg, J. M.; Tymoczko, J. L.; Gatto, G. J.; Stryer, L. Stryer Biochemie. *Stryer Biochemie* **2018**,
- (38) Atkins, P. W.; De Paula, J. *Physical chemistry for the life sciences*; W. H. Freeman: London, UK, 2005; p 590.

- (39) Liao, Q.; Kulkarni, Y.; Sengupta, U.; Petrović, D.; Mulholland, A. J.; van der Kamp, M. W.; Strodel, B.; Kamerlin, S. C. L. Loop Motion in Triosephosphate Isomerase Is Not a Simple Open and Shut Case. *Journal of the American Chemical Society* **2018**, *140*, 15889–15903.
- (40) Park, H.; Lee, G. R.; Heo, L.; Seok, C. Protein loop modeling using a new hybrid energy function and its application to modeling in inaccurate structural environments. *PLoS ONE* **2014**, *9*.
- (41) López-Blanco, J. R.; Canosa-Valls, A. J.; Li, Y.; Chacón, P. RCD+: Fast loop modeling server. *Nucleic Acids Research* **2016**, *44*, W395–W400.
- (42) Karami, Y.; Rey, J.; Postic, G.; Murail, S.; Tufféry, P.; de Vries, S. J. DaReUS-Loop: a web server to model multiple loops in homology models. *Nucleic Acids Research* **2019**, *47*, W423–W428.
- (43) Fiser, A.; Sali, A. ModLoop: automated modeling of loops in protein structures. *Bioinformatics* **2003**, *19*, 2500–2501.
- (44) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **2010**, *78*, 1950–1958.
- (45) Hub, J. S.; de Groot, B. L.; van der Spoel, D. g wham - A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates. *Journal of Chemical Theory and Computation* **2010**, *6*, 3713–3720.
- (46) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38.
- (47) Bakan, A.; Meireles, L. M.; Bahar, I. ProDy: Protein Dynamics Inferred from Theory and Experiments. *Bioinformatics* **2011**, *27*, 1575–1577.

**Supplementary Information:**

**Allosteric communication induced by GTP  
binding sets off a closed-to-open transition in a  
bacterial dynamin-like protein**

Wibke Schumann<sup>†,‡</sup> and Birgit Strodel<sup>\*,¶,‡</sup>

<sup>†</sup>*Institute of Theoretical and Computational Chemistry, Heinrich Heine University Düsseldorf,  
40225 Düsseldorf, Germany*

<sup>‡</sup>*Institute of Biological Information Processing: Structural Biochemistry, Forschungszentrum Jülich,  
52428 Jülich, Germany*

<sup>¶</sup>*Institute of Theoretical and Computational Chemistry, Heinrich Heine University Düsseldorf,  
40225 Düsseldorf, Germany*

E-mail: b.strodel@fz-juelich.de

# 1 Supplementary methods

The structures of BDLP were taken from the Protein Data Bank (2J68 and 2W6D)<sup>1,2</sup> and modeled using RCD+, GalaxyLoop, DaReUs-Loop and ModLoop<sup>3-6</sup>. For the MD simulations, the Amber99SB\*-ILDNP force field(<sup>7-9</sup> TIP3P water<sup>10</sup> and GRO-MACS 2016<sup>11</sup> were used, at a temperature of 303 K (37°C) and a pressure of 1 bar. In Table S1, all MD simulations performed in this study are summarized, which corresponds to a total simulated time of 19.5  $\mu$ s. GTP was parametrized according to the protocol described in<sup>12</sup> and docked to BDLP using AutoDock.<sup>13</sup> Several unrestrained simulation of BDLP were performed in order to explore the conformational space. The system sizes are listed in Table S1. 8 or 10 Na<sup>+</sup> were added in order to neutralize the apo- or holo-protein's charges, respectively, with an additional Mg<sup>2+</sup> being added to the holo simulations. First, an energy minimization was performed using a steepest descent algorithm, followed by a subsequent 0.1 ns NVT equilibration and a 1 ns NpT equilibration. During these steps, the protein's heavy atoms were restrained with a force constant of 10 kJ/mol\*Å<sup>2</sup>. The velocity rescaling thermostat was employed to regulate the temperature in the NVT simulations, while the Nosé-Hoover thermostat<sup>14,15</sup> and the isotropic Parrinello-Rahman barostat<sup>16</sup> were used for the NpT simulations. The particle mesh-Ewald (PME)<sup>17,18</sup> sum method was used to calculate electrostatic interactions with xyz-periodic boundary conditions. The van der Waals (vdW) and short-range Coulombic interaction cutoffs were set to 12 Å, repulsive Lennard-Jones (LJ) interactions were cut at 10 Å. The equations of motion were integrated using the leapfrog method, and the LINCS algorithm<sup>19</sup> was used to constrain all bonds. For the production rund, the same setup as in the NpT equilibration was used, minus the position restraints.

In the apo-BDLP systems, some hydrogen atoms were treated as virtual interaction sites, permitting an integration time step of 4 fs while maintaining energy conservation.<sup>20</sup> For the holo-BDLP systems, a time step of 2 fs was used. Coordinates and velocities were recorded every 20 ps in both cases. In all simulations, we applied position restraints for the rigid C $\alpha$  atoms of the G-domain  $\beta$ -sheets to remove overall translation and rotation, allowing us

to decrease the box size without harming the protein’s flexibilities.<sup>21</sup> To ensure that GTP stayed in its binding pocket, two distance restraints between GTP-atoms O3G/O2S and Ser97/Val243 of BDLP were applied using the GROMACS pull code, when necessary.

Next, we performed a Hamiltonian replica exchange MD (HREMD) simulation<sup>22</sup> with 30 replicas, each 100 ns per replica in length. The energy function of BDLP, and its protein-water interactions were modified in each but the target replica by applying biasing factors of  $310\text{ K}/T$ , with the 30 temperatures  $T$  exponentially distributed between 310 and 370 K ( $1 < \lambda < 0.667$ ). The unbiased target replica at 310 K was used for analysis. The average exchange probability between the replicas was ca. 30%. The HREMD simulations were conducted with Gromacs 2016.4 patched with the PLUMED plugin (version 2.4.1) (<sup>23</sup>). In all HREMD simulations, we used the v-rescale thermostat with canonical sampling and the Parrinello-Rahman barostat.<sup>16</sup> Apart from that, the setup was identical to the regular production runs described above.

**Table S1:** Simulations performed in this study, amounting to a total of 19.5  $\mu\text{s}$ .

simulation type	conformation	apo/holo	system size (atoms)	length
MD unbiased	open	apo	238,657	600 ns
MD unbiased	open	holo	194,959	600 ns
MD unbiased	closed	apo	284,578	600 ns
MD unbiased	closed	holo	194,905	600 ns
HREMD	open	apo	238,657	$30 \times 100$ ns
MD pulling	closed-to-open	apo	238,657	0.8 ns
USMD	–	apo	238,519 to 1,537,576	$74 \times 100$ ns
USMD	–	holo	787,203 to 1,536,653	$67 \times 100$ ns

From the unrestrained simulations, the predominant motion was identified using principal component analysis. This motion was then described using a distance, measured between the  $C_\alpha$  atoms of residues 224 and 453. This order parameter/reaction coordinate is referred to as  $\xi$  and used for the USMD simulations. In the USMD, 100 apo-BDLP conformations along  $\xi$  were simulated for 100 ns each. The force constants for restraining the conformations along  $\xi$  ranged between 3 and 6,200  $\text{kJ}/(\text{mol}\cdot\text{nm}^2)$ , depending on the stability of the different

windows; though the majority of the windows were simulated with 500 kJ/(mol·nm<sup>2</sup>). This setup was repeated for holo-BDLP. Of the 100 windows, only those which stayed in the range of the original window, and those where GTP remained bound at least 50% of the time were used for analysis.

The analysis was mainly performed with GROMACS-internal tools and included following calculations:

**WHAM** WHAM<sup>24</sup> was used as implemented in GROMACS,<sup>11</sup> for 74 (apo) or 67 (holo) trajectories of 100 ns each, at a temperature of 303 K. To ensure convergence, a subset of windows was extended to 300 ns, however, the energy profile did not change anymore after 100 ns. The profile shown in the main text was smoothed with a window size of 5 in reflect mode, while the extended data shows the original profile. Error bars were generated by 1000-fold bootstrapping, as implemented in *gmx wham*.

**Distances** Distance  $\xi$  was measured between the C <sub>$\alpha$</sub>  atoms of residues 224 and 453.

**Angles** Angle  $\alpha$  defining hinge1 was measured between the C <sub>$\alpha$</sub>  atoms of residues 4, 359, and 587. Angle  $\beta$  defining hinge2 was measured between the C <sub>$\alpha$</sub>  atoms of residues 291, 303 and 323.

**RMSF** The *gmx RMSF* tool was used to calculate fluctuations of the C <sub>$\alpha$</sub>  atoms around their time-averaged positions. The resulting values (between 0.1 and 4 Angstrom) were projected onto the structure, with red areas denoting flexible residues (>2 Å) and blue areas rigid ones (<2 Å).

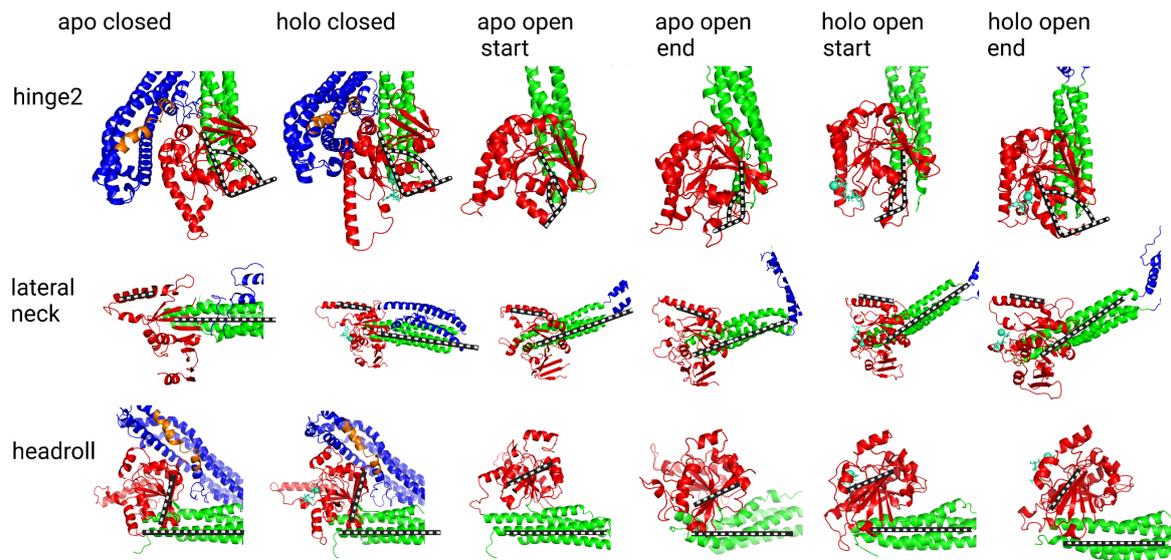
**Spatial distribution** In order to visualize the area sampled by the stalk tip, the *gmx spatial* tool was used to calculate its spatial distribution. First, the trajectory was fitted to the G domain of BDLP, then the position of the C <sub>$\alpha$</sub>  atom of residue 580 was binned at 0.1 nm precision. The resulting distribution was visualized in PyMOL.

**PCA** To observe the largest motions present in the trajectory, principal component analysis of the atomic motions was performed, using the PCA implementation of GROMACS. The trajectory was fitted to the backbone for this analysis. The three main backbone motions were visualized both for the whole protein and the flexible parts of the G domain (flap1, flap2, hinge2, switch2).

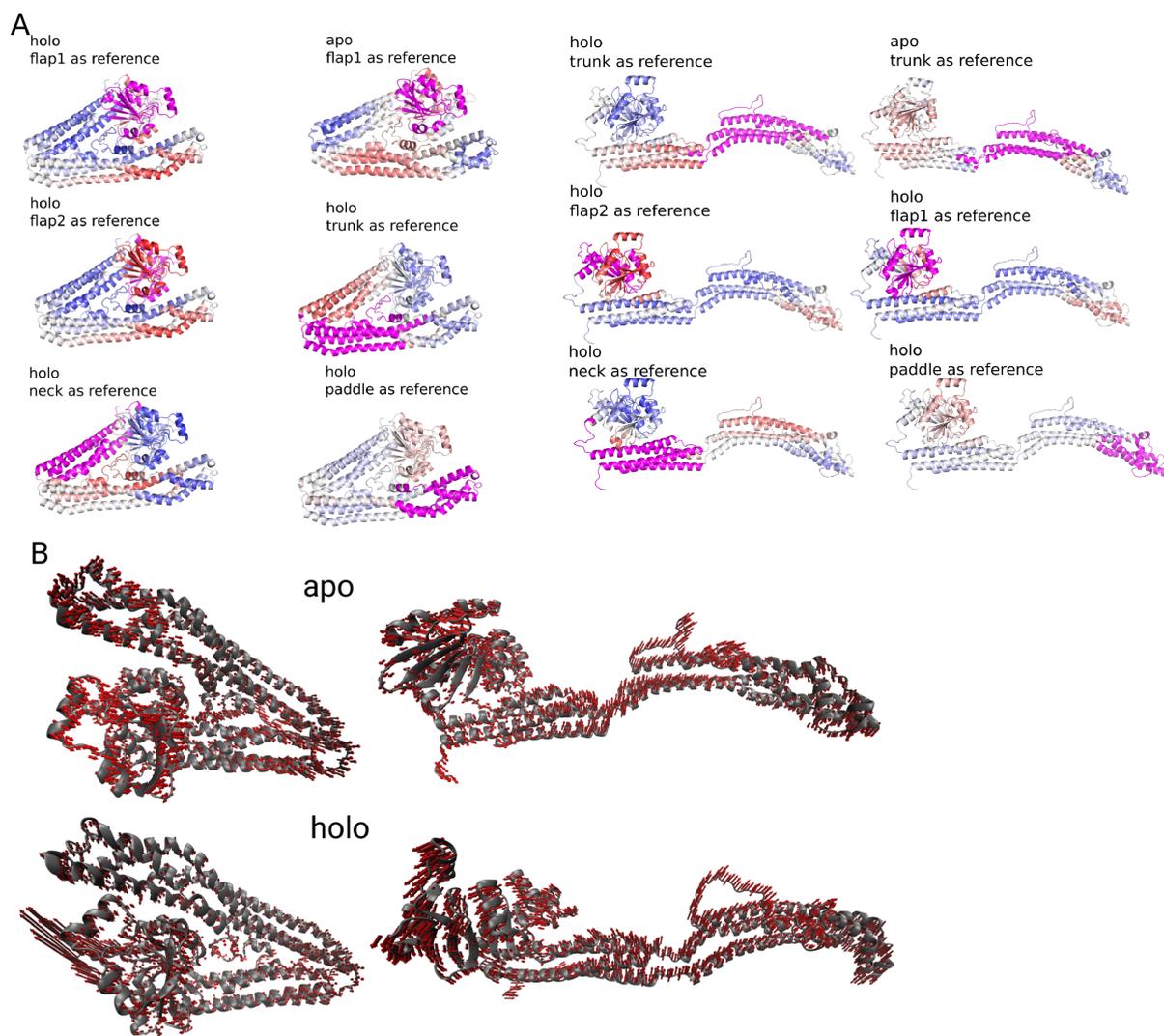
**Cluster analysis** For the clustering of the G domain motifs, the trajectory was first fitted to the  $C_\alpha$  atoms of the G domain with a time step of 60 ps between snapshots. The clustering was then performed only on the motifs (flap1, flap2, switch1, hinge2,  $Mg^{2+}$ , and GTP) with a cutoff 0.25 nm and the *nofit* option.

**Normal mode analysis** The VMD plugin ProDy<sup>25</sup> was used to analyze the normal modes of the initial open and closed BDLP conformations, both in their apo and holo forms. The anisotropic network model was calculated for the  $C_\alpha$  atoms and extended to the backbone, with a cutoff of 19 nm.

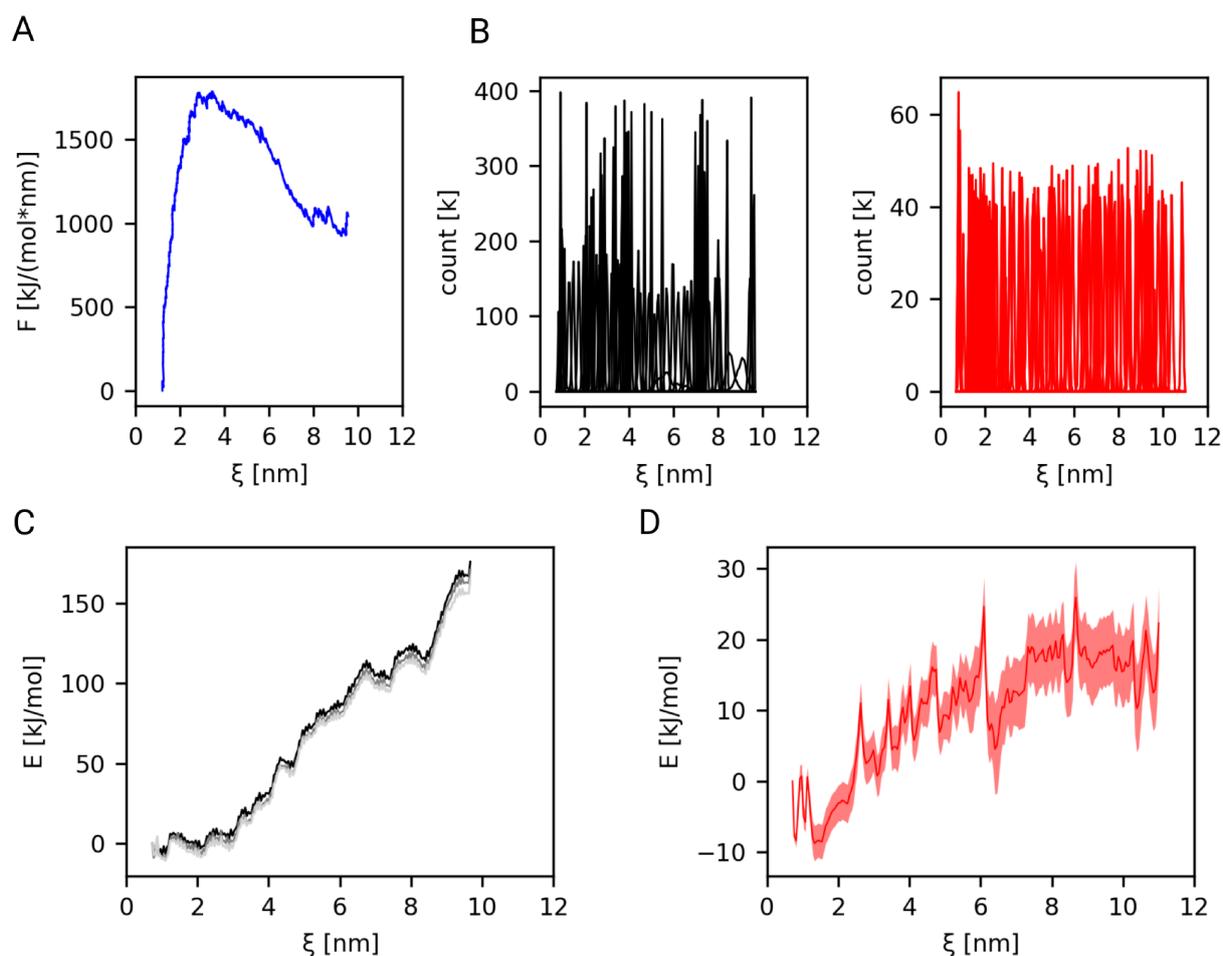
## 2 Supplementary figures



**Figure S1: Snapshots of different hinge2 motions.** Hinge2 is defined by the angle indicated in the top row. A more structured hinge2 region leads to a  $90^\circ$  angle between helix12 and helix13. The lateral motion of the stalk beginning at helix13 becomes apparent in the second row, when using helix10 as a reference. In the third row, we can see how  $\beta$ -strand4 is at a  $90^\circ$  angle to the stalk in the closed form, but more parallel for the open form. These differences could be largely originate from the initial starting structures and are equilibrated after the simulations.



**Figure S2: Normal mode analysis of the BDLP starting conformations.** **A** The first normal mode projected onto the structure as red arrows. **B** The correlation matrix projected onto the structures (blue = anticorrelated motions, red = correlated motion, magenta = reference region). The correlations in the apo state are only shown for one reference region, to illustrate the most striking differences to the holo state.



**Figure S3: Umbrella sampling MD simulations of BDLP.** **A** Force profile during the initial pulling simulation. **B** Coverage of the order parameter  $\xi$  by the windows in the USMD simulations of apo-BDLP (left) and holo-BDLP (right). **C** Convergence test for the USMD simulation of apo-BDLP using increasing simulation lengths: 100 ns (black), 200 ns (grey), 300 ns (light grey). **D** Margin of error for the USMD simulation of holo-BDLP, generated with 1000-fold bootstrapping.

## References

- (1) Low, H. H.; Löwe, J. A bacterial dynamin-like protein. *Nature* **2006**, *444*, 766–769.
- (2) Low, H. H.; Löwe, J. Dynamine architecture — from monomer to polymer. *Current Opinion in Structural Biology* **2010**, *20*, 791–798.
- (3) Park, H.; Lee, G. R.; Heo, L.; Seok, C. Protein loop modeling using a new hybrid energy

- function and its application to modeling in inaccurate structural environments. *PLoS ONE* **2014**, *9*.
- (4) López-Blanco, J. R.; Canosa-Valls, A. J.; Li, Y.; Chacón, P. RCD+: Fast loop modeling server. *Nucleic Acids Research* **2016**, *44*, W395–W400.
  - (5) Karami, Y.; Rey, J.; Postic, G.; Murail, S.; Tufféry, P.; de Vries, S. J. DaReUS-Loop: a web server to model multiple loops in homology models. *Nucleic Acids Research* **2019**, *47*, W423–W428.
  - (6) Fiser, A.; Sali, A. ModLoop: automated modeling of loops in protein structures. *Bioinformatics* **2003**, *19*, 2500–2501.
  - (7) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **2010**, *78*, 1950–1958.
  - (8) Best, R. B.; Hummer, G. Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides. *The journal of physical chemistry. B* **2009**, *113*, 9004.
  - (9) Aliev, A. E.; Kulke, M.; Khaneja, H. S.; Chudasama, V.; Sheppard, T. D.; Lanigan, R. M. Motional timescale predictions by molecular dynamics simulations: Case study using proline and hydroxyproline sidechain dynamics. *Proteins* **2014**, *82*, 195.
  - (10) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **1998**, *79*, 926.
  - (11) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1-2*, 19–25.

- (12) Loschwitz, J.; Steffens, N.; Wang, X.; Schäffler, M.; Pfeffer, K.; Degrandi, D.; Strodel, B. Domain motions, dimerization, and membrane interactions of the murine guanylate binding protein 2. *Sci. Rep.* **2023**, *13*, 679.
- (13) Trott, O.; Olson, A. J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry* **2009**, *31*.
- (14) Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Physical Review A* **1985**, *31*, 1695.
- (15) Nosé, S. A molecular dynamics method for simulations in the canonical ensemble. *Molecular Physics* **1984**, *52*, 255–268.
- (16) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics* **1998**, *52*, 7182.
- (17) T. Darden, D. Y.; Pedersen, L. PARTICLE Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- (18) Essmann, U.; Perera, L.; Berkowitz, M. L. A Smooth PARTICLE Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.
- (19) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463–1472.
- (20) Feenstra, K. A.; Hess, B.; Berendsen, H. J. C. Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. *Journal of Computational Chemistry* **1999**, *20*, 786–798.
- (21) Barz, B.; Loschwitz, J.; Strodel, B. Large-scale, dynamin-like motions of the human guanylate binding protein 1 revealed by multi-resolution simulations. *PLOS Computational Biology* **2019**, *15*, e1007193.

- (22) Bussi, G. Hamiltonian replica exchange in GROMACS: a flexible implementation. *Mol. Phys.* **2014**, *112*, 379–384.
- (23) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New feathers for an old bird. *Comp. Phys. Commun.* **2014**, *185*, 604–613.
- (24) Hub, J. S.; de Groot, B. L.; van der Spoel, D. g wham - A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates. *Journal of Chemical Theory and Computation* **2010**, *6*, 3713–3720.
- (25) Bakan, A.; Meireles, L. M.; Bahar, I. ProDy: Protein Dynamics Inferred from Theory and Experiments. *Bioinformatics* **2011**, *27*, 1575–1577.



Publication 2

**B**

# Integrative modeling of guanylate binding protein dimers

Wibke Schumann,<sup>†,‡,△</sup> Jennifer Loschwitz,<sup>¶,‡,△</sup> Jens Reiners,<sup>§</sup> Daniel Degrandi,<sup>||</sup>  
Klaus Pfeffer,<sup>||</sup> Kai Stühler,<sup>⊥,#</sup> Gereon Poschmann,<sup>⊥</sup> Sander H.J. Smits,<sup>@,§</sup> and  
Birgit Strodel\*,<sup>¶,‡</sup>

<sup>†</sup>*Institute of Theoretical and Computational Chemistry, Heinrich Heine University Düsseldorf,  
40225 Düsseldorf, Germany*

<sup>‡</sup>*Institute of Biological Information Processing: Structural Biochemistry, Forschungszentrum Jülich,  
52428 Jülich, Germany*

<sup>¶</sup>*Institute of Theoretical and Computational Chemistry, Heinrich Heine University Düsseldorf,  
40225 Düsseldorf, Germany*

<sup>§</sup>*Center for Structural Studies Heinrich-Heine-Universität Düsseldorf, 40225 Düsseldorf, Germany*

<sup>||</sup>*Institute of Medical Microbiology and Hospital Hygiene, Heinrich Heine University, Duesseldorf,  
Germany*

<sup>⊥</sup>*Institute of Molecular Medicine, Proteome Research, Medical Faculty and University Hospital  
Düsseldorf, Heinrich-Heine University Düsseldorf, Universitätsstraße 1, 40225 Düsseldorf, Germany*

<sup>#</sup>*Molecular Proteomics Laboratory, Biomedical Research Centre (BMFZ), Heinrich-Heine  
University Düsseldorf, Universitätsstraße 1, 40225 Düsseldorf, Germany*

<sup>@</sup>*Institute für Biochemie, Heinrich-Heine-Universität Düsseldorf, 40225 Düsseldorf, Germany*

<sup>△</sup>*Contributed equally to this work.*

E-mail: b.strodel@fz-juelich.de

## Abstract

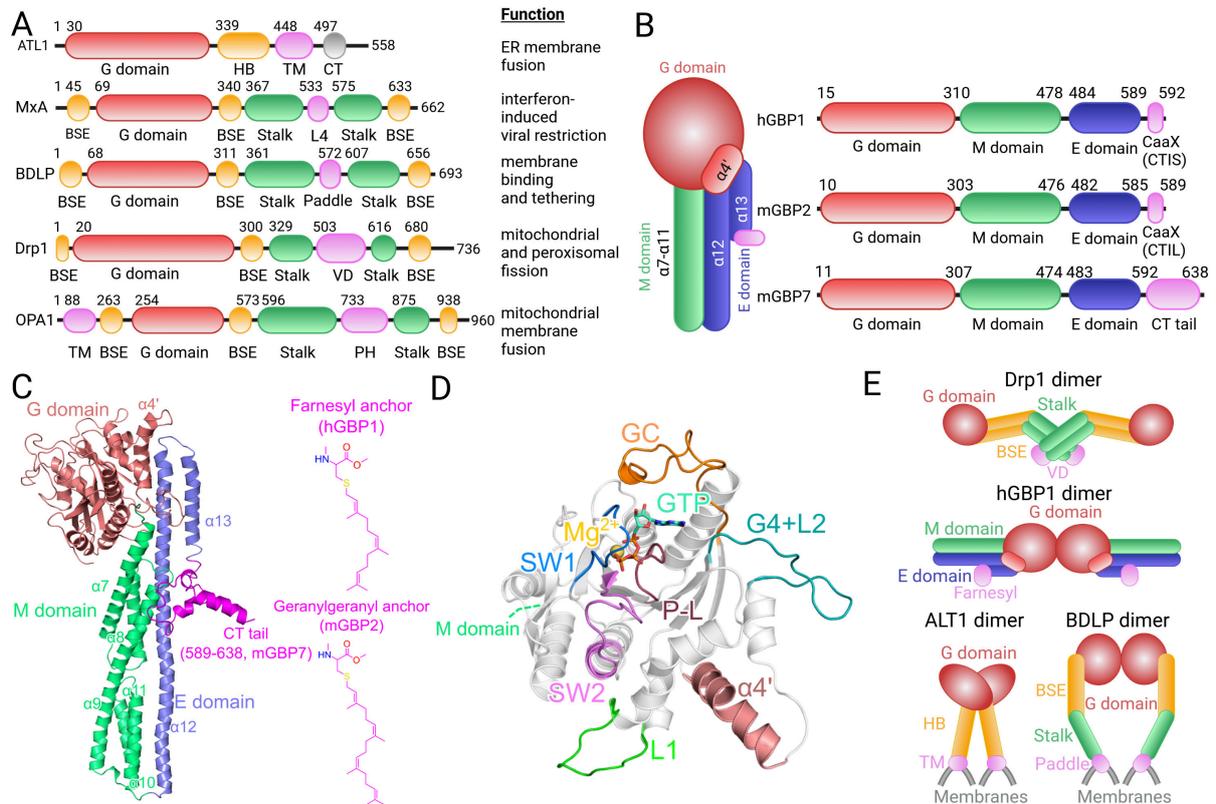
Guanylate binding proteins (GBPs) are interferon- $\gamma$ -activated large GTPases, effective against intracellular pathogens like *Toxoplasma gondii*. Their host-protective functions require oligomerization, however, the oligomer structures have not been completely resolved yet. Here, we provide dimer models for hGBP1 and the murine GBPs 2 and 7 (mGBP2 and mGBP7) based on integrative modeling that involves the crystal structure of the G domain dimer of hGBP1, cross-linking mass spectrometry (XL-MS), small angle X-ray scattering (SAXS), protein-protein docking, and molecular dynamics (MD) simulations of hGBP1, mGBP2 and mGBP7. We first compare the sequences and protein dynamics of the monomeric hGBP1, mGBP2, and mGBP7, finding that the M/E domain of all three proteins is highly mobile featuring a hinge movement, yet this motion is less pronounced in mGBP7 while its GTPase (G) domain is more flexible. These differences can be explained by the variations in the sequences between mGBP7 and hGBP1/mGBP2 and extend to their dimers. While hGBP1 and its close orthologue mGBP2 dimerize via their G domains, mGBP7 shows a variety of possible dimer structures, among them parallel and crossed-stalk conformations. The G domain is only partly involved in mGBP7 dimerization, which provides a rational why mGBP7, unlike hGBP1 and mGBP2, can dimerize in the absence of GTP. The different GBP dimer structures, which still exhibit hinge movements to certain degrees, are expected to encode diverging functions, such as a destabilization of pathogenic membranes or fusion of the parasitophorous vacuole membrane with the autophagic machinery.

**Keywords:** guanylate binding proteins, protein-protein docking, MD simulation, small angle X-ray scattering, crosslinking mass spectrometry

# 1 Introduction

Dynamin superfamily proteins (DSPs) are mechanochemical enzymes, converting GTPase activity into some of the highest torques observed for proteins, which is then used for membrane fusion or fission.<sup>1,2</sup> In contrast to the smaller Rat sarcoma virus (Ras)-like GTPases, DSPs exhibit lower substrate affinity but higher basal hydrolysis rates, which is highly stimulated upon oligomerization.<sup>3</sup> Furthermore, no accessory proteins, such as GTPase-activating proteins or guanine nucleotide exchange-factors are required since these functions are already encoded in the corresponding GTPase effector domain.<sup>3,4</sup> The oligomers of DSPs form ordered rings, lattices or helices, thereby tubulating membranes,<sup>2</sup> and are thus involved in critical cellular functions, such as endocytosis, mitochondrial membrane tubulation, cell division, and vesiculation.<sup>5,6</sup>

Since some of the DSPs have been studied separately, the nomenclature of domains is not unified, although they are functionally homologous and feature a modular composition of domains. Some examples are depicted in Fig. 1A. The GTPase domain (G domain) is the only conserved domain at the sequence level, and it shows the typical motifs of a nucleotide-binding domain (G1–G5, switch1 (SW1), switch2 (SW2)), in total spanning about 300 continuous residues. GTPase activity, although stimulated by multimerization, is usually not dependent upon it.<sup>7,8</sup> In the case of dynamin, an intramolecular interaction can replace intermolecular complementation of the G domains.<sup>2,9</sup> Other domains of DSPs are structurally conserved, even if they are not homologous in sequence. The bundle signaling element (BSE or neck or helical bundle 1) is an elongated helical domain that is in contact with the G domain via the so-called hinge 2 and connects to the stalk (or trunk or middle/effector-domain or helical bundle 2) domain via hinge 1. The membrane interaction can be mediated by a pleckstrin homology (PH) domain, the L4 loop, a variable domain (VD), a paddle or a C-terminal region (Fig. 1A and B).<sup>2</sup>



**Figure 1: Structural organization and dimerization modes of DSPs and GBPs.** (A) Schematic domain organization of selected DSPs: atlastin 1 (ATL1), myxovirus resistance protein 1 (MxA), bacterial dynamin like protein (BDLP), dynamin-related protein 1 (Drp1), OPA1 (from optic atrophy gene 1).<sup>2,6,10</sup> DSPs exhibit a highly conserved large GTPase domain (G domain, red), a structural element or domain for membrane binding (violet) which can be a transmembrane (TM) domain, the L4 loop of the G domain, a paddle domain, a variable (VD) domain, or a pleckstrin homology (PH) domain, two elongated  $\alpha$ -helical bundle domains called stalk (green), and a bundle signaling element/helical bundle (BSE/HB, orange). The function of each DSPs is given on the right. (B) Schematic domain organization of GBPs. They are split into three domains: the large GTPase (G) domain (red), the middle (M) domain (green), and the effector (E) domain (blue), followed by a region responsible for membrane binding. It is either a CaaX box for isoprenylation or an elongated C-terminal (CT) tail. (C) The structure of mGBP7 is shown as cartoon. The helices discussed in this work ( $\alpha 4'$ ,  $\alpha 7$ – $13$ ) are indicated. At the C-terminus, mGBP7 features a helical CT tail, while hGBP1 and mGBP2 become isoprenylated with either a farnesyl (hGBP1) or a geranylgeranyl (mGBP2) group (2D structures on the right). (D) Close-up of the G domain of mGBP7. The four conserved GTP-binding site motifs and other important structural elements are highlighted: P-L or G1 (dark red), SW1 or G2 (blue), SW2 or G3 (magenta), G4 plus loop L2 (G4+L2, turquoise), loop L1 (green), guanine cap (GC, red), and  $\alpha 4'$  (red). The bound nucleotide GTP is shown with sticks and colored by atom type and Mg<sup>2+</sup> is shown as a yellow sphere. (E) Known DSP dimerization modes are shown for Drp1 (PDB ID 4BEJ),<sup>11</sup> hGBP1 (PDB ID 2BC9)<sup>3</sup> featuring the G–G domain dimer,<sup>2</sup> ATL1 (PDB ID 3Q5D),<sup>12</sup> and BDLP (PDB ID 2W6D).<sup>13</sup> In general, the dimerization occurs via the stalk regions (Drp1) or via the G domains (ATL1, hGBP1, BDLP). Panels A, B, and E were created with BioRender.

One subfamily of DSPs is the guanylate binding protein family (GBPs, Fig. 1B), which are known, among others, for their anti-*Toxoplasma gondii*,<sup>14-17</sup> anti-HIV,<sup>18</sup> and anti-*Chlamydia trachomatis*<sup>19,20</sup> activity. Using bioinformatic/phylogenetic methods, seven human GBPs (hGBP1 to hGBP7) and 11 murine GBPs (mGBP1 to mGBP11) have been identified.<sup>15,21</sup> In this work, we will concentrate on hGBP1, mGBP2, and mGBP7, continuing our previous studies of hGBP1,<sup>22</sup> mGBP2,<sup>23</sup> and mGBP7.<sup>24</sup> The GBPs exhibit three domains (Fig. 1): (1) a GTPase domain (G domain) for GTP binding and hydrolysis (Fig. 1D), (2) a middle domain (M domain) for regulation, and (3) a GTP effector domain (E domain) for interactions with lipid membranes. The M and E domain (called M/E domain herein) give the GBPs an elongated shape. The E domain can contain a CaaX motif at the C-terminus, which is the case for hGBP1 and mGBP2, resulting in an isoprenyl lipid anchor after post-translational modification, while in the case of mGBP7 the protein harbors about 50 more residues at the C-terminus, which were shown to be essential for membrane binding of the protein (Fig. 1B and C).<sup>24</sup> mGBP2 is the murine orthologue of the well-studied hGBP1. It has been shown that mGBP2 binds to the membrane of the *T. gondii* parasitophorous vacuole before mGBP7; however, the deletion of mGBP7 – compared to mGBP2 – results in a higher lethality of mice infected by the parasite.<sup>16,17</sup> The membrane of the parasitophorous vacuole originates from lipids of the host cell and the parasite hides within the vacuole to escape from the host cell defense response. GBPs cycle between monomeric, dimeric, and polymeric states, with the polymer being formed by thousands of subunits on the parasitophorous vacuole membrane of *T. gondii*.<sup>25</sup>

To achieve oligomerization, DSPs use up to four interfaces in the helical (stalk) region, and only once a full helical turn around the invaginated membrane is completed, the G domains come into contact with the membrane („grip“). The substrate GTP, which has previously been loaded, is now processed to GDP, which leads to a conformational change („pull“), where the dynamin filaments slide against each other, constricting the membrane. GDP is weaker in promoting oligomerization, possibly leading to dissociation once the cycle is

complete.<sup>26–28</sup> While nucleotides are necessary for oligomerization of most DSPs, mGBP7 can form dimers in the absence of GTP, whereas mGBP2 dimerization is mostly dependent on nucleotide binding.<sup>24</sup> Nucleotide-independent oligomerization has also been found for OPA1 (from the optic atrophy gene 1, Fig. 1A).<sup>10</sup> Moreover, GBPs are able to hydrolyze GDP further to GMP, which suggests that they remain in an oligomeric state even after hydrolysis. The second hydrolysis takes place either in a tetrameric state, or in a drastically extended conformation of the dimer.<sup>29,30</sup> According to Sistemich et al.,<sup>31</sup> binding of GTP to hGBP1 leads to solvent exposure of the farnesyl anchor, then dimerization via the G domains takes place, and in the transition state, the helix  $\alpha$ 12 folds out and/or further oligomerization occurs. When GTP is scarce, these polymers dissociate, indicating that GMP cannot stabilize the polymer.<sup>31</sup> Oligomers are also more strongly associated with membranes, due to the high local concentration of membrane-binding motifs.<sup>32</sup> Several studies suggest functional heterodimerization as a key concept for DSPs,<sup>33–35</sup> especially proteins that are localized on the same chromosome and are thus transcribed together tend to associate.<sup>25,36</sup> With regards to GBPs, this specifically applies to mGBP1/mGBP2 and mGBP3/mGBP7.

Membrane fission is likely to require several continuous cycles of constriction, until the membrane tubule thickness is thin enough for the two eventually separating bilayers to touch each other. Examples of DSPs enabling fission are dynamin, dynamin-related protein 1 (Drp1), and Dnm1p, a yeast protein homologous to Drp1. Fusion, on the other hand, requires the association of at least two proteins to opposing membranes (tethering) and subsequent conformational change, putting them in close spatial proximity to initiate the fusion process.<sup>27</sup> Fusion DSPs, such as atlastin, Sey1p, mitofusins, OPA1, MGM1p, and LeoA/B/C, have permanent membrane binding ability, function as dimers, and have a fused BSE/stalk domain.<sup>27</sup> Intriguingly, some proteins can serve both functions, i.e., fission and fusion, such as Vps1p and NosBDLP.<sup>37</sup> It is likely that different dimerization motifs could confer different functions. Dimer structures of DSPs that are deposited in the Protein Data Bank (PDB) display G domain dimers (hGBP1, Irga6, dynamin), crossed-stalk dimers (dynamin), parallel

(BDLP, Sey1p) and antiparallel (SX9) stalk dimers (Fig. 1E), and higher-order oligomers that combine several of these interfaces.

Whether GBPs actively destabilize pathogenic membranes, including parasitophorous vacuole membranes, or fuse them within the autophagic machinery, or both, is still unknown. However, information about their dimerization patterns and dynamics could yield important hints. Thus far, only two partial dimer structures of GBPs are available.<sup>3,38</sup> The structure by Ghosh et al.<sup>3</sup> is a dimer only formed by the G domains of hGBP1, while in the structure reported by Cui et al. two C-terminally truncated human hGBP5 molecules, comprising the G and M domains, interact lengthwise, with a cross formed at the base of the G domains. This dimerization mode is similar to the one known for atlastin, which cycles between this and the G domain dimer structure as seen for hGBP1.<sup>39,40</sup> Based on molecular dynamics (MD) simulations, we have previously unraveled a large-scale hinge motion taking place in monomeric and dimeric hGBP1<sup>22</sup> and mGBP2,<sup>23</sup> which is also present in the membrane-bound protein form. For the dimer structures, we assumed the presence of the G domain dimer as reported for hGBP1.<sup>3</sup> Here, we test for the existence of alternative dimer structures of hGBP1, mGBP2, and mGBP7 by applying a combination of protein-protein docking, small-angle X-ray scattering (SAXS), and cross-linking mass spectrometry (XL-MS). The stability of the resulting dimer models is analyzed using MD simulations, which at the same time allows us to contrast the dynamics in the GBP monomers and dimers as well as to pinpoint the important residues in the dimerization interfaces. In this study, we provide evidence, that mGBP7 prefers different dimerization modes as compared to hGBP1 and mGBP2, which ties in with some differences in their sequences and dynamics that are reported here too.

## 2 Results

### 2.1 Sequence, structural and biochemical comparison of hGBP1, mGBP2 and mGBP7

First, we compare the protein sequences, molecular structures and biochemical properties of hGBP1, mGBP2, and mGBP7. Their most relevant properties are summarized in Table 1.

**Membrane binding and protein assembly** All three GBPs are composed of a G, M, and E domain,<sup>41</sup> with the M and E domain sometimes grouped together as an M/E domain (Fig. 1B and C). mGBP2 is a close orthologue of hGBP1 with 66.2% sequence identity and both feature a CaaX motif (residues 589–592) leading to isoprenylation. The mGBP7, on the other hand, contains 49 additional residues at the C-terminus (590–638, called CT tail) compared to mGBP2, which mediate the protein binding to a membrane (Fig. 1C).<sup>24,36</sup> Residues 598–620 were predicted as a transmembrane (TM) helical region using TMpred<sup>42</sup> and mutation experiments indeed confirmed the CT tail to be essential for membrane binding of mGBP7.<sup>24</sup> Thus, all three proteins can bind to lipid membranes. Moreover, in the cytosol, they form vesicle-like structures (VLS), which can also result from hetero-assembly of mGBP2 co-localizing with mGBP1/3 or mGBP7 with mGBP3.<sup>17,25,31</sup> However, mGBP2 and mGBP7 do not co-localize in VLS.<sup>25</sup> For hGBP1 it has been shown that in the presence of GTP and the farnesyl lipid anchor it polymerizes with the E domain being folded out, leading to the formation of hGBP1 disks that assemble into tubes and *in vitro*, they can tether membranes.<sup>43–45</sup> Binding of GTP to hGBP1 shifts the monomer–dimer equilibrium in favor of the dimer, which enhances the GTP hydrolysis.<sup>45</sup> However, this was suggested to loosen the contacts between the G domains defining the dimer, provoking E domain contacts to be formed, which is followed by GDP hydrolysis and dimer dissociation.<sup>30,32,46</sup>

**Table 1: Summary of properties of hGBP1, mGBP2, and mGBP7.** Abbreviations: GBP – guanylate binding protein; VLS – vesicle-like structure; h – human; m – murine,  $K_D$  – dissociation constant; CT tail – additional C-terminal residues; 3-letter code for amino acids.

Property	hGBP1	mGBP2	mGBP7
Sequence length (UniProt ID)	592 (P32455)	589 (Q9Z0E6)	638 (Q91Z40)
Sequence identity [%]	mGBP2: 66.2 mGBP7: 50.2	hGBP1: 66.2 mGBP7: 46.6	hGBP1: 50.2 mGBP2: 46.6
Sequence similarity [%]	mGBP2: 76.3 mGBP7: 63.0	hGBP1: 76.3 mGBP7: 58.8	hGBP1: 63.0 mGBP2: 58.8
Lipid anchor	yes (CTIS; farnesyl)	yes (CTIL; geranylgeranyl)	no (elongated CT tail)
GTP binding for dimerization	avored	yes	no
GTP → GDP → GMP	yes	yes	yes
GTP affinity ( $K_D$ in $\mu\text{M}$ )	1.1 <sup>47</sup>	0.45 <sup>48</sup>	0.22 <sup>24</sup>
GTP binding residues (G4)	TLRD	TLRD	TVRD
G domain stabilization by GTP	yes <sup>22</sup>	yes <sup>23</sup>	no (in this paper)
Cytosol organization	VLS	VLS	VLS
Colocalization with other GBPs	not known	mGBP1/3 <sup>25</sup>	mGBP3 <sup>17</sup>

Abbreviations: GBP – guanylate binding protein; VLS – vesicle-like structure; h – human; m – murine,  $K_D$  – dissociation constant; CT tail – additional C-terminal residues; 3-letter code for amino acids.

**GTP binding and hydrolysis** The G domain contains the four G motifs (G1–G4) needed for the GTP hydrolysis, the guanine cap, the loop 1 (L1), and the  $\alpha 4'$  helix involved in membrane interactions (Fig. 1D). All three GBPs can hydrolyze GTP to GMP in a two-step mechanism, which is a big difference to other DSPs where the hydrolysis stops at GDP.<sup>2,49,50</sup> However, none of the GBPs is able to directly bind GDP, i.e., it can only be processed as an intermediate following GTP hydrolysis.<sup>24,48,49</sup> The GTP binding affinity is slightly different between the proteins, with mGBP7 having the highest and hGBP1 the lowest affinity (Table 1). To understand the differences in the GTP binding, we aligned the sequences of

hGBP1, mGBP2, and mGBP7 using T-coffee<sup>51,52</sup> and analyzed their amino acid compositions in terms of physicochemical properties (Fig. S1). This extends to a representation of their electrostatic potential surface (EPS) for the whole protein and the GTP binding site in particular (Fig. S2). All three GBPs are rather negatively charged on one side of the protein along the M/E domain, yet feature a mix of positive and negative charges on the opposite side, yet with a higher share of positive charges in the G domain on that protein side. Of the three proteins, mGBP7 is the most negative on either side, followed by mGBP2. However, the GTP binding region of mGBP2 is more positive than in hGBP1, which might explain mGBP2's higher affinity for GTP.<sup>47,48</sup> Though this reasoning should be taken with caution, as the GTP affinity of mGBP7 is even higher, yet its GTP binding region is not more positive but contains mixed charges.

The G1 motif, also called phosphate-binding or P-loop (P-L), is highly conserved and is key for GTP binding and hydrolysis. The two most important residues are R48 serving as arginine finger and K51 being responsible for GTP hydrolysis. The K51A mutant drastically impairs GTP binding: in hGBP1 by ~50-fold ( $K_D = 53 \mu\text{M}$ ),<sup>47</sup> in mGBP2 by ~100-fold ( $K_D = 44.1 \mu\text{M}$ ),<sup>48</sup> and mGBP7 lost its GTP binding ability altogether.<sup>24</sup> Moreover, only negligible GTP hydrolysis and no dimerization takes place for the K51A mutant of hGBP1 and mGBP2, while the mGBP7 mutant can still dimerize as it does not require GTP for this. The R48A mutant of hGBP1 showed a slightly increased GTP binding affinity ( $K_D = 0.34 \mu\text{M}$ ) but a decreased GTP hydrolysis, and dimerization still takes place.<sup>48</sup> Kravets et al. further revealed in 2016 that the R48A mutant has a decreased capacity for recruitment to the parasitophorous vacuole membrane, while the K51A mutant lacks this ability completely, which is accompanied by its inability to polymerize/localize in a VLS and to control *T. gondii* growth at all.<sup>25</sup> For hGBP1, Zhu et al. demonstrated in 2013 that the antiviral function against influenza virus A is reduced by the K51A mutant.<sup>53</sup> These observations underscore the importance of the G1 motif for the function of these GBPs.

The sequences of the next two G motifs, G2 and G3 (also called switches SW1 and SW2),

feature like G1 only minor changes among the three proteins. Only for the surrounding residues of SW1 we identified changes, such as Q72 and H74 in hGBP1 and mGBP2 became R72 and E74 in mGBP7. In hGBP1, the mutants Q72A, H74A, T75A and E99A led to a decreased GTP binding and hydrolysis with reduced GMP production.<sup>47</sup> Another function of H74 was recently identified, as it was found to play a critical role as a distant molecular switch for the E domain releasing from the G domain of hGBP1.<sup>44</sup> The SW2 of mGBP7 harbors more hydrophobic and also more positively charged residues, while in mGBP2 these are more polar and negatively charged than in hGBP1 and mGBP7. However, the key residues T75 of SW1 needed for interactions with Mg<sup>2+</sup> as well as S73 of SW1 and E99 of SW2 required for coordinating the nucleophilic water molecule during the hydrolysis reaction are the same.<sup>3,47</sup> For mGBP2, only data of an E99A mutant is known, which shows decreased GTP hydrolysis, a low multimerization capability, and a reduced binding to parasitophorous vacuole membranes.<sup>25,48</sup> For mGBP7, there are no such mutation data available yet.

The G4 motif, which is responsible for mediating nucleotide specificity together with the loop 2 (G4+L2), is more positively charged in mGBP7 than in the other two GBPs. The first residues of that motif are not very different among the three proteins. An interesting change occurred at position 182 and 184 in mGBP2/7 and hGBP1, respectively, where a valine is found in mGBP7 as compared to a leucine in hGBP1 and mGBP2. The mutations D182N and D184N in hGBP1 and mGBP2, respectively, resulted in a 20 fold lower affinity for GTP and led to a reduction in the GTP hydrolysis rate, cooperativity, and GMP formation.<sup>25,47</sup> Moreover, the D182N mutant of mGBP2 might have a similar effect on multimerization and controlling *T. gondii* growth as the K51A mutant, which is dysfunctional with regard to GTP hydrolysis.<sup>25,48</sup> The position 182 is therefore an interesting candidate for future mutational studies of mGBP7 to assess its involvement in GTP binding and hydrolysis.

**G domain loops** The most important element for dimerization of hGBP1 is the guanine cap, which is an unstructured loop region.<sup>3,22</sup> In hGBP1, two arginine residues, R240 and

R244 exhibit a key role for the dimerization.<sup>54</sup> Interestingly, mGBP7 has oppositely charged residues in that region. Examples are R244 in hGBP1 and R242 in mGBP2 that converted to D242 in mGBP7, and K246 in hGBP1 and Q244 in mGBP2 became E244 in mGBP7 (Fig. S1). As a result, the overall charge of the guanine cap in mGBP7 is  $-3$  while it is neutral in hGBP1 and mGBP2. The other larger loop region of the G domain is the loop 1 (L1), which in mGBP2 is more hydrophobic and polar but less charged than in hGBP1 and mGBP7. It can therefore interact with the geranylgeranyl lipid anchor and also the helix  $\alpha 12$  of the E domain.<sup>23</sup> The  $\alpha 4'$  helix, which belongs to the G domain, also interacts with the  $\alpha 12$  as well as with  $\alpha 13$  via salt bridges, which keeps the E domain close to the G domain.<sup>32</sup> For hGBP1 it was demonstrated that this "lock" can be broken by GTP hydrolysis, releasing the whole E domain from the G domain in order for hGBP1 to polymerize and tether membranes.<sup>43,44</sup> Interestingly, mutating the hGBP1 residues F171 and F175 in L1 and F229 in  $\alpha 4'$  to alanine leads to a higher rate in GDP hydrolysis, even though these residues are quite far away from the active site.<sup>44</sup> The same is true for the E227A/K228A mutant, which disrupts the intramolecular interactions between  $\alpha 4'$  and  $\alpha 12/13$ , enabling release of the E domain from the G domain. Not much is known yet about the biochemical importance of the same residues in mGBP2 and mGBP7, yet their involvement in the protein dynamics and dimerization will be investigated within this study.

**M/E domain** All three proteins harbor a considerable amount of charged residues in the M/E domain. A major difference is that the M domain of mGBP7 has an overall charge of  $+10$ , whereas that charge is  $-2$  and  $0$  in hGBP1 and mGBP2, respectively (Fig. S1B). This might be of relevance for the membrane binding of mGBP7, especially since these positive charges are preferentially found on one side of the protein (Fig. S2A) and are augmented by an overall positive charge of  $+3$  of the CT tail, which was already demonstrated to be essential for membrane interaction.<sup>24</sup> Another common feature of the three protein sequences is that positively or negatively charged residues are repeated twice or more often and the

opposite charges tend to cluster together. Examples are (i) triply repeated residues of the same charge, like <sup>417/419</sup>EEE<sup>419/421</sup> in mGBP2/hGBP1 and <sup>439</sup>RKK<sup>441</sup> in mGBP7; (ii) mixture of positively and negatively charged residues, often in repeats, like <sup>389</sup>EKKRDD<sup>394</sup> in hGBP1, <sup>389</sup>KRD<sup>391</sup> in mGBP2, and <sup>388</sup>EEKRED<sup>393</sup> in mGBP7. The second variant occurs more often in all three GBPs. In mGBP7, however, there are more breaks by polar and hydrophobic residues between the charged stretches, which are therefore more distributed along the helices. The helices  $\alpha$ 12/13 are not amphipathic in either of the three GBPs, as the charged residues are not only solvent-exposed, some of them are also buried (Fig. S2B). For hGBP1 we showed that some of the charged residues of  $\alpha$ 12 form salt bridges with the M domain that are crucial for the stability of that long helix.<sup>22</sup> The  $\alpha$ 13 helices of hGBP1 and mGBP7 feature clusters of positive charges, <sup>582</sup>KMRRRK<sup>587</sup> and <sup>567</sup>KRK<sup>569</sup>, respectively. Alanine-scanning mutagenesis of the three arginines in  $\alpha$ 13 helix of hGBP1 ablated the farnesylated protein's ability to bind to bacterial outer membranes.<sup>55</sup> The mGBP2 does not have repeated lysine or arginine residues in  $\alpha$ 13 directly before the CaaX motif, only a single lysine residue (K585), which one can expect to help in membrane binding together with the longer geranylgeranyl lipid anchor compared to the farnesyl group in hGBP1. The sequence <sup>567</sup>KRK<sup>569</sup> in mGBP7 is directly before the CT tail, which was demonstrated to be essential for membrane binding.<sup>24</sup> The CT tail has a +3 charge, coming from six positively and three negatively charged residues. Of the other 40 CT tail residues, there are 27 aromatic/hydrophobic ones. This combination of charges and hydrophobicity allows the CT tail to take over the role of a membrane anchor, replacing the function of the isoprenyl group of hGBP1 and mGBP2.

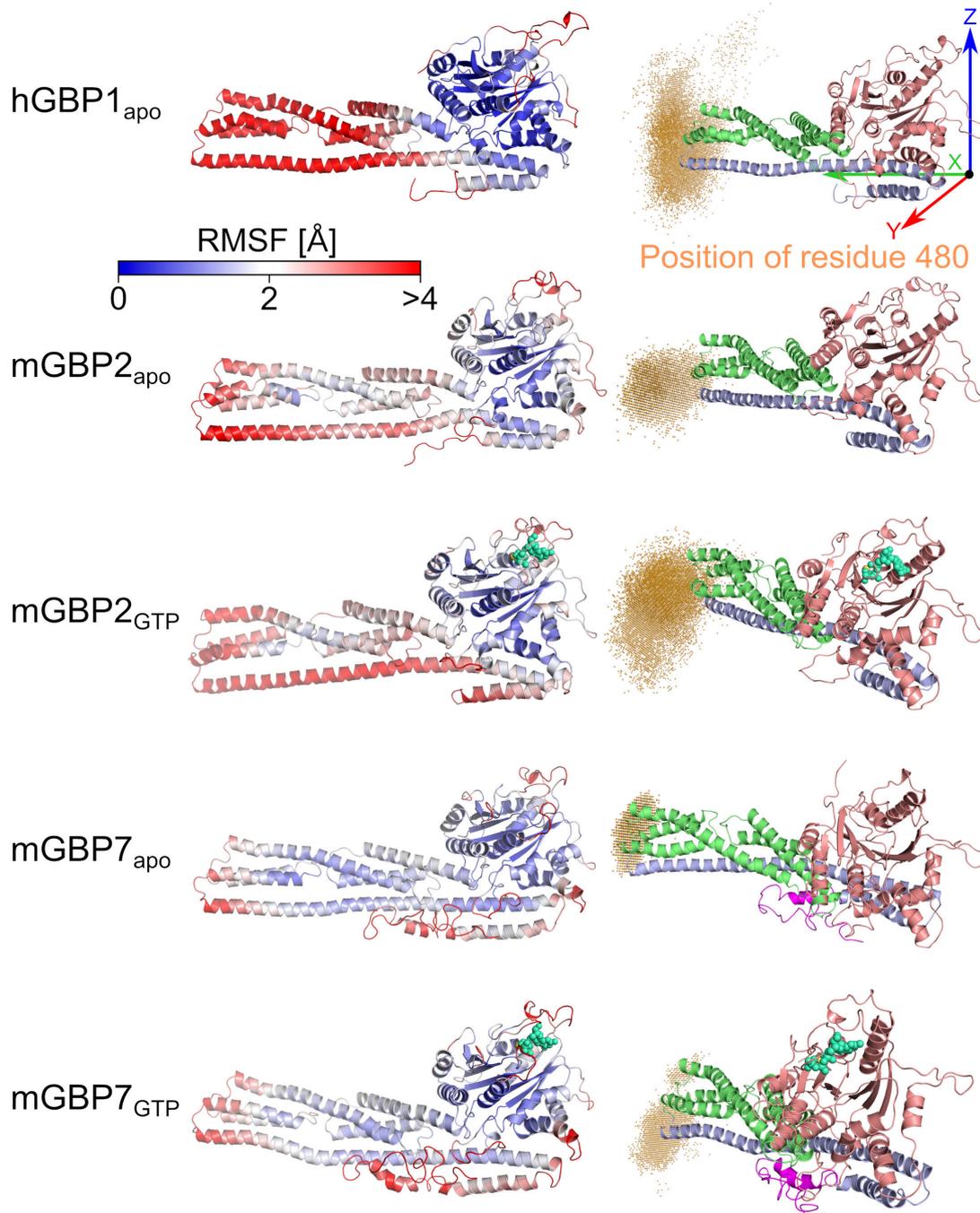
## 2.2 Molecular dynamics of monomeric GBPs

Before establishing dimer models of the three GBPs under investigation, we compare the dynamics of their monomers, on the one hand to further elucidate possible differences between hGBP1, mGBP2, and mGBP7, and on the other hand to have the monomer dynamics as a reference for the dynamics of the proteins when being oligomerized. The dynamics of

hGBP1 and mGBP2 was discussed in detail in our previous works,<sup>22,23</sup> while the dynamics of mGBP7 is being unraveled in this study. In all three cases, Hamiltonian replica exchange molecular dynamics (HREMD) simulations were applied to the monomeric proteins. All three proteins were simulated in their apo-state (denoted hGBP1<sub>apo</sub>, mGBP2<sub>apo</sub>, and mGBP7<sub>apo</sub> here) and for mGBP2 and mGBP7 we also considered the GTP-bound state (mGBP2<sub>GTP</sub>, mGBP7<sub>GTP</sub>). For hGBP1 and mGBP7 we used  $30 \times 400$  ns MD replicas and for mGBP2 it was  $16 \times 200$  ns MD replicas (see Table 2 for a list of all simulations). For mGBP2 we had shown that the smaller HREMD setup was sufficient as no further protein states were sampled in a  $40 \times 400$  ns HREMD simulation.<sup>23</sup> For the analysis, we considered the protein conformations collected in the HREMD target replica that had no modifications applied to the potential energy function of either hGBP1, mGBP2, or mGBP7.

**Overall protein flexibility** The flexibility of the protein monomers was first quantified by the root mean square fluctuations (RMSF) of the  $C_\alpha$  atoms (Fig. 2 and S3). Among the three proteins, hGBP1<sub>apo</sub> is the most flexible with RMSF values of up to 15.8 Å in the M/E domain, followed by mGBP2 with maximal RMSF values of 5.1 Å at residue 481, which is at the tip of the M/E domain, in the apo-state and 6.4 Å at residue 422 of  $\alpha 10$  in the GTP-bound state, while mGBP7 is the least flexible with maximal RMSF values at residue 588 of  $\alpha 13$ , reaching 4.1 Å in mGBP7<sub>apo</sub> and 4.9 Å in mGBP7<sub>GTP</sub>. The different protein flexibilities are also visible in Figure 2 where the protein structures are colored according to the RMSF values.

**G domain motions** In mGBP2, binding of GTP stabilizes the G motifs and loops of the G domain, especially also the guanine cap, which is not so much the case for mGBP7. Except for the P-L motif, these motifs and loops in mGBP7<sub>GTP</sub> remain flexible with RMSF values above 2 Å (Figs. 2 and S3).



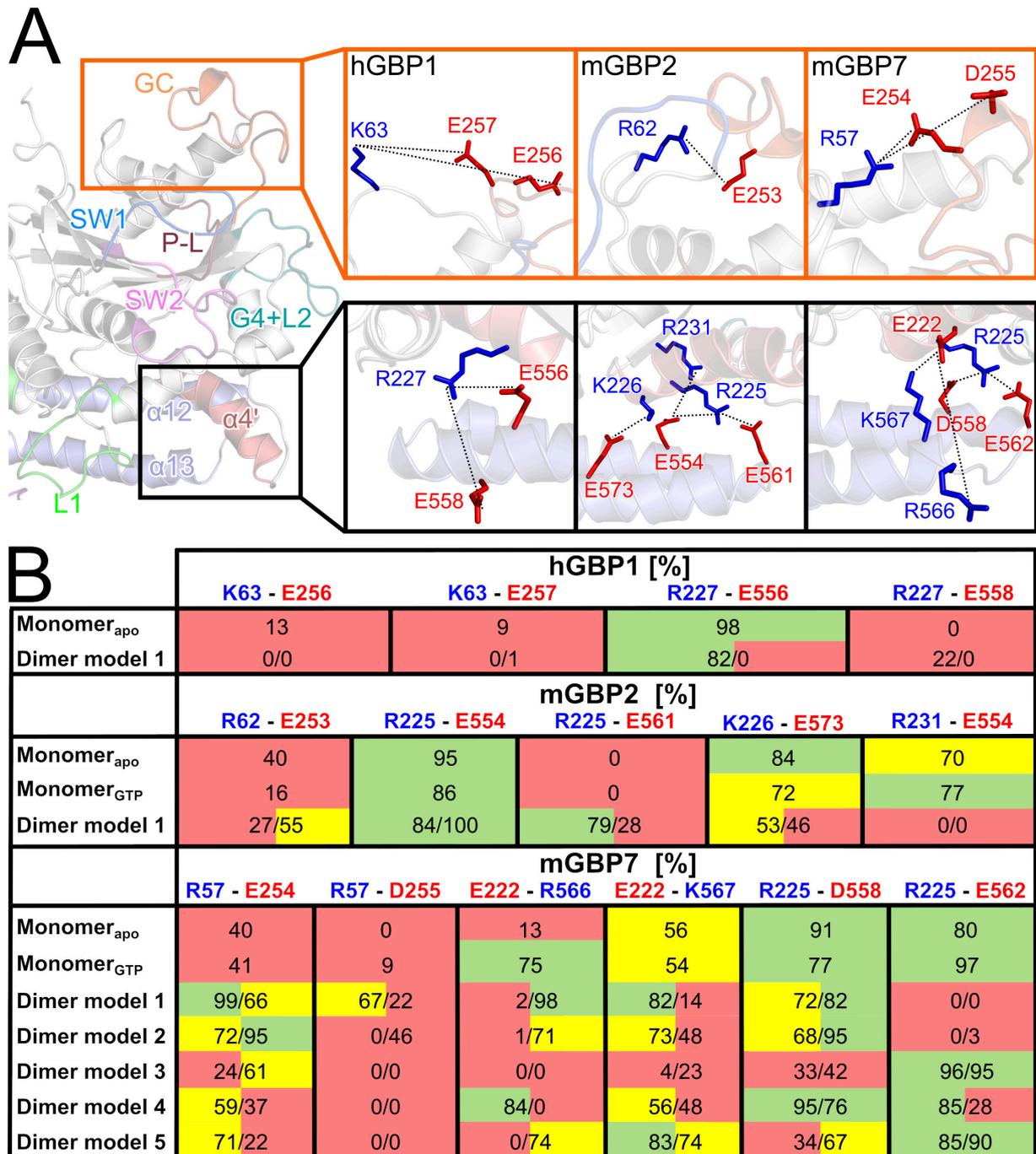
**Figure 2: Structural fluctuations of the monomeric hGBP1, mGBP2, and mGBP7.** (Left) The protein flexibility, as quantified by the RMSF of the  $C_{\alpha}$  atoms during the respective HREMD simulation, is projected onto the initial structure of hGBP1<sub>apo</sub>, mGBP2<sub>apo</sub>, mGBP2<sub>GTP</sub>, mGBP7<sub>apo</sub> and mGBP7<sub>GTP</sub>. Rigid residues are colored in blue and flexible residues are shown in red, according to the color scale at the top. (Right) The spatial distribution of the residue 480 (i.e., the tip of the M/E domain) is shown as orange cloud to illustrate the hinge motion of the corresponding GBP monomer. Taking hGBP1<sub>apo</sub> as an example, the definition of the coordinate system is shown for the quantification of the hinge motion using the motions of the residue 480 with respect to the initial structure, as measured by  $\Delta x$ ,  $\Delta y$ ,  $\Delta z$  and  $d_{480}$ .

To further characterize the motions of the G domain motifs and loops, we performed a clustering analysis of the G domains, with the simulation snapshots being fitted onto the  $\beta$ -sheets of the corresponding G domain. The resulting numbers of clusters, their populations, and maximal structural differences between the clusters are summarized in Table S1. The cluster data show that among the apo-states, mGBP2 has the most stable G domain, which becomes even stiffer in mGBP2<sub>GTP</sub>, except for L1, which shows only an insignificant RMSF decrease, and  $\alpha 4'$ , whose flexibility even increased slightly (3 instead 2 clusters). In mGBP7, the G motifs and loops are generally more flexible than in mGBP2 and the stabilization by GTP bindings is also smaller. This is most significant for the guanine cap, for which the number of clusters decreases only by 14% from mGBP7<sub>apo</sub> to mGBP7<sub>GTP</sub> as compared to a 90% reduction in mGBP2. The other structural G domain elements of mGBP7 experience larger decreases in the number of clusters, yet apart from P-L and G4+L2 they remain rather flexible in mGBP7<sub>GTP</sub>. This can also be seen in the structural presentation of the clusters in Fig. S4. The different behavior in these mGBP7 motifs and loops as compared to hGBP1 and mGBP2 largely correlates with the amount and type of changes in their amino acid composition. In particular the negative charge of the guanine cap in mGBP7, as opposed to charge neutrality of that region in the other two proteins, might cause its generally higher flexibility and electrostatic repulsion from GTP. This in turn prevents its stabilization in mGBP7<sub>GTP</sub> as seen in mGBP2<sub>GTP</sub> where the guanine cap becomes more ordered (Fig. S4). Interestingly, the relatively high flexibility of the G domain of mGBP7 does not affect the binding affinity of GTP, which is even somewhat higher for mGBP7 than for hGBP1 and mGBP2 (Table 1). It thus seems to be sufficient if the amino acid residues directly involved in GTP binding, such as R48 and K51 become rigid if GTP is present in order to enable a stable GTP binding site. It is interesting to note that mGBP7 shows a higher dynamics in the G domain independent of the GTP loading state, while its M/E domain is more rigid than in the other two GBPs.

**Hinge motion of the M/E domain** The dynamics of the M/E domain is characterized as a hinge motion, as first discovered for hGBP1<sup>22</sup> and then confirmed for mGBP2.<sup>23</sup> To quantify that motion, we computed the motions of residue 480 at the tip of the M/E domain relative to its position in the crystal structure or homology model during the HREMD simulations. These motions of residue 480 are provided as changes in its Cartesian coordinates ( $\Delta x$ ,  $\Delta y$ ,  $\Delta z$ , see Fig. 2 top right for the definition of the coordinate system) as well as its Euclidean distance between the current and reference position ( $d_{480}$ ). The resulting data are plotted as statistical box plots in Figs. S5 and S6, and are illustrated by the spatial distribution of residue 480 (Fig. 2). The difference in the three GBPs is mainly described by  $\Delta y$ , while  $\Delta x$  and  $\Delta z$  are similar. For hGBP1, motions in all directions are possible, but motions in  $+\Delta y$  and  $-\Delta z$  are favored and resemble a jack-knife where the tip of the M/E domain is moving towards the G domain. In hGBP1, the absolute motion of residue 480 is the highest of all three GBPs, with the  $d_{480}$  values reaching up 69.8 Å. The tip of the M/E domain of mGBP2 also favors motions into  $+\Delta y$  direction, in both its apo and GTP-bound state, which is accompanied by  $-\Delta x$  motions as a result of the jack-knifing of the domain. With regard to motions along the  $z$  coordinate, we see a displacement towards  $-\Delta z$  in mGBP2<sub>apo</sub>, while with GTP both directions of  $\Delta z$  are equally achieved. We revealed that the motions in the  $+\Delta z$  direction are accompanied by a stabilization of the salt bridges between  $\alpha 4'$  and  $\alpha 12/13$ , which became possible due to a displacement of  $\alpha 4'$  towards the E domain as a result of GTP binding.<sup>23</sup> The maximal motion of residue 480 in mGBP2 is slightly lower than in hGBP1, but can also reach values of about 60 Å (mGBP2<sub>apo</sub>: max.  $d_{480} = 62.4$  Å; mGBP2<sub>GTP</sub>: max.  $d_{480} = 58.9$  Å). In mGBP7<sub>apo</sub>, the motion into  $-\Delta y$  and  $+\Delta z$  direction is preferred, while in mGBP7<sub>GTP</sub> the M/E domain tip chooses to move into the opposite direction, i.e., into  $+\Delta y$  and  $-\Delta z$  direction. This implies that, as in mGBP2, GTP binding affects the direction of the hinge motion, yet in the opposite way, which correlates with the different structural effects that GTP binding has on the G domains of mGBP7 and mGBP2. Another difference is that  $\Delta x$  is almost zero in mGBP7. This results from the overall smaller jack-knifing in mGBP7,

as revealed by the maximal displacement of residue 480 (mGBP7<sub>apo</sub>: max.  $d_{480} = 35.2 \text{ \AA}$ ; mGBP7<sub>GTP</sub>: max.  $d_{480} = 37.5 \text{ \AA}$ ). This can be also seen from the spatial distribution of that residue in Fig. 2.

**Salt bridges** We finish the analysis of the monomeric GBP simulations by assessing the stability of intramolecular salt bridges that are thought to be relevant during protein dimerization, in particular salt bridges within the guanine cap and between  $\alpha 4'$  and  $\alpha 12/13$ . The salt bridges in question are shown in Fig. 3A and their occurrences during the simulations are summarized in Fig. 3B. In hGBP1<sub>apo</sub>, the salt bridge between K63 and E256 or E257 in the guanine cap has an occurrence below 15%, while a very stable salt bridge with 98% probability formed between R227 of  $\alpha 4'$  and E556 or E558 of  $\alpha 12$ . In the corresponding simulation of mGBP2<sub>apo</sub>, the salt bridge in the guanine cap (R62–E253) was more stable, with a population of 40%. However, upon GTP binding this probability decreased to 16%, which can be explained by the ordering of the guanine cap in mGBP2<sub>GTP</sub>, allowing the residue R62 to also interact with E249 with 77% probability (only 1.7% in mGBP2<sub>apo</sub>). The salt bridges K226–E573 and R231–E554 tethering  $\alpha 12/13$  to  $\alpha 4'$ , are stable in both mGBP2 states with populations of over 70%, while the salt bridge R225–E561 does not exist in either monomer state. In mGBP7, the salt bridge R57–E254 in the guanine cap is stable independent of the GTP binding state, which agrees to the observation that also the structure of the mGBP7 guanine cap is hardly affected by GTP binding. The other possible salt bridge, R57–D255, is only formed in mGBP7<sub>GTP</sub>, yet also only with 9% probability. With regard to the  $\alpha 4'$ – $\alpha 12/13$  interactions in mGBP7, there are several possibilities for salt bridges, in particular between E222 and R566/K567 as well as R225 and D558/E562. They were all stable in both mGBP7<sub>apo</sub> and mGBP7<sub>GTP</sub> with occupancies above 50%. The only exception is the E222–R566 salt bridge in mGBP7<sub>apo</sub> for which only a 13% occupancy was found.



**Figure 3: Intramolecular salt bridges relevant for dimerization.** (A) On the left, parts of the G and the E domain are shown. The location of the salt bridges within the guanine cap and between  $\alpha 4'$  and  $\alpha 12/13$  are indicated by an orange and a black box, respectively. On the right, the zoomed views of these areas are shown for hGBP1, mGBP2, and mGBP7. The possible salt bridges are indicated by dotted lines between the residues being involved, which are labeled and their side chains highlighted as red and blue sticks for negatively and positively charged residues, respectively. (B) Occupancy (in %) of the salt bridges during the simulations. The results for chain 1 and chain 2 are given separately for the dimers. The results are colored based on the occupancy: 0–49%, red; 50–74%, yellow; 75–100%, green.

An interesting difference between mGBP7 and the other two GBPs is that  $\alpha 4'$  of mGBP7 donates a positively and a negatively charged residue for the formation of salt bridges with  $\alpha 12/13$ , whereas in hGBP1 and mGBP2 only positively charged residues in  $\alpha 4'$  are available for electrostatic interactions with  $\alpha 12/13$ . Apart from that, the overall conclusion is that in all three proteins there are stable salt bridges between the G and the E domain that kept the three helices  $\alpha 4'$ ,  $\alpha 12$  and  $\alpha 13$  closely together independent of the GTP loading state.

## 2.3 Dimer models of GBPs

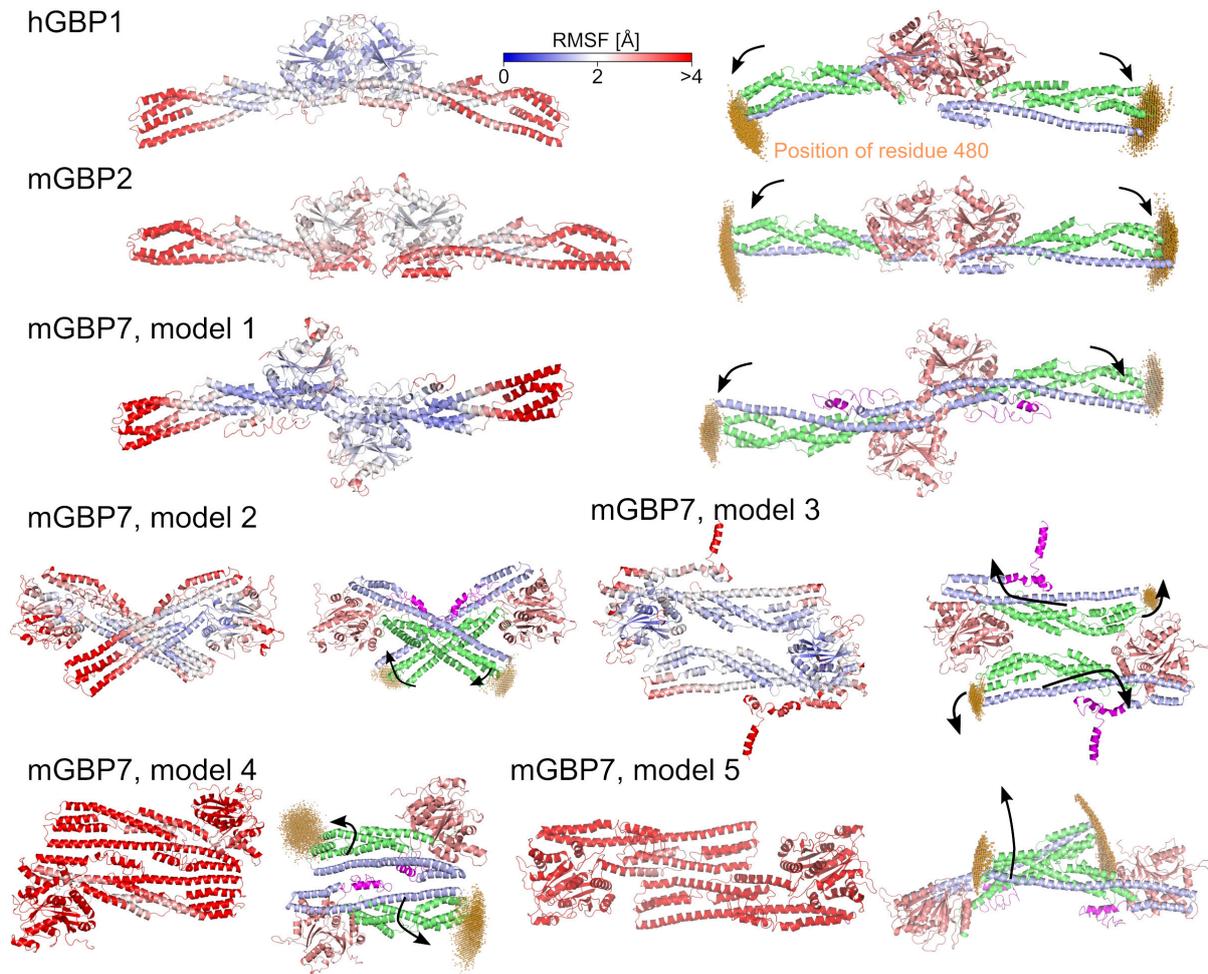
Our long-term goal is to elucidate the mode of action of membrane-bound GBP multimers involving up to thousands of subunits. To reach that goal, we first need structures for the polymerized GBPs and an understanding of their dynamics, especially also in comparison to the dynamics of their smallest organizational subunits, the monomers, which we just presented. Here, we continue by providing possible structure models for the smallest oligomer, the dimer, and assess the stability and motions of these GBP dimers.

### 2.3.1 Model generation

We created structural models for the dimers of hGBP1, mGBP2, and mGBP7 based on experimental data, taken from the literature if available or produced within this study, combined with protein–protein docking. In the case of hGBP1, we constructed the dimer model using the crystal structure that exists for the dimer of its G domain (PDB IDs: 2B8W, 2B92 and 2BC9<sup>3</sup>) and aligning the G domains of two hGBP1 molecules with that structure (Fig. 4). It should be noted that we modeled the hGBP1 and mGBP2 dimers with GTP being bound as both proteins only notably dimerize in the presence of GTP,<sup>3,41,48</sup> whereas the mGBP7 was considered in its apo-state as the dimerization of this protein is not affected by GTP<sup>24</sup> and the experimental mGBP7 data obtained here was also recorded without GTP. We tested for the existence of alternative hGBP1 dimerization motifs using ClusPro,<sup>56</sup> which is a physics-based protein-protein docking program. As such, it does not incorporate structural

information from the PDB.<sup>57</sup> Nonetheless, the best hGBP1 dimer structure proposed by ClusPro is very similar to the hGBP1 dimer created from the structure of the G domain dimer. Their RMSD from each other is only 5.0 Å, which is small for the system size under consideration, and their topological similarity based on the so-called TM-score calculated with MMalign<sup>58</sup> is rather high with a value of 0.73. The TM-score can range between 0 and 1, with 1 corresponding to a perfect match between two structures, a TM-score of > 0.5 indicating that two structures have a similar topology, and a TM-score < 0.17 suggesting that the structural similarity is close to random. The ClusPro docking scores for the other hGBP1 dimer models were low, and we therefore did not consider them in our further study.

A similar picture emerged for mGBP2 as a result of its 76.3% sequence similarity with hGBP1. The most likely mode of dimerization predicted by ClusPro is via the G domains, resulting in an mGBP2 dimer model very similar to the crystal structure-based one for hGBP1 (Fig. 4). The following five best mGBP2 dimer models are also all G domain dimers, with different rotations of the two proteins with respect to each other, suggesting that this interface is quite robust. However, these models had considerably lower docking scores than the top model. The application of AlphaFold-Multimer<sup>59</sup> to mGBP2 produced the same outcome, i.e., only G domain dimers were suggested.<sup>23</sup> Therefore, only the top mGBP2 dimer model produced by ClusPro was considered here.



**Figure 4: Structural fluctuations of the hGBP1, mGBP2, and mGBP7 dimers.** The dimer models considered in this work are shown. The overall stability of the dimers was quantified by the RMSF of the  $C_{\alpha}$  atoms during the MD simulations of these dimers. The RMSF was calculated by aligning the whole dimer and is projected onto the initial dimer model, where rigid residues are colored in blue and flexible residues in red (color scale at the top). The degree of the hinge motion of the M/E domains was determined by the spatial distributions of the residue 480, which are shown as orange clouds. The coordinates used for measuring the hinge motion is the same as in Fig 2, with each chain of the dimers being aligned individually in that coordinate system. The black arrows indicate the main motions resulting from PCA analysis.

For mGBP7, the situation is different. First, ClusPro predicts three dimerization modes of equal likelihood for mGBP7. While one of them agrees nicely to the XL-MS data gathered here (discussed below), none of them could be satisfactorily fit into the molecular envelope reconstructed from our SAXS data. Therefore, we generated two further mGBP7 dimer

models by fitting two mGBP7 proteins into molecular envelopes corresponding to the SAXS data using two different fitting techniques (discussed below). In total, we thus produced five mGBP7 dimer models, of which models 1–3 are from the ClusPro prediction and models 4 and 5 are from the SAXS data fitting (Fig 4). None of the mGBP7 dimers considered here involves the G domain interface seen for hGBP1 and predicted for mGBP2. ClusPro did not produce such a dimer model. Interestingly, the application of AlphaFold-Multimer<sup>59</sup> to mGBP7 creates only G domain dimers (Fig. S7), yet their DockQ scores are low with respect to mGBP2 as reference. This indicates a poor match of amino acids at the dimer interface,<sup>60</sup> which can be rationalized with the different amino acid compositions of the guanine caps of both proteins. Since the G domain dimer model is also not in agreement with our experimental data for mGBP7 dimers, we did not consider it further.

In the following, the G domain dimer of hGBP1 and mGBP2 and dimer models 1–5 of mGBP7 are discussed in detail. To assess the stability and flexibility of the dimers, we performed 100 ns MD simulations per dimer and calculated the RMSF and stalk tip motions. Moreover, a principal component analysis (PCA) was applied to the MD data to identify the most prominent collective motions. To assess the dimer interface, we calculated the dimer contact area, analyzed selected salt bridges (see Fig. 3A), and determined the energies of the intermolecular contacts.

### 2.3.2 Structural details of the hGBP1 and mGBP2 dimers

The dimerization via the G domains as considered for hGBP1 and mGBP2 allows to bury a large protein area. Of all dimer models discussed here, the hGBP1 dimer has the largest interface, with 5,314 Å<sup>2</sup>, and also the mGBP2 dimer model is stabilized by a large interaction area of 3,129 Å<sup>2</sup>.

To assess the dimer stability and flexibilities, we make use of the RMSF again. When aligning the whole dimer for the RMSF calculation, we obtain a measure for the overall dimer stability (Fig. 4), which reveals dimer model for hGBP1 to be more rigid than for mGBP2.

The flexibilities of the chains within the dimers (Fig. S8), calculated after alignment of each chain, are very similar to the RMSF profiles of the corresponding monomers, suggesting that dimerization does not have a large effect on protein movements. As for the monomers, the stalk tips are the most flexible, with RMSF values of up to 8 Å. Notably, in the monomer state mGBP2 is less flexible than hGBP1, while it is the other way round for the dimers. Other flexible regions in both dimers are L1, G4+L2,  $\alpha 4'$ , and the guanine cap. Nonetheless, compared to the monomers, the guanine cap as well as SW1 are stabilized, which correlates with salt bridge formation between the two chains (Fig. 3B). The two hinge motions per dimer, as described by the spatial distribution of residue 480 (Fig. 4) and the motions of this residue relative to its starting structure as defined by the  $\Delta x, y, z$  values (Fig. S5) are similar, but they are more pronounced in mGBP2 as evidenced by higher  $d_{480}$  values due to larger motions along almost all axes. The preferred directions of the motions appear along  $+\Delta y / \pm \Delta z$ . The two hinge motions in either dimer are correlated with each other, as revealed by the PCA. The first principal motion can be likened to a butterfly motion (see arrows in Fig. 4), leading to dimer geometries that are more curved or C-shaped, which is accompanied by the formation of an intramolecular contact between residues K62 and D255 (Fig. 3) and unfurling of the C-terminal regions (yet more so in one of the two chains per dimer). The motion requires a certain structural flexibility at the dimer interface, especially in all GTPase motifs. The second principal motion is comparable to the first one, yet involves more screwing of the M/E domain in the lateral dimension.

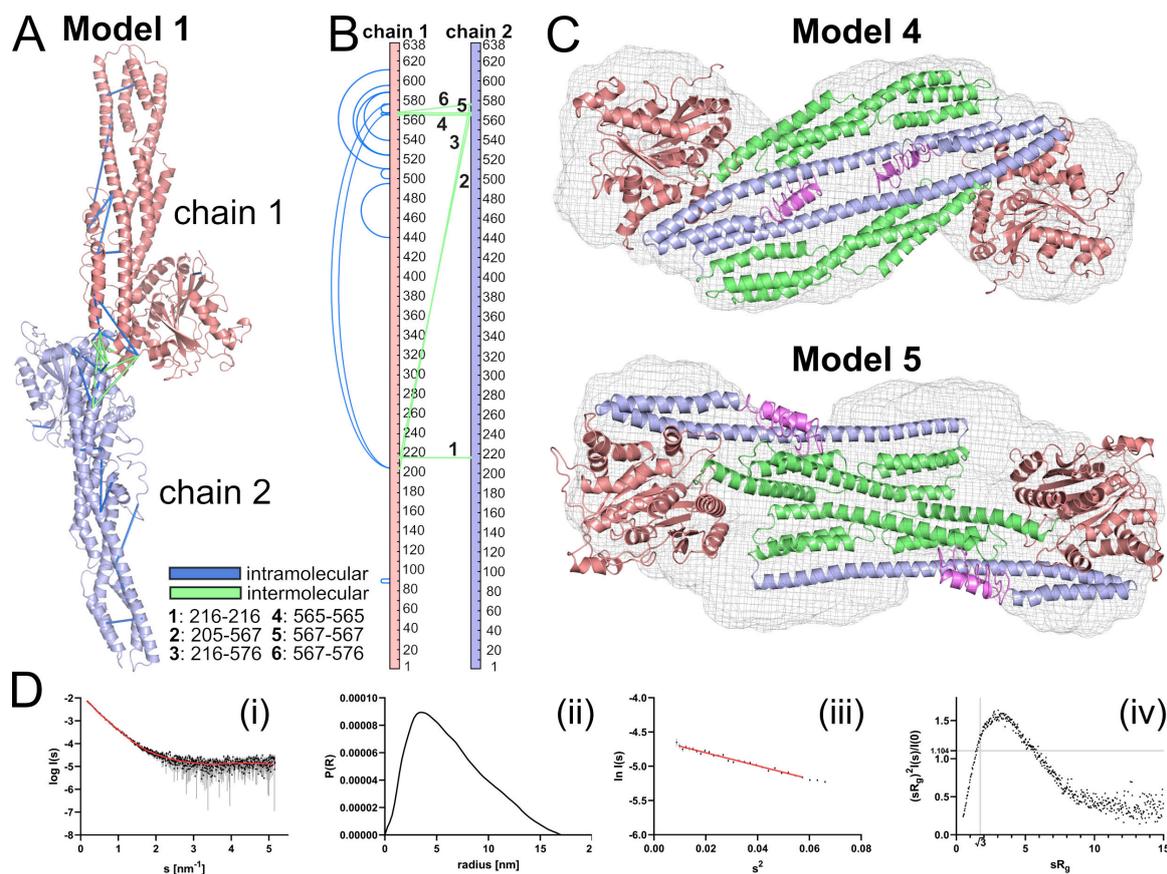
The structure of the guanine cap is stabilized in both dimers compared to the monomeric proteins, yet the stabilization originates from partly different interactions in the hGBP1 and mGBP2 dimer (Fig. 3). The occurrence of the intramolecular salt bridge K63–E256/E257, which already in the monomer only occurred with a probability of  $\sim 10\%$ , is abolished in the hGBP1 dimer. The corresponding salt bridge R62–E253 in the mGBP2 dimer gained in strength with respect to the mGBP2<sub>GTP</sub> monomer, reaching a chain-averaged occurrence similar to the 40% observed for the mGBP2<sub>apo</sub> monomer. The  $\alpha 4'$ – $\alpha 12/13$  lock is overall

weakened in both dimers, but more so in the hGBP1 dimer. In mGBP2, where more salt bridges between  $\alpha 4'$  and  $\alpha 12/13$  are possible, the R225–E554 interaction is of particular stability and remains intact with a  $> 84\%$  probability in all mGBP2 systems studied here. With regard to intermolecular interactions, the hGBP1 dimer harbors strong (about  $-80$  kJ/mol) contacts involving both guanine caps, such as E261–K252', K252–D239', and R245–E251'/E256' as well as E105–K209' (where the prime indicates that these residues belong to the other chain), which are mostly Coulombic in nature. The mGBP2 dimer is also held together by interactions between both guanine caps, but further involves interactions of residues in the 130ies and of the G4+L2 motif that form contacts with their respective counterpart in the other chain. The single strongest interaction is that of D237–R238', with an interaction energy of  $-120$  kJ/mol (Fig. S9).

### 2.3.3 Experimental data for mGBP7 dimers

**XL-MS** The aim of XL-MS was to obtain information on the mGBP7 dimerization interface. To this end, recombinant mGBP7 was incubated in solution with either of two crosslinkers: bis(sulfosuccinimidyl) suberate (BS3, 11.4 Å) or disuccinimidyl sulfoxide (DSSO, 10.1 Å). Both crosslinkers react mainly with lysine residues and can link two lysines within a distance of 12 Å. Potential mGBP7 dimers and monomers were separated in a polyacrylamide gel and separately analyzed by mass spectrometry. Crosslinked peptide pairs were subsequently mapped to the different mGBP7 dimer models created with ClusPro and to the mGBP7 monomer (summarized in Table S2). A focus of the analysis was the identification of potential intermolecular contact sites within the homodimer, for which especially bridges between the same residues (like 216–216', 565–565', 567–567') – as revealed by BS3 crosslinking – were informative. Of the three mGBP7 dimer models that resulted from docking, only model 1, which has an elongated shape with the dimer interface being formed via the two G domains (Fig. 4), agrees with our crosslink information (Fig. 5A and B), while the G domain dimers predicted by AlphaFold disagree. The XL-MS intermolecular contacts can be mapped to the

$\alpha 4'$  of the G domain and  $\alpha 13$  of the E domain. Moreover, the interactions 205–567', 216–576' and 567–576' for BS3 as well as 106–216', 205–554', 205–565' and 557–567' for DSSO are in agreement with the intermolecular contacts in this model. In conclusion, this is the most probable model that fits to the XL-MS results.



**Figure 5: Results of the XL-MS and SAXS data of mGBP7 dimer.** (A,B) Visualization of the crosslinks matching model 1 for the mGBP7 dimer. Mapped intermolecular crosslinks (numbers of the linked residues are given) are shown in green, while intramolecular crosslinks are colored in blue. Crosslinks are indicated in the 3D model in (A) and in a linear sequence plot (created with xiNET<sup>61</sup>) in (B). (C) Calculated SAXS envelope for mGBP7 with dimer models fit into it: (top) model 4 created with GASBOR<sup>62</sup> and (bottom) model 5 created with SASREF.<sup>63</sup> (D) (i) SAXS curve showing the intensity as a function of momentum transfer  $s$ . The experimental data curve is given as black dots with grey error bars, and the red line is for the GASBOR *ab initio* model fit ( $\chi^2$  of 1.14). (ii) The distance distribution, shown as the  $p(r)$  function, provided a maximum particle diameter ( $D_{\max}$ ) of 17.0 nm. (iii) The Guinier plot revealed a Guinier region (red line) in the range of  $sR_g < 1.3$ , leading to a  $R_g$  of 5.35 nm. (iv) The Kratky plot agrees with a compact shape for the mGBP7 dimer with certain flexibility.

**SAXS** With SAXS we also aimed to reveal the mGBP7 dimer organization. We used a protein concentration range of 1.66–6.85 mg/ml and merged the low  $s$  data from the 1.66 mg/ml concentration with the high  $s$  data from the 6.85 mg/ml concentration. The SAXS data are summarized in Table S3. We calculated an *ab initio* dimer model with GASBOR,<sup>62</sup> using a  $P2$  symmetry, yielding the dimer model shown in Fig. 5C (denoted as model 4 here) which has a tail–tail interface between the E domains of the two mGBP7 chains. We calculated the theoretical intensity of the mGBP7 dimer model with CRY SOL<sup>64</sup> and compared it with the experimental scattering data, resulting in a  $\chi^2$  of 1.89 (Fig. 5D(i)). With the Guinier approximation,<sup>65</sup> we determined the radius of gyration ( $R_g$ ) as 5.35 nm and used the distance distribution, shown as the  $p(r)$  function, to specify the maximum particle dimension ( $D_{\max}$ ) as 17 nm (Fig. 5D(ii) and (iii)). The dimensionless Kratky plot showed a compact elongated shape for the mGBP7 dimer (Fig. 5D(iv)).

Since the  $\chi^2$  value for model 4 is somewhat large, we created a further dimer model for mGBP7 (denoted as model 5) using SASREF,<sup>63</sup> which performs quaternary structure modeling of a complex formed by subunits with known atomic structure against the SAXS data set. The resulting model has larger  $R_g$  and  $D_{\max}$  values compared to model 4, yet the  $\chi^2$  is closer to 1 with a value of 1.27 (Table S4). As model 4, this dimer has a tail–tail interface, yet the dimerization is predicted to occur via the M domains (Fig. 5C, bottom). We further tested how good the three ClusPro models (models 1 to 3) are in agreement with the SAXS data. To this end, we docked them into the calculated molecular envelope. Table S4 lists the resulting  $\chi^2$ ,  $R_g$  and  $D_{\max}$  values. A  $\chi^2$  close to 1 is desirable, yet also large deviations from the experimentally determined values of  $R_g = 5.35$  nm and  $D_{\max} = 17$  nm render a dimer model as disagreeing with the SAXS data. Indeed, more emphasis is given on the agreement for the physical dimensions  $R_g$  and  $D_{\max}$  than on  $\chi^2$ . Based on these considerations, none of the three mGBP7 models created by ClusPro fully complies with the SAXS data; only model 3 might still be a suitable model. The least fitting is model 1, as its overall length is clearly larger than the SAXS envelope. A head–head interface via the G domains as predicted

by AlphaFold (Fig. S7) is for the same reason also unlikely as its  $D_{\max}$  value is in conflict with the calculated  $p(r)$  function.

### 2.3.4 Structural details of the mGBP7 dimers

In contrast to hGBP1 and mGBP2, mGBP7 shows a strong tendency to dimerize via its stalks in an antiparallel arrangement (Fig. 4). This is seen for models 2 and 3 created by ClusPro, and also applies to the next five best docking predictions not further discussed here, and models 4 and 5 resulting from fitting to the SAXS data. The crossed-stalks motif present in model 2 and to some extent also in model 5 is reminiscent of the dimerization pattern observed for other members of the dynamin superfamily.<sup>66,67</sup> Only model 1, the most probable docking prediction, dimerizes via the G domains and some contacts involving  $\alpha 13$  (Fig. 4). However, this dimerization mode is different from the G domain dimers observed for hGBP1 and mGBP2. This structure is further interesting since it fulfills 3 of the 4 constraints imposed by XL-MS. Regarding the fourth constraint it should be noted that it clashes with the other three constraints, as revealed by a DisVis analysis using the HADDOCK webserver,<sup>68</sup> indicating that no dimer model would be able to fulfill all four XL-MS constraints at the same time. It is possible that this constraint is due to an artifact, a transient conformational change, or that more than one dimer shape can be assumed. With regard to the dimer interface, model 3 buries the largest protein surface with 3,414 Å<sup>2</sup>, followed by model 1 (2,678 Å<sup>2</sup>) and model 2 (2,581 Å<sup>2</sup>). The smallest contact areas are observed for model 4 (2,224 Å<sup>2</sup>) and model 5 (1,560 Å<sup>2</sup>), which further decreased during the MD simulations. This indicates instability of these models. Moreover, model 4 had to be slightly adjusted to remove atom clashes before the simulation could even be started. Model 2, on the other hand, notably increased its dimer contact area during the simulation, suggesting stability and optimization of the dimer structure. Each mGBP7 dimer model is now discussed in more detail.

**Model 1** The flexibilities of the M/E domains of mGBP7 in the dimer model 1 differ from the structural fluctuations (measured by the RMSF) of the mGBP7 monomer and are more similar to those of the mGBP2 dimer (Fig. S8). In particular  $\alpha_{10}$  and  $\alpha_{12}$  are more flexible than in the mGBP7 monomer. The guanine cap is stabilized, involving stronger intramolecular salt bridges between R57 and E254/D255 than in either the apo- or GTP-bound mGBP7 monomer (Fig. 3). Within both chains,  $\alpha_{4'}$  and  $\alpha_{12}$  remain connected via a network of four redundant salt bridges of which in either chain two remain highly occupied (Fig. 3). However, the salt bridge R225–E562 completely disappeared. The first principal component of dimer model 1 is a butterfly motion, wherein the hinges move in a correlated fashion to form a more curved dimer, while the CT tail unfurls from its pocket. The second principal component combines the butterfly motion (yet in a reduced form) with a slight screwing motions of the M/E domains. The directions of the hinge motions in both chains are the same as in the hGBP1 and mGBP2 monomers and dimers, with  $+\Delta y/+ \Delta z$  as preferred direction and  $-\Delta y/+ \Delta z$  being possible too. The maximum distance of  $d_{480}$  is 25–31 Å and on average it is  $\approx 10$  Å, which is comparable to the motions in the mGBP7 monomer (Fig. S6). The strongest intermolecular interactions in the mGBP7 dimer model 1 are between the L1 and  $\alpha_{4'}$  (E161–K213', S211–D164'), between  $\alpha_{4'}$  and the E domain (G209–K567', R222–R566'), and between the two E domains (R566–E562') with interaction energies in the range of  $-80$  to  $-40$  kJ/mol (Fig. S9). These are partly the same residues as those involved in intramolecular salt bridges, showing the competition between intra- and intermolecular salt bridges, and explaining that the intramolecular salt bridge R225–E562 completely disappeared while those involving R566 or K567 were absent in one of either chain of the dimer.

**Model 2** Interestingly, the protein motions within model 2 are very similar to those in model 1, even though the dimerization interface is completely different. The M/E domains in models 1 and 2 are similarly flexible. The intramolecular salt bridges in the guanine cap strengthened in both dimers compared to the monomer, while the number of salt bridges

formed between  $\alpha 4'$  and  $\alpha 12$  is smaller, but the overall contact remains intact. As in model 1, the R225–E562 salt bridge is not formed in either chain of model 2 and E222–R566 is also not present in chain 1. This is also the chain displaying a higher tip flexibility, and we can thus conclude that the E222–R566 salt bridge strength is anticorrelated with the flexibility of the M/E domain. The butterfly motion resulting from the synchronous hinge motions of both chains is also present in model 2. Nonetheless, the first principal motion is decidedly asymmetric, with the stalks twisting around each other, which results in a contact between one G domain to the stalk tip of the other chain. This can be correlated with the observation of dissolved salt bridges between  $\alpha 4'$  and  $\alpha 12$  and a higher stalk tip flexibility. The second principal motion is similar, albeit symmetric, and here the motion originates not from the stalk-middle but from the base of the G domain. In analogy to DSP nomenclature, this could be termed hinge 2, which is not strongly visible in the GBP monomers. The directions of the hinge 1 motion in model 2 are the same as in model 1 (Fig. S6). The strongest intermolecular interactions are found for the two electrostatic interactions E389–K508' and E399–R396' reaching about  $-110$  kJ/mol, with supporting weaker interactions around residues 510 and 610.

**Model 3** In model 3, the mGBP7 proteins are less mobile compared with the other dimer models and also the protein monomer (Fig. S8). Only the CT tail remains flexible. The occupancy of the intramolecular salt bridge in the guanine cap is similar in the mGBP7 monomer and this dimer model, and thus smaller than in the other dimers. Moreover, also three of the four salt bridges tacking  $\alpha 4'$  and  $\alpha 12$  together have an occupation below 50%, indicating an increased tendency of the E domain to detach from the G domain. However, the fourth salt bridge, R225–E562, remains present to a similar degree as in the monomer, whereas in models 1 and 2 this particular salt bridge was absent (Fig. 3B). The first principal motion of model 3 results in a saddle-like structure where the rather planar shape of the dimer becomes more curved due to both stalk tips moving into the same direction. The tip

of one of the M/E domains even detaches from the G domain of the other chain, enabling a larger range of motions for the E domains that together almost form a cross shape. The main parts of the helices of the M domains, on the other hand, remain parallel with respect to each other, meaning that the  $\alpha_{12}$  helices become slightly angled against the rest of the protein. The motion is accompanied by the CT tails curling up against the G domains. The second principal motion is very similar to the saddle-formation motion, yet with only one of the CT tails curling up, while the other one stays compact. The tips of the M/E domains move mostly in  $+\Delta y / -\Delta z$  and  $-\Delta y / +\Delta z$  direction, and the range of the motions is larger than in the monomer (Figs. 4 and S6). The maximum amplitude of that motion is 25–38 Å, with the average values being between 10 and 20 Å, which is comparable to the motions in models 1 and 2. However, all three mGBP7 dimer models discussed so far move generally less than the mGBP2 dimer, which replicates the findings for the mGBP7 and mGBP2 monomer. The intermolecular interactions that contribute the most to the dimer stability are R33–D411' and K26–E419', reaching  $-120$  kJ/mol and  $-60$  kJ/mol, respectively, which are supported by weaker interactions between the guanine cap and the stalk tip (Fig S9). Moreover, the N-terminal E8 is also of relevance as it interacts with several residues of  $\alpha_{10}$  of the other chain, summing up to an additional  $-140$  kJ/mol of interaction energy.

**Model 4** Model 4 is generally very flexible, which is caused by the overall instability of this dimer model, involving the loss of one of the two main interprotein contact interfaces (Fig. 4). In terms of intraprotein flexibilities, the SW2, G4 motif, and guanine cap are more mobile than in models 1 to 3, while the  $\alpha_{10/12}$  region is only slightly more flexible than in models 1 and 2. The guanine cap is of similar flexibility as in the mGBP7 monomer, which is also reflected in a comparable occupancy of the intramolecular salt bridge within that region. The considerable dissolution of the intramolecular salt bridges connecting  $\alpha_{4'}$  and  $\alpha_{12/13}$  in chain 2 of model 4 (Fig. 3B) could be at fault for the increased mobility of the dimer, as  $\alpha_{13}$  provides a large part in the contact area between both chains. As in model 3,

the first principal component can be described as a saddle formation due to a bending of both chains. The second main motion is a sliding of the stalks against each other along the M/E domain axis, which likely results from the accumulation of negatively charged amino acid residues at the dimer interface and the involvement of the highly flexible CT tails therein. Model 4 nominally has a greater range of motions compared to the other dimers, but due to the disruption of the dimer, this is less meaningful. The strongest interprotein contacts are K588–E501' and E501–R636' with energies of about  $-50$  kJ/mol, which is weaker than for any of the primary interaction pairs in the other dimer models. Moreover, the first principal motion even results in an opening of the contact site around E501-K637'. The conclusion therefore is that model 4 is not a stable dimer.

**Model 5** The final mGBP7 dimer model displays a crosswise interaction via the top parts of the two M/E domains, which prevents contact formation between the stalk tip of one chain and the G domain of the other chain as in models 2 and 3. However, this interaction interface is not particularly stable as within the 100 ns MD simulation, it is broken and replaced by an asymmetrical stalk tip–G domain interaction. Reasons for that asymmetry are the higher flexibility observed for one of the CT tails and a generally great accumulation of negative charges at the original dimer interface. The motion leading to the change in interaction interface is contained within the first principal component, while the second principal component is an asymmetric motion caused by one chain swinging out and rotating against the other. The individual chains display the same flexibility pattern observed before. The range of the hinge motion is comparable to that in model 3, but as for model 4 this is not of relevance due to instability of model 5. For the latter reason, the overall flexibility of this dimer model is as high as for model 4. The intramolecular salt bridge in the guanine cap is not particularly strong (20–30% population on average), while the  $\alpha 4'$ – $\alpha 12/13$  lock is still engaged (50–76% on average). The only strong intermolecular contact that is left from the original interface is E336–R418', contributing  $-100$  kJ/mol.

### 3 Discussion and Conclusion

We compared the sequence and biochemical properties of three GBPs, the human GBP1 and the murine GBP2 and GBP7. The main conclusion from that juxtaposition is that mGBP2 is more similar to hGBP1 than mGBP7 is to mGBP2. Both hGBP1 and mGBP2 are post-translationally modified by isoprenylation where instead mGBP7 exhibits 49 additional C-terminal residues for membrane binding. Moreover, the sequence of mGBP7 differs markedly at various places from those of hGBP1 and mGBP2. The different distribution of charged residues gives rise to different protein dynamics, as we revealed by HREMD simulations of the monomeric proteins. We generated dimer models with ClusPro for the three GBPs, where hGBP1 and mGBP2 involve the G–G domain dimer, also called head-to-head dimer in the following, that was already solved as crystal structure for the G domain of hGBP1.<sup>3</sup> For mGBP7, we obtained three dimer models, of which model 1 agrees with our XL-MS data. We generated two further mGBP7 dimer models by reconstruction from SAXS data.

#### 3.1 General comparison of the dimer models

In order to understand the dimer dynamics, we first simulated the monomers of the three GBPs in solution serving as a reference. For mGBP2 and mGBP7 both the apo- and GTP-bound state were considered, while the hGBP1 monomer was only simulated as apo-protein. In the apo-state of the GBP monomers, the different motifs and loops of the G domains are very flexible, one reason being that R48, which serves as arginine finger, has no stable contact without GTP. The simulations further revealed that all three GBPs feature the characteristic hinge motion involving the M/E domain that we had originally uncovered for hGBP1.<sup>22</sup> Among them, hGBP1 exhibits the largest hinge motion and mGBP7 the smallest one. A probable reason for this is that the CT tail has a stabilizing effect on the M/E domain of mGBP7. Another difference is that the G domain of mGBP7 does not stabilize so much upon GTP binding as in mGBP2, especially the guanine cap does not adopt a structured

conformation and remains flexible as in the apo-state. There are differences in the sequence of the mGBP7 guanine cap compared to hGBP1 and mGBP2. In particular, this motif is negatively charged in mGBP7, which would explain that binding of the negatively charged GTP has no stabilizing effect on it. This in turn further unfolds why none of the three techniques used for determining the mGBP7 dimer structures, i.e., XL-MS, SAXS, and protein–protein docking, predicted the dimerization to occur via the two guanine caps. It is either too dynamic to allow a stable mGBP7 dimer to be formed and/or the amino acid composition does not allow strong enough interactions between the two guanine caps to evolve. In hGBP1 and mGBP2, the stabilization of the guanine cap via GTP binding seems to be a prerequisite for dimer formation, either to reduce the entropy penalty for the dimerization or to have a well defined interaction surface. Since the guanine cap seems not to be involved in mGBP7 dimerization, this further explains why this protein can be a transient dimer in the cytosol without GTP.<sup>24</sup>

Five alternative mGBP7 dimer models were predicted instead, where the protein–protein interface is either between the G domain and  $\alpha$ 12/13 (model 1), between the M/E domains (models 2 and 5), the two M domains (model 3), or the two E domains (model 4). The two mGBP7 models derived from fitting of the SAXS data, models 4 and 5, turned out not to be stable. Model 5 adopted an asymmetric arrangement very quickly. Model 4 might become more stable in the presence of a lipid membrane where binding to the membrane would remove the flexible and this interaction-disturbing CT tail from the interface area.<sup>24</sup> However, this is a speculation for the time being. Based on the current results, our conclusion is that the mGBP7 dimer structures created by protein–protein docking with ClusPro are more sound. Common observations for these three dimer models are that the intramolecular salt bridges locking  $\alpha$ 4' and  $\alpha$ 12/13 never completely dissolved, due to a redundancy of four salt bridges of which at least one is always present with >50% occupancy. The G domain dimers (hGBP1, mGBP2, and model 1 for mGBP7) show an increase in the stability of the intramolecular salt bridge within the guanine cap, which coincides with a structural

stabilization of the guanine cap, while this salt bridge is not stabilized in the stalk dimers of mGBP7 (models 2 and 3). It should be noted though that the mGBP2 monomer already has a strong salt bridge in the guanine cap. A special observation for the mGBP2 dimer is that in one of the chains all salt bridges between  $\alpha 4'$  and  $\alpha 12$  completely dissolved, which on longer time scales could give rise to a folding out of the E domain, as was already observed for hGBP1 multimers.<sup>31,45,69</sup> All dimers are stabilized by intermolecular salt bridges, with one or two of them being of particular strength with interaction energies below  $-100$  kJ/mol, which are supported by several surrounding weaker interactions. The biological relevance of the different findings are discussed in the following sections.

### 3.2 Linking structural and mutation data to the current dimer models

**Dimer interface size** According to Bahadur et al., a biologically relevant dimer interface should comprise at least  $1,000 \text{ \AA}^2$ .<sup>70</sup> All six dimer models studied here fulfill this criterion. Negi et al.<sup>71</sup> had analyzed more than 70 protein complexes with known 3D structures and revealed that dimer contacts are often mediated by „hot spot“ residues, meaning that only very few amino acids significantly contribute towards the binding energy. This is also reflected in our GBP dimer models. The interaction area is largest for the hGBP1 G domain dimer, indicating that this might be the „strong initial interaction“ around which the polymer lattice on a lipid membrane is built.<sup>28</sup> Previous dimers reported for DSPs (some of which may only be due to crystal-packing) include the hGBP1 dimer by Prakash et al.,<sup>41</sup> burying  $2,890 \text{ \AA}^2$  of surface area for head-to-tail dimerization via G domain- $\alpha 10/12$  interactions, and  $2,140 \text{ \AA}^2$  for the head-to-head arrangement. In 2006, Ghosh et al.<sup>3</sup> described a hGBP1 dimer that has an even larger interface of  $3,900 \text{ \AA}^2$  and includes residues that are conserved in hGBP isoforms and the G domain motifs. The G domain dimer of OPA1, on the other hand, buries only a surface area of  $1,573 \text{ \AA}^2$ ;<sup>10</sup> however, as the authors did not provide the formula how the interface area was determined, it might be larger by a factor of two. The interfaces buried by

our dimers is within or even beyond that range, with values between 5,314 Å<sup>2</sup> (hGBP1) and 1,560 Å<sup>2</sup> (mGBP7, model 5). These interfaces thus fulfill the necessary size requirement for a stable dimer. The largest interface sizes were observed for the G domain dimers, which were also identified as the most stable dimers, further confirming the existence of this GBP dimerization mode.

**Salt bridges at the dimer interface** According to Praefcke et al.,<sup>47</sup> the residues 48–52 of hGBP1 are necessary for dimerization, especially R48 for building tetramers. Further residues involved, yet to a lesser degree, are 72, 76, 99, 103 106, 109 and 112.<sup>47</sup> We now understand the relevance of these residues for dimerization due their importance in GTP binding and subsequent stabilization of the G domain, in particular the guanine cap. Moreover, our simulations of mGBP2 also revealed direct engagement in dimerization for two of the residues in that region, namely for R48, which has a small, but noticeable dimerization energy contribution, and E102, which shows a prominent interaction to a number of residues. In a study of hGBP1, where the production of GMP was used as an indirect measure of dimerization, it was demonstrated that the  $\alpha 6$  helix plays a critical role in dimerization, especially the residues 289–308, as well as residues 103–108.<sup>72</sup> While the former region was not close enough to the dimer interface in any of the dimers studied here, we can confirm that the latter region has a high energetic contribution to dimer formation of both hGBP1 and mGBP2.

Vöpel et al.<sup>32</sup> observed that two positively charged residues of hGBP1, R227 and K228, located on  $\alpha 4'$ , as well as four glutamate residues, 556, 563, 568, and 575 located in  $\alpha 12/13$ , form intraprotein salt bridges and that loosening their interaction facilitates dimerization and tetramerization. This can either mean that (i)  $\alpha 4'$  and/or  $\alpha 12/13$  are part of the interface causing the intraprotein salt bridges to break, (ii) that they need to be broken in order to reveal the interface, or (iii) that dimerization could cause these salt bridges to break via an allosteric effect if  $\alpha 4'$  and  $\alpha 12/13$  should not be directly involved in the dimerization. The

effect (i) was observed in two of our mGBP7 dimer models, where these salt bridges loosened and either both  $\alpha 4'$  and  $\alpha 12$  (model 1) or only  $\alpha 12$  (model 4) contributes to dimer formation. In another study it was found that deleting the hGBP1 region after residue 481 abolishes tetramerization,<sup>32</sup> which is broadly reflected in the mGBP7 dimer models 1, 2, and 4, where the region 481–638 is involved in the dimer interface. In 2012, Wehner et al.<sup>54</sup> suggested that residues 105, 186, 245, and 259 have only minor effects on the dimerization of hGBP1, while residues 240 and 244 (i.e., the guanine cap) were identified to be essential for it. Our results for the head-to-head dimer of hGBP1 revealed R48, E105, K209, D237, D239, R242, K243, R245, K247, E251, K252, E256, E259, and E261 as main players in the dimerization, underlining the importance of the guanine cap. These residues have further in common that they are all charged, and under that aspect, R243 and K244 appear interchangeable. It is interesting to note that all of these residues are conserved between hGBP1 and mGBP2, and often also mGBP7. Although ClusPro did not propose a dimerization of mGBP7 that would correspond to the G domain dimer of hGBP1, we do not want to completely rule out its existence, also considering that AlphaFold-Multimer predicted this as the only mGBP7 dimer model.

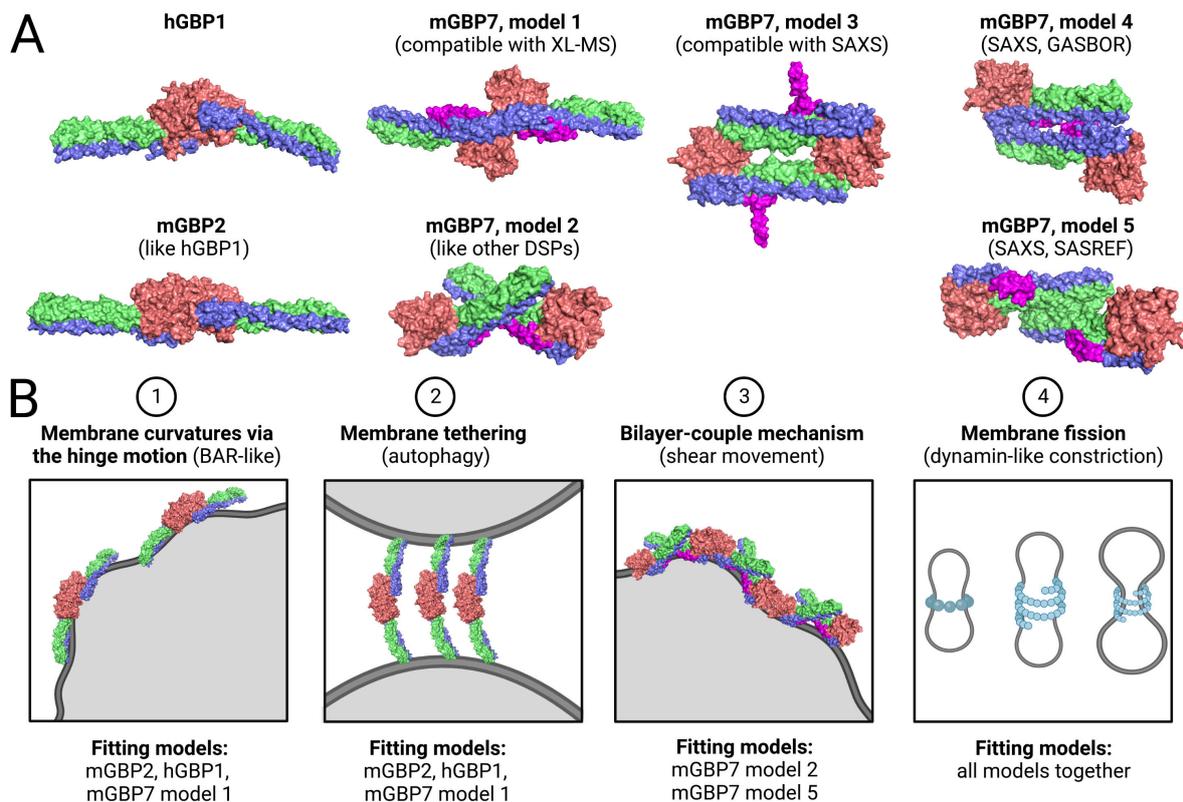
In general, we can confirm the importance of residues in the guanine cap (230–245) for dimerization and have indications for the relevance of  $\alpha 13$  (560–590) for tetramerization. The salt bridge 62–255 that enhances dimerization shows a slightly increased formation in the dimer, and the salt bridge 227–558, which supposedly inhibits tetramerization, shows a slightly decreased probability, especially in dimer models 2 and 4 of mGBP7. Energy analysis of interfacing residues suggests that the intramolecular salt bridges might be replaced by intermolecular interactions.

### 3.3 Biological relevance of different dimer structures

In order to understand why different GBP dimer structures may be adopted, it is instructive to recall their conformational dynamics. Upon dimerization, the hinge motions of the two

proteins forming the dimer give rise to an overall correlated motion. In the cases of the mGBP2 dimer and model 1 of mGBP7, the two chains do not show a greater individual range of motions than their monomeric counterparts, but the extended length of the two proteins moving in a concerted fashion could introduce more stress on a membrane than a single GBP protein is able to do, comparable to the observations made for BAR (Bin/Amphiphysin/Rvs) domain proteins.<sup>73–75</sup> In contrast to that, the motions of dimer model 2 of mGBP7 do not lead to an increased overall curvature, so that a possible mechanism of introducing membrane stress/shearing force would require both the CT tails and the stalk tips to be inserted into the membrane. The insertion of a single (angled or short) helix can result in a hydrophobic mismatch that facilitates membrane fission,<sup>76</sup> comparable to a wedge driven into the lipid bilayer. The dimer models 3 and 4 of mGBP7 involve a more flexible stalk tip than in the monomeric mGBP7 form, but the protein–protein interaction along the length of the M/E domains constricts the overall dynamic effect.

While these three notable structures seem initially to be at odds with each other, several studies of other DSPs are in support of the existence of different DSP multimer arrangements based on up to four intermolecular interfaces.<sup>75,77</sup> In the case of OPA1, dimerization can be both nucleotide-dependent and -independent, involve multiple interfaces apart from the G domain, and lead to higher order oligomers.<sup>10</sup> It is possible that the structures spatially complement each other to form a lattice, a multimer consisting of several different dimers. On the other hand, the dimers may also temporally complement each other during the GTPase cycle, with rearrangements triggered by nucleotide binding or hydrolysis. Moreover, some dimer structures may not be functional and only be present during the aggregated/storage phase.



**Figure 6: Summary of dimer structures:** (A) Seven dimer models of hGBP1, mGBP2, and mGBP7 were studied in this work. (B) Different dimerization modes suggest various mechanisms of action. A curving of the dimer could induce membrane curvature by the scaffold mechanism, similar to BAR domain proteins or the COP (cytoplasmic coat proteins for vesicle transport) machinery. The G domain dimer of hGBP1, mGBP2, or model 1 of mGBP7 could serve for tethering the parasitophorous vacuole membrane to the autophagosome. Insertion of tilted helices, which would be possible with mGBP7 dimer models 2 and 5, would give rise to the so-called local spontaneous curvature mechanism or bilayer-couple mechanism<sup>73</sup> and ruffle the membrane or generate shearing forces. Finally, by combining several of these dimerization modes, even a dynamamin-like constriction would be imaginable. This figure was created with BioRender.

If we translate the structural information that is available for DSP multimers to the GBPs studied here, we conclude that two of the five mGBP7 dimer models (models 1 and 5) can be combined to form a long string of half-moon shapes, comparable to the BAR domain proteins (Fig. 6). The primary, most stable interaction would happen via the G domains, with the underside of the E domains forming the second interface. The alternating rise and fall of

the proteins could induce local curvature of the membrane, especially when considering the hinge motion of the M/E domains. Adding a third model (model 3 or 4), the dimers can be integrated into a helical-ring-model, reminiscent of other dynamins. While the first turn of the helix consists of G and E domain interactions, the remaining M domain would form the inter-rung contacts. A conformational change of the M/E domain would be needed to „unlock“ the fourth interface, the crossed-stalks dimer. This further requires that the G domain interface is flexible enough that a tilting of the spherical G domains against each other is still stable. The docking results indicate that such non-standard G domain interfaces are possible. Such a ratchet-like tightening of the dynamin collar has been previously suggested.<sup>2</sup>

Furthermore, with more than one interface, the dissolution of less stable interactions could be prevented by a lattice of surrounding interactions, and additional binding of the CT tails to the membrane. The small sequence variations between hGBP1, mGBP2 and mGBP7 seem indeed to be sufficient to tip the scale in favor of other dimer structures. It might thus be possible that some of the isoforms act as chain-starters (with maybe only one stable interface), some as continuous elongators, and some as chain-breakers (with no further interfaces), a scenario in which heterooligomers of seemingly redundant proteins would be necessary. For bacterial DSPs, this has already been proposed.<sup>33</sup> We further know that mGBP2 and mGBP7 co-localize at the parasitophorous vacuole membrane and that mGBP2 attaches first, followed by mGBP7, and that mGBP7 deletion is more lethal than mGBP2 deletion.<sup>17</sup> Perhaps mGBP7's various polymerization options allow a larger carpet of GBPs to form, which can then more effectively remodel membranes, something that should be addressed by future studies.

## 4 Methods and materials

All details about the experimental methods are provided in the Supplementary Information.

### 4.1 MD simulations and docking

**General aspects of the MD simulations** The structures of hGBP1, mGBP2, mGBP7 were prepared for the simulations as previously described, which includes the parametrization of GTP and the geranylgeranyl group.<sup>22-24</sup> The hGBP1 dimer was created using MM-align (<https://zhanglab.ccmb.med.umich.edu/MM-align/><sup>58</sup>). For all MD simulations, the Amber99SB\*-ILDNP force field<sup>78-80</sup> was used for modeling the proteins, the TIP3P model was used for water,<sup>81</sup> and GROMACS 2016<sup>82,83</sup> was employed for running the simulations. The simulation temperature and pressure were 310 K (37 °C) and 1 bar, respectively. In Table 2, all MD simulations included in this study are listed.

**HREMD simulation of the mGBP7 monomer** In our previous study of mGBP7,<sup>24</sup> we already obtained the structure of mGBP7 by using homology modeling and simulated it as mGBP7<sub>GTP</sub> for 100 ns. The most populated cluster structure of that simulation was used here as starting point for the HREMD simulations. To identify the role of GTP and the CT tail, two further systems, mGBP7<sub>GTP</sub> and mGBP7<sub>holo</sub> were created, using an identical approach as described on our work of mGBP2.<sup>23</sup> The protein was always placed in a rectangle simulation box of 10 nm × 10 nm × 18.0 nm dimensions, ~56,000 water molecules were added, as well as 10 (apo), 13 (GTP), or 13 Na<sup>+</sup> (holo) for the neutralization of the systems, resulting in a total number of ~178,000 atoms. A similar HREMD protocol as applied in our study of mGBP2 dynamics was used and is described in detail in the Supplementary Information.

**Table 2: Summary of simulations included in this work.<sup>a</sup>**

Simulation	System	Size in atoms	Runs	Length	Cumulated time
HREMD <sup>b</sup>	hGBP1 <sub>apo</sub>	335,553	1 × 30	400 ns	12 $\mu$ s
HREMD <sup>c</sup>	mGBP2 <sub>apo</sub>	131,493	1 × 16	200 ns	3.2 $\mu$ s
HREMD <sup>c</sup>	mGBP2 <sub>GTP</sub>	129,116	1 × 16	200 ns	3.2 $\mu$ s
HREMD	mGBP7 <sub>apo</sub>	178,800	1 × 30	400 ns	12 $\mu$ s
HREMD	mGBP7 <sub>GTP</sub>	178,019	1 × 30	400 ns	12 $\mu$ s
MD	hGBP1 dimer	556,008	1	100 ns	100 ns
MD	mGBP2 dimer	556,729	1	100 ns	100 ns
MD	mGBP7 dimer, model 1	538,254	1	100 ns	100 ns
MD	mGBP7 dimer, model 2	308,874	1	100 ns	100 ns
MD	mGBP7 dimer, model 3	308,763	1	100 ns	100 ns
MD	mGBP7 dimer, model 4	308,514	1	100 ns	100 ns
MD	mGBP7 dimer, model 5	308,763	1	100 ns	100 ns
<b>Total simulation time</b>					<b>36.7 <math>\mu</math>s</b>

<sup>a</sup> The simulations reported in this work were either performed in the supercomputer JURECA at the Jülich Supercomputing Centre,<sup>84</sup> in SuperMUC-NG at the Leibniz-Rechenzentrum (LRZ) in Munich, or in HILBERT at Heinrich Heine University Düsseldorf.

<sup>b</sup> This trajectory originates from reference<sup>22</sup>.

<sup>c</sup> This trajectory originates from reference<sup>23</sup>.

**Protein–protein docking** To obtain GBP dimer structures, we performed protein–protein docking and employed ClusPro (<https://cluspro.bu.edu>)<sup>56</sup> for this purpose, taking favorable assessments of this program into account.<sup>85,86</sup> The multimer docking mode of ClusPro was used, both in combination with SAXS data and/or crosslinking data, and without additional restraints. The inclusion of additional restraints did not result in new structures, only a reordering of the results happened. From the top 10 results of the different setups, the most promising candidates were selected for MD simulations.

**MD simulations of the GBP dimers** The dimers were simulated in boxes with sizes, depending on the structure, varying from 27.0 nm × 15.5 nm × 13.5 nm to 18.5 nm × 12.5 nm × 13.5 nm, with water added for solvation and NaCl added for neutralization. A two-step equilibration was performed as described for the HREMD simulation, yet without an additional  $NpT$  equilibration. The hGBP1 and mGBP2 dimers were simulated with GTP being bound, using a 2 fs time step, while the GTP-free mGBP7 dimers were simulated using a 4 fs time step. The reason for simulating mGBP7 in its apo-form is two-fold: firstly, it was shown by Legewie et al.<sup>24</sup> that mGBP7 can form dimers in the absence of GTP; and secondly, without GTP, virtual sites can be applied enabling a larger time step. In all simulations, distance restraints of 1,000 kJ mol<sup>-1</sup> nm<sup>-2</sup> were applied between the  $\beta$ -sheets of the G domains, in order to inhibit overall rotation, allowing us to keep the box size smaller. The Parrinello-Rahman barostat and Nosé-Hoover thermostat were used. Each dimer structure was simulated for 100 ns, where frames were saved every 20 ps and analyzed. In order to assess the dimer-dimer interactions, another 100 ns MD trajectory was produced per dimer, where additional energy groups had been added to the GROMACS index file. For this, the first 100 ns trajectory was analyzed, and all residues within 7.5 Å of the other chain were included. This allowed us to identify the most strongly interacting residues in the second trajectory, which were energetically analyzed in detail.

## 4.2 Analysis of the MD simulations

To generate figures of the 3D protein structures, we applied PyMol.<sup>87</sup> If not stated otherwise, the analyses were performed using GROMACS 2016. For the analysis of the HREMD simulations, only the target replica was used.

**Sequence alignment** The alignment of the sequences of hGBP1, mGBP2, and mGBP7 was done with T-Coffee<sup>51</sup> using the default settings as available at the European Bioinformatics Institute (EMBL-EBI) webserver<sup>52</sup> (<https://www.ebi.ac.uk/services>). The resulting

sequence alignment was visualized using Jalview 2<sup>88</sup> where we colored the residues according to their residue type.

**Root mean square fluctuations (RMSF)** For quantifying the flexibility of the GBPs, we calculated the RMSF of the C<sub>α</sub> atoms around their average positions for each residue. A residue with a value over 2 Å is considered as flexible. For visualization, the RMSF values were color-mapped and projected onto the corresponding start structure of the simulations using red colors for flexible regions (RMSF  $\gtrsim$  2.5 Å) and blue colors for rigid parts (RMSF  $\lesssim$  1.5 Å).

**Electrostatic potential surface (EPS)** The Adaptive Poisson-Boltzmann Solver as available as plugin APBS 2.1 in PyMol was used for the calculation of the electrostatic potential.<sup>89</sup> To generate the input files (\*.pqr and \*.pot) needed for this, the web server <https://server.poissonboltzmann.org> with default settings and with pH 7 for the  $pK_a$  calculation was employed.<sup>90</sup> For visualization, the electrostatic potential was color-mapped between -5 (red) and +5 kT/e (blue) and projected onto the surface presentation of the proteins.

**Movements of residue 480** To quantify and characterize the hinge motion of the M/E domain, the motions of residue 480 were monitored relative to the start structure of the simulations. For this, the changes in the Cartesian coordinates ( $\Delta x$ ,  $\Delta y$ ,  $\Delta z$ ) as well as the absolute distance ( $d_{480}$ ) of that residue were determined using our own VMD<sup>91</sup> script. To do this, we fitted the trajectory frames to the G domain of the start structure. To illustrate the hinge motion, we computed the spatial distribution of residue 480 using the *gmx spatial* tool of GROMACS.

**Principal component analysis (PCA)** To extract the main motions of the proteins, principal component analyses as implemented in GROMACS were performed. The covariance matrix was created with *gmx covar*, and the first two eigenvectors, also called principal compo-

nents or principal motions (PC1 and PC2), were determined using *gmx anaeig*. The trajectory was fitted to the C<sub>α</sub> atoms, and for the generation of conformations along eigenvectors 1 and 2, the backbone was selected.

**Salt bridge analysis** To determine the occupancies of selected salt bridges, the distances between the atoms involved were determined. Following atoms of charged residues were used for the calculations: Lys – NZ; Arg – NE, NH1, NH2; Asp – OD1, OD2; Glu – OE1, OE2. A salt bridge was assumed to be present if the distance in question was within 4.5 Å. Salt bridge occupancies were calculated as time-averaged probabilities.

**Clustering analysis.** To characterize the flexibilities of the G domains, we performed clustering analyses for the G motifs and loops using the algorithm of Daura et al.<sup>92</sup> as implemented in GROMACS. The clustering was applied to all atoms of the desired structural element with a C<sub>α</sub>-RMSD cutoff of 2.5 Å to identify cluster membership. Before the clustering, the trajectory was fitted onto the β-sheets in the G domain.

**Protein–protein interface size** The size of the protein-protein interfaces was calculated using the difference in the solvent-accessible surface areas (SASA):

$$\text{SASA}_{\text{interface}} = (\text{SASA}_{\text{chain 1}} + \text{SASA}_{\text{chain 2}}) - \text{SASA}_{\text{dimer}}$$

For the SASA calculation, the GROMACS tool *gmx sasa* was used, with a probe radius of 1.4 Å.

**Analysis of dimer interaction energies** The energy groups of interacting residues as determined in the 100 ns MD simulations of the GBP dimers were added to the GROMACS index files and a second 100 ns MD simulation per dimer was performed. The resulting residue–residue energies were presented as a matrix, and the most negative energies were further analyzed by decomposing them into Coulomb and Lennard-Jones short-ranged interaction energies as calculated by GROMACS.

## 5 Funding and Acknowledgments

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)-project number 267205415-CRC 1208 project A07 (J.L., W.S, B.S) as well as and Z01 (K.S.). J.L., W.S and B.S gratefully acknowledge the computing time granted through JARA-HPC (project JICS6A) on the supercomputer JURECA at the Forschungszentrum Juelich.<sup>84</sup> The Center for Structural studies is funded by the DFG (Grant number 417919780 and INST 208/761-1 FUGG to S.S.). The authors gratefully acknowledge the Gauss Centre for Supercomputing e.V. ([www.gauss-centre.eu](http://www.gauss-centre.eu)) for funding this project by providing computing time on the GCS Supercomputer SuperMUC-NG at Leibniz Supercomputing Centre ([www.lrz.de](http://www.lrz.de)). W.S. gratefully acknowledges funding received from Jürgen Manchot Promotionsstipendium. Additional computational infrastructure and support were provided by the Centre for Information and Media Technology at Heinrich Heine University Düsseldorf.

## 6 Data availability

We uploaded the SAXS data to the Small Angle Scattering Biological Data Bank (SASBDB),<sup>93,94</sup> with the accession code SASDLY9. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE<sup>95</sup> partner repository with the dataset identifier PXD026979.

## 7 Author Contribution Statement (draft)

W.S. performed, analyzed and wrote the GBP dimer part. J.L. performed, analyzed and wrote the GBP monomer part. G.P. performed, analyzed and wrote the XL-MS part. J.R. and S.S. performed, analyzed and wrote the SAXS part. D.D. and K.P. provided infrastructure, supervision and protein material. K.S. provided infrastructure and supervision. B.S. conceived the research project and provided infrastructure and extensive supervision.

## References

- (1) Morlot, S.; Galli, V.; Klein, M.; Chiaruttini, N.; Manzi, J.; Humbert, F.; Dinis, L.; Lenz, M.; Cappello, G.; Roux, A. Membrane Shape at the Edge of the Dynamin Helix Sets Location and Duration of the Fission Reaction. *Cell* **2012**, *151*, 619.
- (2) Ford, M. G. J.; Chappie, J. S. The structural biology of the dynamin-related proteins: New insights into a diverse, multitasking family. *Traffic* **2019**, *20*, 717–740.
- (3) Ghosh, A.; Praefcke, G. J. K.; Renault, L.; Wittinghofer, A.; Herrmann, C. How guanylate-binding proteins achieve assembly-stimulated processive cleavage of GTP to GMP. *Nature* **2006**, *440*, 101–104.
- (4) Narayanan, R.; Leonard, M.; Song, B. D.; Schmid, S. L.; Ramaswami, M. An internal GAP domain negatively regulates presynaptic dynamin in vivo: a two-step model for dynamin function. *J. Cell Biol.* **2005**, *169*, 117.
- (5) Praefcke, G. J. K.; McMahon, H. T. The dynamin superfamily: universal membrane tubulation and fission molecules? *Nat. Rev. Mol. Cell Biol.* **2004**, *5*, 133–147.
- (6) Gao, S.; Hu, J. Mitochondrial Fusion: The Machineries In and Out. *Trends Cell Biol.* **2021**, *31*, 62–74.
- (7) Stowell, M. H.; Marks, B.; Wigge, P.; McMahon, H. T. Nucleotide-dependent conformational changes in dynamin: evidence for a mechanochemical molecular spring. *Nature cell biology* **1999**, *1*, 27–32.
- (8) Warnock, D. E.; Hinshaw, J. E.; Schmid, S. L. Dynamin Self-assembly Stimulates Its GTPase Activity \*. *Journal of Biological Chemistry* **1996**, *271*, 22310–22314.
- (9) Chappie, J. S.; Acharya, S.; Liu, Y.-W.; Leonard, M.; Pucadyil, T. J.; Schmid, S. L. An Intramolecular Signaling Element that Modulates Dynamin Function In Vitro and In Vivo. *Mol. Biol. Cell.* **2009**, *20*, 3561.

- (10) Yu, C.; Zhao, J.; Yan, L.; Qi, Y.; Guo, X.; Lou, Z.; Hu, J.; Rao, Z. Structural insights into G domain dimerization and pathogenic mutation of OPA1. *J. Cell Biol.* **2020**, *219*.
- (11) Fröhlich, C.; Grabiger, S.; Schwefel, D.; Faelber, K.; Rosenbaum, E.; Mears, J.; Rocks, O.; Daumke, O. Structural insights into oligomerization and mitochondrial remodelling of dynamin 1-like protein. *The EMBO J.* **2013**, *32*, 1280–1292.
- (12) Byrnes, L. J.; Sondermann, H. Structural basis for the nucleotide-dependent dimerization of the large G protein atlastin-1/SPG3A. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 2216–2221.
- (13) Low, H. H.; Sachse, C.; Amos, L. A.; Löwe, J. Structure of a Bacterial Dynamin-like Protein Lipid Tube Provides a Mechanism For Assembly and Membrane Curving. *Cell* **2009**, *139*, 1342–1352.
- (14) Taylor, G. A.; Feng, C. G.; Sher, A. p47 GTPases: regulators of immunity to intracellular pathogens. *Nat. Rev. Immunology* **2004**, *4*, 100–109.
- (15) Degrandi, D.; Konermann, C.; Beuter-Gunia, C.; Kresse, A.; Würthner, J.; Kurig, S.; Beer, S.; Pfeffer, K. Extensive Characterization of IFN-Induced GTPases mGBP1 to mGBP10 Involved in Host Defense. *J. Immunol.* **2007**, *179*, 7729–7740.
- (16) Degrandi, D.; Kravets, E.; Konermann, C.; Beuter-Gunia, C.; Klümpers, V.; Lahme, S.; Wischmann, E.; Mausberg, A. K.; Beer-Hammer, S.; Pfeffer, K. Murine Guanylate Binding Protein 2 (mGBP2) controls *Toxoplasma gondii* replication. *PNAS* **2013**, *110*, 294–299.
- (17) Steffens, N.; Beuter-Gunia, C.; Kravets, E.; Reich, A.; Legewie, L.; Pfeffer, K.; Degrandi, D. Essential Role of mGBP7 for Survival of *Toxoplasma gondii* Infection. *mBio* **2020**, *11*.

- (18) Krapp, C.; Hotter, D.; Gawanbacht, A.; McLaren, P. J.; Kluge, S. F.; Stürzel, C. M.; Mack, K.; Reith, E.; Engelhart, S.; Ciuffi, A.; Hornung, V.; Sauter, D.; Telenti, A.; Kirchhoff, F. Guanylate Binding Protein (GBP) 5 Is an Interferon-Inducible Inhibitor of HIV-1 Infectivity. *Cell Host Microbe* **2016**, *19*, 504–514.
- (19) Xavier, A.; Al-Zeer, M. A.; Meyer, T. F.; Daumke, O. hGBP1 Coordinates Chlamydia Restriction and Inflammasome Activation through Sequential GTP Hydrolysis. *Cell Rep.* **2020**, *31*, 107667.
- (20) Lindenberg, V.; Mölleken, K.; Kravets, E.; Stallmann, S.; Hegemann, J. H.; Degrandi, D.; Pfeffer, K. Broad recruitment of mGBP family members to Chlamydia trachomatis inclusions. *PloS one* **2017**, *12*.
- (21) Olszewski, M. A.; Gray, J.; Vestal, D. J. In silico genomic analysis of the human and murine guanylate-binding protein (gbp) gene clusters. *J. Interferon Cytokine Res.* **2006**, *26*, 328–352.
- (22) Barz, B.; Loschwitz, J.; Strodel, B. Large-scale, dynamin-like motions of the human guanylate binding protein 1 revealed by multi-resolution simulations. *PLoS Comput. Biol.* **2019**, *15*, 1–29.
- (23) Loschwitz, J.; Steffens, N.; Wang, X.; Schäffler, M.; Pfeffer, K.; Degrandi, D.; Strodel, B. Domain motions, dimerization, and membrane interactions of the murine guanylate binding protein 2. *bioRxiv* **2022**,
- (24) Legewie, L.; Loschwitz, J.; Steffens, N.; Prescher, M.; Wang, X.; Smits, S. H.; Schmitt, L.; Strodel, B.; Degrandi, D.; Pfeffer, K. Biochemical and structural characterization of murine GBP7, a guanylate binding protein with an elongated C-terminal tail. *Biochem. J.* **2019**, *476*, 3161–3182.
- (25) Kravets, E.; Degrandi, D.; Ma, Q.; Peulen, T.-O.; Klümpers, V.; Felekyan, S.; Kühne-

- muth, R.; Weidtkamp-Peters, S.; Seidel, C. A.; Pfeffer, K. Guanylate binding proteins directly attack *Toxoplasma gondii* via supramolecular complexes. *eLife* **2016**, *5*, e11479.
- (26) Faelber, K.; Posor, Y.; Gao, S.; Held, M.; Roske, Y.; Schulze, D.; Haucke, V.; Noé, F.; Daumke, O. Crystal structure of nucleotide-free dynamin. *Nature* **2011**, *477*, 556–560.
- (27) Rennie, M. L.; McKelvie, S. A.; Bulloch, E. M.; Kingston, R. L. Transient dimerization of human MxA promotes GTP hydrolysis, resulting in a mechanical power stroke. *Structure* **2014**, *22*, 1433–1445.
- (28) Daumke, O.; Praefcke, G. J. K. Invited review: Mechanisms of GTP hydrolysis and conformational transitions in the dynamin superfamily. *Biopolymers* **2016**, *105*, 580–593.
- (29) Britzen-Laurent, N.; Bauer, M.; Berton, V.; Fischer, N.; Syguda, A.; Reipschläger, S.; Naschberger, E.; Herrmann, C.; Stürzl, M. Intracellular Trafficking of Guanylate-Binding Proteins Is Regulated by Heterodimerization in a Hierarchical Manner. *PLoS ONE* **2010**, *5*, e14246.
- (30) Ince, S.; Kutsch, M.; Shydlovskiy, S.; Herrmann, C. The human guanylate-binding proteins hGBP-1 and hGBP-5 cycle between monomers and dimers only. *FEBS J.* **2017**, *284*, 2284–2301.
- (31) Sistemich, L.; Kutsch, M.; Hämisch, B.; Zhang, P.; Shydlovskiy, S.; Britzen-Laurent, N.; Stürzl, M.; Huber, K.; Herrmann, C. The Molecular Mechanism of Polymer Formation of Farnesylated Human Guanylate-binding Protein 1. *J. Mol. Biol.* **2020**, *432*, 2164–2185.
- (32) Vöpel, T.; Syguda, A.; Britzen-Laurent, N.; Kunzelmann, S.; Lüdemann, M.-B.; Dovengerds, C.; Stürzl, M.; Herrmann, C. Mechanism of GTPase-Activity-Induced Self-Assembly of Human Guanylate Binding Protein 1. *J. Mol. Biol.* **2010**, *400*, 63–70.
- (33) Liu, J.; Noel, J. K.; Low, H. H. Structural basis for membrane tethering by a bacterial dynamin-like pair. *Nat. Commun.* **2018**, *9*.

- (34) Michie, K. A.; Boysen, A.; Low, H. H.; Møller-Jensen, J.; Löwe, J. LeoA, B and C from Enterotoxigenic *Escherichia coli* (ETEC) Are Bacterial Dynamins. *PLoS ONE* **2014**, *9*, e107211.
- (35) DeVay, R. M.; Dominguez-Ramirez, L.; Lackner, L. L.; Hoppins, S.; Stahlberg, H.; Nunari, J. Coassembly of Mgm1 isoforms requires cardiolipin and mediates mitochondrial inner membrane fusion. *J. Cell Biol.* **2009**, *186*, 793–803.
- (36) Kresse, A.; Konermann, C.; Degrandi, D.; Beuter-Gunia, C.; Wuerthner, J.; Pfeffer, K.; Beer, S. Analyses of murine GBP homology clusters based on in silico, in vitro and in vivo studies. *BMC Genomics* **2008**, *9*.
- (37) Jimah, J. R.; Hinshaw, J. E. Structural Insights into the Mechanism of Dynamin Superfamily Proteins. *Trends Cell Biol.* **2019**, *29*, 257–273.
- (38) Cui, W. et al. Structural basis for GTP-induced dimerization and antiviral function of guanylate-binding proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2021**, *118*, e2022269118.
- (39) Bian, X.; Klemm, R. W.; Liu, T. Y.; Zhang, M.; Sun, S.; Sui, X.; Liu, X.; Rapoport, T. A.; Hu, J. Structures of the atlastin GTPase provide insight into homotypic fusion of endoplasmic reticulum membranes. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 3976–3981.
- (40) Byrnes, L. J.; Sondermann, H. Structural basis for the nucleotide-dependent dimerization of the large G protein atlastin-1/SPG3A. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 2216–2221.
- (41) Prakash, B.; Praefcke, G.; Renault, L.; Wittinghofer, A.; Herrmann, C. Structure of human guanylate-binding protein 1 representing a unique class of GTP-binding proteins. *Nature* **2000**, *403*, 567.
- (42) Hofmann, K. TMbase-A database of membrane spanning proteins segments. *Biol. Chem. Hoppe-Seyler* **1993**, *374*, 166.

- (43) Shydlovskiy, S.; Zienert, A. Y.; Ince, S.; Dovengerds, C.; Hohendahl, A.; Dargazanli, J. M.; Blum, A.; Günther, S. D.; Kladt, N.; Stürzl, M.; Schauss, A. C.; Kutsch, M.; Roux, A.; Praefcke, G. J. K.; Herrmann, C. Nucleotide-dependent farnesyl switch orchestrates polymerization and membrane binding of human guanylate-binding protein 1. *Proc. Natl. Acad. Sci. U.S.A.* **2017**, *114*, E5559–E5568.
- (44) Ince, S.; Zhang, P.; Kutsch, M.; Krenczyk, O.; Shydlovskiy, S.; Herrmann, C. Catalytic activity of human Guanylate-Binding Protein 1 coupled to the release of structural restraints imposed by the C-terminal domain. *FEBS J.* **2020**, febs.15348.
- (45) Sistemich, L.; Dimitrov Stanchev, L.; Kutsch, M.; Roux, A.; Günther Pomorski, T.; Herrmann, C. Structural requirements for membrane binding of human guanylate-binding protein 1. *FEBS J.* **2021**, *288*, 4098–4114.
- (46) Syguda, A.; Kerstan, A.; Ladnorg, T.; Stüben, F.; Wöll, C.; Herrmann, C. Immobilization of Biotinylated hGBP1 in a Defined Orientation on Surfaces Is Crucial for Uniform Interaction with Analyte Proteins and Catalytic Activity. *Langmuir* **2012**, *28*, 6411–6418.
- (47) Praefcke, G.; Kloep, S.; Benschaid, U.; Lilie, H.; Prakash, B.; Herrmann, C. Identification of Residues in the Human Guanylate-binding Protein 1 Critical for Nucleotide Binding and Cooperative GTP Hydrolysis. *J. Mol. Biol.* **2004**, *344*, 257–269.
- (48) Kravets, E.; Degrandi, D.; Weidtkamp-Peters, S.; Ries, B.; Konermann, C.; Felekyan, S.; Dargazanli, J.; Praefcke, G.; Seidel, C.; Schmitt, L.; Smits, S.; Pfeffer, K. The GTPase Activity of Murine Guanylate-binding Protein 2 (mGBP2) Controls the Intracellular Localization and Recruitment to the Parasitophorous Vacuole of *Toxoplasma gondii*\*. *J. Biol. Chem.* **2012**, *287*, 27452–27466.
- (49) Schwemmle, M.; Staeheli, P. The interferon-induced 67-kDa guanylate-binding protein

- (hGBP1) is a GTPase that converts GTP to GMP. *J. Biol. Chem.* **1994**, *269*, 11299–11305.
- (50) Praefcke, G. J.; Geyer, M.; Schwemmler, M.; Robert Kalbitzer, H.; Herrmann, C. Nucleotide-binding characteristics of human guanylate-binding protein 1 (hGBP1) and identification of the third GTP-binding motif. *Journal of Molecular Biology* **1999**, *292*, 321–332.
- (51) Notredame, C.; Desmond, G.; Jaap, H. T-coffee: a novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **2000**, *302*, 205–217.
- (52) Madeira, F.; Park, Y. m.; Lee, J.; Buso, N.; Gur, T.; Madhusoodanan, N.; Basutkar, P.; Tivey, A.; Potter, S.; Finn, R.; Lopez, R. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* **2019**, *47*, W636–W641.
- (53) Zhu, Z.; Shi, Z.; Yan, W.; Wei, J.; Shao, D.; Deng, X.; Wang, S.; Li, B.; Tong, G.; Ma, Z. Nonstructural Protein 1 of Influenza A Virus Interacts with Human Guanylate-Binding Protein 1 to Antagonize Antiviral Activity. *PLoS ONE* **2013**, *8*, 1–11.
- (54) Wehner, M.; Kunzelmann, S.; Herrmann, C. The guanine cap of human guanylate-binding protein 1 is responsible for dimerization and self-activation of GTP hydrolysis. *FEBS J.* **2012**, *279*, 203–210.
- (55) Zhu, S.; Bradfield, C. J.; Mamińska, A.; Park, E.-S.; Kim, B.-H.; Kumar, P.; Huang, S.; Zhang, Y.; Bewersdorf, J.; MacMicking, J. D. Cryo-ET of a human GBP coatomer governing cell-autonomous innate immunity to infection. *bioRxiv* **2021**,
- (56) Kozakov, D.; Hall, D.; Xia, B.; Porter, K.; Olson, A.; Padhorny, D.; Yueh, C.; Beglov, D.; Vajda, S. The ClusPro web server for protein–protein docking. *Nat. Protoc.* **2017**, *12*, 255–278.

- (57) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.
- (58) Mukherjee, S.; Zhang, Y. MM-align: a quick algorithm for aligning multiple-chain protein complex structures using iterative dynamic programming. *Nucleic Acids Res.* **2009**, *37*, e83–e83.
- (59) Evans, R. et al. Protein complex prediction with AlphaFold-Multimer. *bioRxiv* **2021**, 10.1101/2021.10.04.463034v1.
- (60) Basu, S.; Wallner, B. DockQ: A Quality Measure for Protein-Protein Docking Models. *PLoS ONE* **2016**, *11*, 1–9.
- (61) Combe, C. W.; Fischer, L.; Rappsilber, J. xiNET: cross-link network maps with residue resolution. *Mol. Cell Proteomics* **2015**, *14*, 1137–1147.
- (62) Svergun, D. I.; Petoukhov, M. V.; Koch, M. H. Determination of Domain Structure of Proteins from X-Ray Solution Scattering. *Biophys. J.* **2001**, *80*, 2946–2953.
- (63) Petoukhov, M. V.; Svergun, D. I. Global Rigid Body Modeling of Macromolecular Complexes against Small-Angle Scattering Data. *Biophys. J.* **2005**, *89*, 1237–1250.
- (64) Svergun, D.; Barberato, C.; Koch, M. H. J. *CRY SOL* – a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *J. Appl. Crystallogr.* **1995**, *28*, 768–773.
- (65) Guinier, A. Studies on transformation of *Escherichia coli* with plasmids. *Ann. Phys.* **1939**, *12*, 161–237.
- (66) Noel, J.; Noé, F.; Daumke, O.; Mikhailov, A. Polymer-like Model to Study the Dynamics of Dynamin Filaments on Deformable Membrane Tubes. *Biophys. J.* **2019**, *117*, 1870–1891.

- (67) Reubold, T.; Faelber, K.; Plattner, N.; Posor, Y.; Ketel, K.; Curth, U.; Schlegel, J.; Anand, R.; Manstein, D.; Noé, F.; Haucke, V.; Daumke, O.; Eschenburg, S. Crystal structure of the dynamin tetramer. *Nature* **2015**, *525*, 404–408.
- (68) van Zundert, G.; Bonvin, A. DisVis: quantifying and visualizing accessible interaction space of distance-restrained biomolecular complexes. *Bioinformatics* **2015**, *31*, 3222–3224.
- (69) Lorenz, C.; Ince, S.; Zhang, T.; Cousin, A.; Batra-Safferling, R.; Nagel-Steger, L.; Herrmann, C.; Stadler, A. M. Farnesylation of human guanylate-binding protein 1 as safety mechanism preventing structural rearrangements and uninduced dimerization. *FEBS J.* **2020**, *287*, 496–514.
- (70) Bahadur, R. P.; Zacharias, M. The interface of protein-protein complexes: Analysis of contacts and prediction of interactions. *Cell. Mol. Life Sci.* **2008**, *65*, 1059–1072.
- (71) Negi, S. S.; Braun, W. Statistical analysis of physical-chemical properties and prediction of protein-protein interfaces. *J. Mol. Model.* **2007**, *13*, 1157–1167.
- (72) Abdullah, N.; Balakumari, M.; Sau, A. Dimerization and Its Role in GMP Formation by Human Guanylate Binding Proteins. *Biophys. J.* **2010**, *99*, 2235–2244.
- (73) McMahon, H. T.; Gallop, J. L. Membrane curvature and mechanisms of dynamic cell membrane remodelling. *Nature* **2005**, *438*, 590–596.
- (74) Zimmerberg, J.; Kozlov, M. M. How proteins produce cellular membrane curvature. *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 9–19.
- (75) Daumke, O.; Roux, A.; Haucke, V. BAR Domain Scaffolds in Dynamin-Mediated Membrane Fission. *Cell* **2014**, *156*, 882–892.
- (76) Fattal, D. R.; Ben-Shaul, A. A molecular model for lipid-protein interaction in membranes: the role of hydrophobic mismatch. *Biophys. J.* **1993**, *65*, 1795–1809.

- (77) Chen, Y.; Zhang, L.; Graf, L.; Yu, B.; Liu, Y.; Kochs, G.; Zhao, Y.; Gao, S. Conformational dynamics of dynamin-like MxA revealed by single-molecule FRET. *Nat. Commun.* **2017**, *8*, 15744.
- (78) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **2010**, *78*, 1950–1958.
- (79) Best, R. B.; Hummer, G. Optimized Molecular Dynamics Force Fields Applied to the Helix-Coil Transition of Polypeptides. *J. Phys. Chem. B* **2009**, *113*, 9004–9015.
- (80) Aliev, A. E.; Kulke, M.; Khaneja, H. S.; Chudasama, V.; Sheppard, T. D.; Lanigan, R. M. Motional timescale predictions by molecular dynamics simulations: Case study using proline and hydroxyproline sidechain dynamics. *Proteins* **2014**, *82*, 195–215.
- (81) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Theory Comput.* **1983**, *79*, 926–935.
- (82) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**,
- (83) Abraham, M. J.; van der Spoel, D.; Lindahl, E.; Hess, B. the GROMACS development team GROMACS User Manual Version 2016.4. **2017**,
- (84) Krause, D.; Thörnig, P. JURECA: Modular supercomputer at Jülich Supercomputing Centre. *JLSRF* **2018**, *4*, A132.
- (85) Torchala, M.; Moal, I.; Chaleil, R.; Fernandez-Recio, J.; Bates, P. SwarmDock: a server for flexible protein–protein docking. *Bioinformatics* **2013**, *29*, 807–809.

- (86) Porter, K.; Padhorny, D.; Desta, I.; Ignatov, M.; Beglov, D.; Kotelnikov, S.; Sun, Z.; Alekseenko, A.; Anishchenko, I.; Cong, Q.; Ovchinnikov, S.; Baker, D.; Vajda, S.; Kozakov, D. Template-based modeling by ClusPro in CASP13 and the potential for using co-evolutionary information in docking. *Proteins* **2019**, *87*, 1241–1248.
- (87) PyMol, The PyMOL Molecular Graphics System, Version 1.8. 2015; Schrödinger, LLC.
- (88) Waterhouse, A. M.; Procter, J. B.; Martin, D. M. A.; Clamp, M.; Barton, G. J. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **2009**, *25*, 1189–1191.
- (89) Baker, N.; Sept, D.; Joseph, S.; Holst, M.; McCammon, J. Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 10037–10041.
- (90) Jurrus, E. et al. Improvements to the APBS biomolecular solvation software suite. *Protein Sci.* **2018**, *27*, 112–128.
- (91) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38".
- (92) Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; van Gunsteren, W.; Mark, A. Peptide Folding: When Simulation Meets Experiment. *Angew. Chem. Int. Ed.* **1999**, *38*, 236–240.
- (93) Valentini, E.; Kikhney, A. G.; Previtali, G.; Jeffries, C. M.; Svergun, D. I. SASBDB, a repository for biological small-angle scattering data. *Nucleic Acids Res.* **2014**, *43*, D357–D363.
- (94) Kikhney, A. G.; Borges, C. R.; Molodenskiy, D. S.; Jeffries, C. M.; Svergun, D. I. SASBDB: Towards an automatically curated and validated repository for biological scattering data. *Protein Sci.* **2020**, *29*, 66–75.

- (95) Perez-Riverol, Y. et al. The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* **2018**, *47*, D442–D450.

# Supplementary Information:

## Integrative modeling of guanylate binding protein dimers

Wibke Schumann,<sup>†,‡,△</sup> Jennifer Loschwitz,<sup>¶,‡,△</sup> Jens Reiners,<sup>§</sup> Daniel Degrandi,<sup>||</sup>  
Klaus Pfeffer,<sup>||</sup> Kai Stühler,<sup>⊥,#</sup> Gereon Poschmann,<sup>⊥</sup> Sander H.J. Smits,<sup>@,§</sup> and  
Birgit Strodel<sup>\*,¶,‡</sup>

<sup>†</sup>*Institute of Theoretical and Computational Chemistry, Heinrich Heine University Düsseldorf,  
40225 Düsseldorf, Germany*

<sup>‡</sup>*Institute of Biological Information Processing: Structural Biochemistry, Forschungszentrum Jülich,  
52428 Jülich, Germany*

<sup>¶</sup>*Institute of Theoretical and Computational Chemistry, Heinrich Heine University Düsseldorf,  
40225 Düsseldorf, Germany*

<sup>§</sup>*Center for Structural Studies Heinrich-Heine-Universität Düsseldorf, 40225 Düsseldorf, Germany*

<sup>||</sup>*Institute of Medical Microbiology and Hospital Hygiene, Heinrich Heine University, Duesseldorf,  
Germany*

<sup>⊥</sup>*Institute of Molecular Medicine, Proteome Research, Medical Faculty and University Hospital  
Düsseldorf, Heinrich-Heine University Düsseldorf, Universitätsstraße 1, 40225 Düsseldorf, Germany*

<sup>#</sup>*Molecular Proteomics Laboratory, Biomedical Research Centre (BMFZ), Heinrich-Heine  
University Düsseldorf, Universitätsstraße 1, 40225 Düsseldorf, Germany*

<sup>@</sup>*Institute für Biochemie, Heinrich-Heine-Universität Düsseldorf, 40225 Düsseldorf, Germany*

<sup>△</sup>*Contributed equally to this work.*

E-mail: b.strodel@fz-juelich.de

# 1 Methods and materials

## 1.1 HREMD simulation of mGBP7 monomer in solution

The starting structure for all replicas was the same and it was prepared in the following way. The homology model of mGBP7 was solvated and  $\text{Na}^+$  ions were added for the neutralization of the system. An energy minimization of the solvated system was then performed using the steepest descent algorithm, followed by three equilibration MD simulations. First, equilibration in an  $NVT$  ensemble was carried out for 0.1 ns, and then an  $NpT$  equilibration for 1 ns position took place. In both of these equilibration steps, the protein atoms were restrained to their positions with a force constant of  $10 \text{ kJ mol}^{-1} \text{ \AA}^{-2}$  to equilibrate the solvent around the protein and to reach the temperature of 310 K and the pressure of 1 atm. The final equilibration, which was for 20 ns in an  $NpT$  ensemble, was without position restraints on most parts of the proteins, apart from the rigid  $\beta$ -sheets of the G domain. These restraints were kept, also in the HREMD simulation, to avoid overall rotation and translation of the highly prolate protein, which would otherwise require a significantly larger simulation box.<sup>1,2</sup> Thereafter, the HREMD simulations<sup>3</sup> with 30 replicas each were performed for mGBP7<sub>apo</sub> and mGBP7<sub>GTP</sub>. The protein was treated as hot region by modifying its energy function, including the mGBP7–water interactions. To this end, a biasing factor of  $310 \text{ K}/T$  was applied to each of the 30 replicas, where  $T$  is the temperature of the replica in question and which were exponentially distributed between 310 and 450 K. This includes one unbiased replica, called target replica at 310 K. Exchanges between neighbored replicas were attempted every 2 ps, and an average exchange probability of  $\sim 30\%$  was reached. Each replica simulation was 400 ns long, which leads to an accumulated simulation time of  $3.2 \mu\text{s}$  per HREMD simulation.

To ensure that GTP stayed in its binding pocket in mGBP7<sub>GTP</sub>, we applied distant restraints between GTP and K51, Y53, and D97 using the pull code of GROMACS. The HREMD simulations were conducted with GROMACS 2016.4 in combination with the PLUMED plugin (version 2.4.1 from <https://github.com/GiovanniBussi/plumed2/tree/>

v2.4).<sup>4</sup> For the temperature and pressure regulation, a velocity rescaling thermostat with canonical sampling<sup>5</sup> and an isotropic Parrinello-Rahman barostat<sup>6</sup> were used. The particle-mesh Ewald (PME) method<sup>7,8</sup> was applied for the calculation of electrostatic interactions in conjunction with periodic boundary conditions. The cutoff value of the Lennard-Jones and short-range electrostatic interaction was 12 Å. We used the leapfrog stochastic dynamics integrator for the integration of equations of motion and the LINCS algorithm<sup>9</sup> to constrain all bond lengths. For the mGBP7<sub>apo</sub>, we treated certain hydrogen atoms as virtual interaction sites, which permitted an integration time step of 4 fs while maintaining energy conservation.<sup>10</sup> In case of mGBP7<sub>GTP</sub>, a time step of 2 fs was applied. The coordinates and velocities were saved every 20 ps.

## 1.2 Protein expression for mGBP7

The expression and purification of the mGBP7 protein based on the protocol of Legewie et. al 2019.<sup>11</sup> For the expression of the mGBP7 protein, competent *E. coli* Rosetta 2 (DE3) pLysS (Novagen) cell were transformed with the pQE-80L vector (Qiagen), containing the n-terminal hexa-histidine tagged mgbp7 gene.<sup>12</sup> We prepared 4 L 2YT (16 g/l tryptone, 10 g/l yeast extract, 5 g/l sodium chloride) media, supplemented with 3.4 µg/ml chloramphenicol and 10 µg/ml ampicillin. We inoculate the media with the mGBP7 expression cell with a starting OD<sub>600</sub> of 0.1 and incubate them at 37 °C and 180 rpm shaking to an OD<sub>600</sub> of 0.5. The protein expression was started by induction with 150 µM IPTG and further incubation at 37°C and 180 rpm for 4h. The cells were harvested at 8000g for 30 min and the supernatant was discarded. The resulting cell pellet was resuspended with buffer (50 mM Tris pH 8.0, 300 mM NaCl, 5 mM MgCl, 10 mM Imidazol, 1 mM DTT and 0.25 mM Pefabloc). The cell disruption was done with 2.7 kbar on a Constant Cell Disruption System in 3 cycles followed by a high spin centrifugation step (100000 g, 1h, 4°C). The supernatant was incubated overnight with Ni-NTA-Agarose beads (Qiagen) at 4°C. After the overnight incubation, the Ni-NTA-Agarose beads were washed four times with wash buffer (50 mM

Tris pH 8.0, 150 mM NaCl, 5 mM MgCl, 10 mM Imidazol, 1 mM DTT) and the final elution was done with elution buffer (50 mM Tris pH 8.0, 150 mM NaCl, 5 mM MgCl, 300 mM Imidazol, 1 mM DTT). The elution fraction was concentrated to 5 ml with a 50 kDa cut-off filter and loaded onto a HiLoad<sup>®</sup> 26/600 Superdex<sup>®</sup> 200 pg column (Cytiva), preequilibrated with SEC buffer (50 mM Tris pH 8.0, 5 mM MgCl, 2 mM DTT). Elution peak fraction were concentrated and used for further experiments.

### 1.3 Cross linking mass spectrometry of mGBP7

Here, 6  $\mu$ g of mGBP7 was crosslinked in a final volume of 10  $\mu$ l using 0.5 mM bis(sulfosuccinimidyl) suberate (BS3) or 0.5 mM disuccinimidyl sulfoxide (DSSO) for 30 minutes at room temperature in 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid containing aqueous buffer. For control reactions, the crosslinkers were omitted. The reaction was stopped by adding an aqueous solution of 0.5  $\mu$ l 1 M tris(hydroxymethyl)aminomethane pH 7.5 for 15 minutes. Samples were separated in 4-12% Bis-Tris polyacrylamidegels. After staining with Coomassie blue, mGBP7 monomer and dimer containing bands were cut-out and processed for mass spectrometric analysis essentially as described.<sup>13</sup> Briefly, the protein was reduced with dithiothreitol, alkylated with iodoacetamide and digested with trypsin. Resulting peptides and crosslinked peptides were resuspended in 0.1% trifluoroacetic acid and analyzed by liquid chromatography coupled mass spectrometry. Peptides were separated for one hour on an Ultimate 3000 Rapid Separation Liquid Chromatography system on a 25 cm length C18 column as described<sup>14</sup> and analyzed by a Fusion Lumos mass spectrometer, online coupled via a nano-electrospray interface.

BS3 crosslinked samples were analyzed as follows: A survey spectra was recorded in the Orbitrap analyser (scan range 400–1800 m/z, resolution 60000, maximum injection time 50 ms, AGC target 100000) and subsequently, 2-10 fold charged precursors were selected (minimum intensity 50000, maximum intensity 1E20, 1.6 m/z isolation window), fragmented with collisional induced dissociation (CID) and independently with higher energy collisional

dissociation (HCD). Fragment spectra were recorded in the Orbitrap (resolution 30000, maximum injection time 100 ms, AGC target 50000). The cycle time was set to 2 seconds and already fragmented precursors were excluded from further isolation for the next minute. For DSSO crosslinked samples, survey scans were carried out with following parameters: scan range 350–1600 m/z, resolution 60000, maximum injection time 50 ms, AGC target 400000. Next, 3–8 fold charged precursors were selected (minimum intensity 20000, maximum intensity 1E20, 1.6 m/z isolation window), fragmented by CID (collision energy 25%) and analyzed in the Orbitrap (resolution 30000, maximum injection time 100 ms, AGC target 50000). Subsequently, MS3 scans of two MS2 precursors were carried out in the ion trap for masses matching DSSO induced differences (isolation window 2.5 m/z, MS2 isolation window 2 m/z, scan rate: rapid, maximum injection time 120 ms, AGC target 20000). Finally, an MS2 scan was carried out in the Orbitrap after ETD fragmentation (isolation window 1.6 m/z, resolution 50000, maximum injection time 150 ms, AGC target 200000). The cycle time was set to 4 seconds.

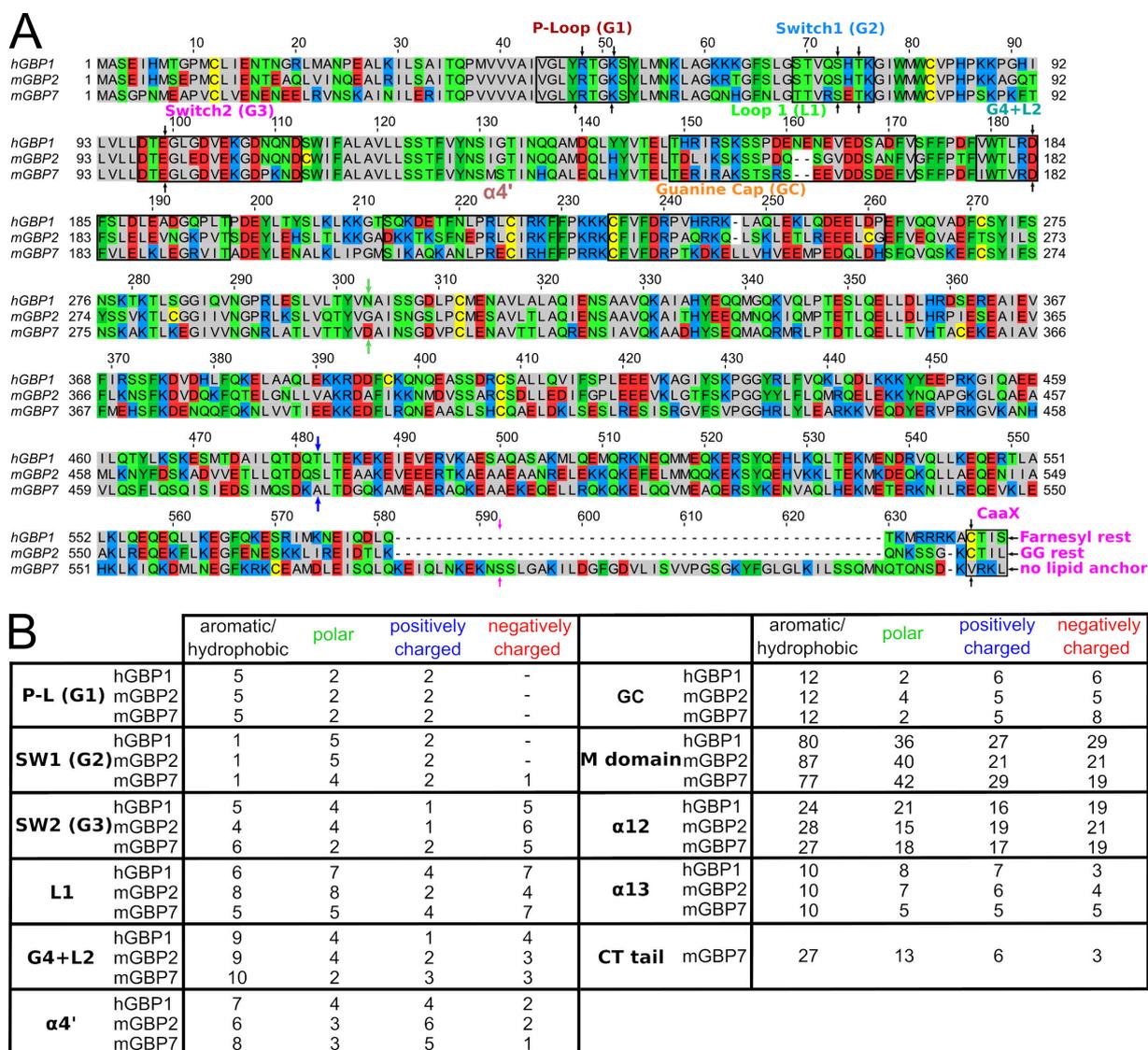
Data analysis of BS3 crosslinked samples was carried out using the mGBP7 amino acid sequence with MeroX (version 2.0.2.4)<sup>15</sup> considering C<sub>8</sub>H<sub>10</sub>O<sub>2</sub> as mass shift for the crosslinks between lysine residues. Methionine oxidation was considered as fixed and cysteine carbamidomethylation as variable modification and up to three missed tryptic cleavage sites. Precursor precision was set to 5 ppm and fragment precision to 10 ppm. Crosslinked peptides were reported at a false discovery rate of 1%. For analysis of DSSO crosslinks, the Proteome Discoverer Software (version 2.3.0.523) including XlinkX was used applying tryptic cleavage specificity with a maximum of two missed cleavage sites, carbamidomethylation on cysteines as fixed and methionine oxidation as variable modifications. Spectra associated with potentially crosslinked peptides were filtered (XlinkX detect, +158.004 Da crosslink modification between lysines) and searched by XlinkS (precursor mass tolerance 10 ppm, Orbitrap fragment spectra mass tolerance 20 ppm, ion trap fragment mass tolerance 0.5 Da). Non-crosslinked peptides associated spectra (CID and EThdD) were subjected to a Sequest

HT based searches including hydrolyzed DSSO (+176.014 Da) at lysine residues as additional variable modification. Precursor tolerances were 10 ppm and fragment spectra tolerances 0.02 Da. Identified peptides and crosslinks were accepted at a false discovery rate of 1%. Only crosslinks were reported which were identified in two independent experiments of a sample group.

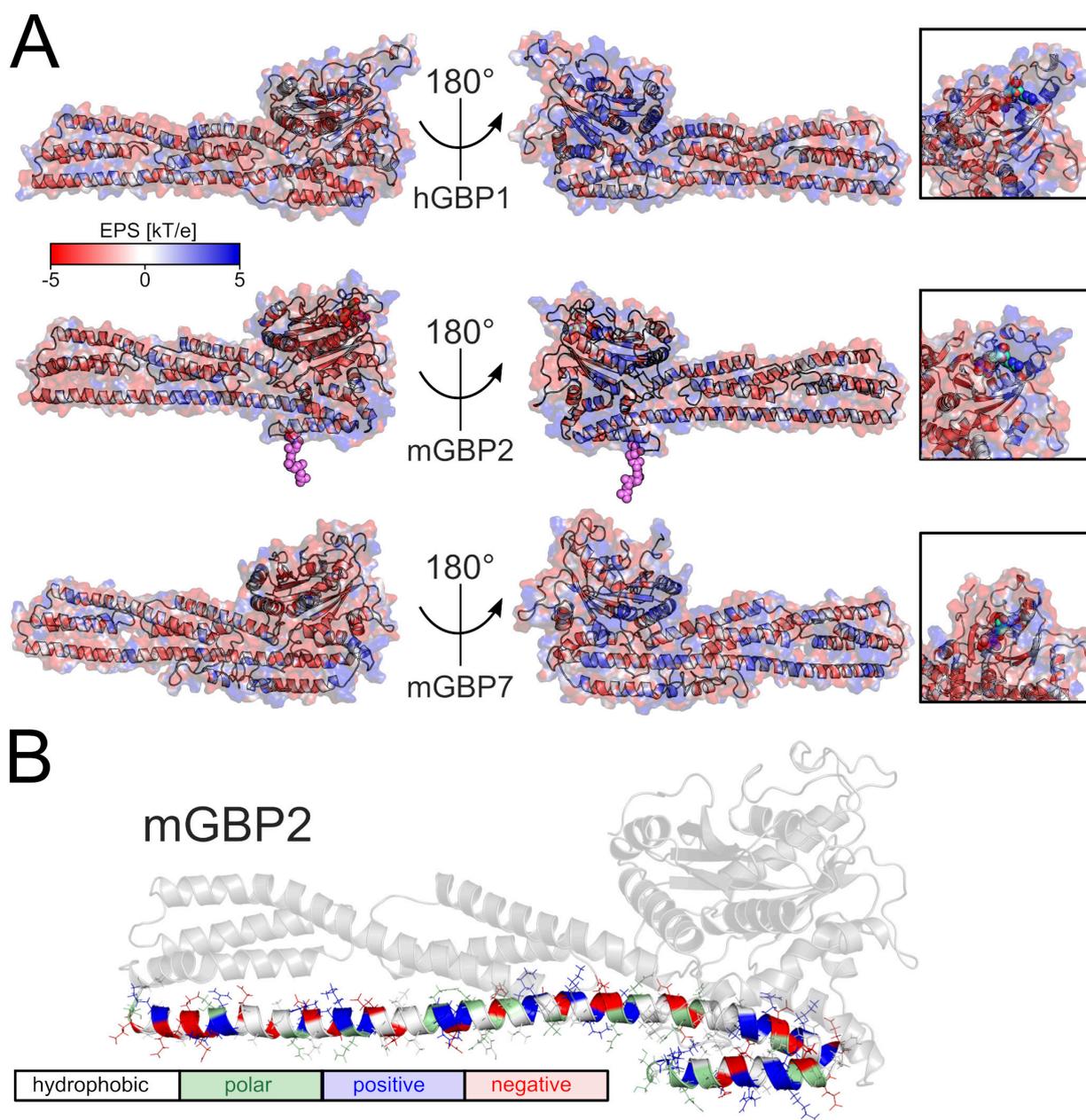
## 1.4 SAXS of mGBP7

We performed the small-angle X-ray scattering (SAXS) measurements of mGBP7 on our Xeuss 2.0 Q-Xoom system (Xenocs). This system is equipped with a GENIX 3D CU Ultra Low Divergence x-ray beam delivery system (Xenocs) and a PILATUS 3 R 300K detector (Dectris). The chosen sample to detector distance for this experiment was 0.55 m, results in an achievable q-range of  $0.05 - 6.5 \text{ nm}^{-1}$ . The measurement was performed at  $10^\circ\text{C}$  with a protein concentration range of  $1.66 - 6.85 \text{ mg/ml}$ . The system autosampler injected the mGBP7 samples in the Low Noise Flow Cell (Xenocs). We collect six frames with an exposure time of ten minutes/frame and scaled the data to absolute intensity against water. The radial averaging of the scattering data was done with Foxtrot (v.3.4.9, Soleil/Xenocs). All other used programs for data processing were part of the ATSAS Software package (Version 3.0.3).<sup>16</sup> Primary data reduction (merging of data and background subtraction) was performed with the program PRIMUS.<sup>17</sup> The forward scattering  $I(0)$  as well as the radius of gyration ( $R_g$ ) was determined with the Guinier approximation.<sup>18</sup> The program GNOM<sup>19</sup> was used to estimate the maximum particle dimension ( $D_{max}$ ), based on the pair-distribution function  $p(r)$ . Low resolution ab initio models were calculated with GASBOR<sup>20</sup> with a P2 symmetry. Rigid body modeling of the mGBP7 dimer was done with SASREF.<sup>21</sup> Superimposings of the mGBP7 dimer models were done with the program SUPCOMB.<sup>22</sup> The agreement of the mGBP7 dimer models were checked with the program CRY SOL.<sup>23</sup> The complete data are summarized in Tab. S3.

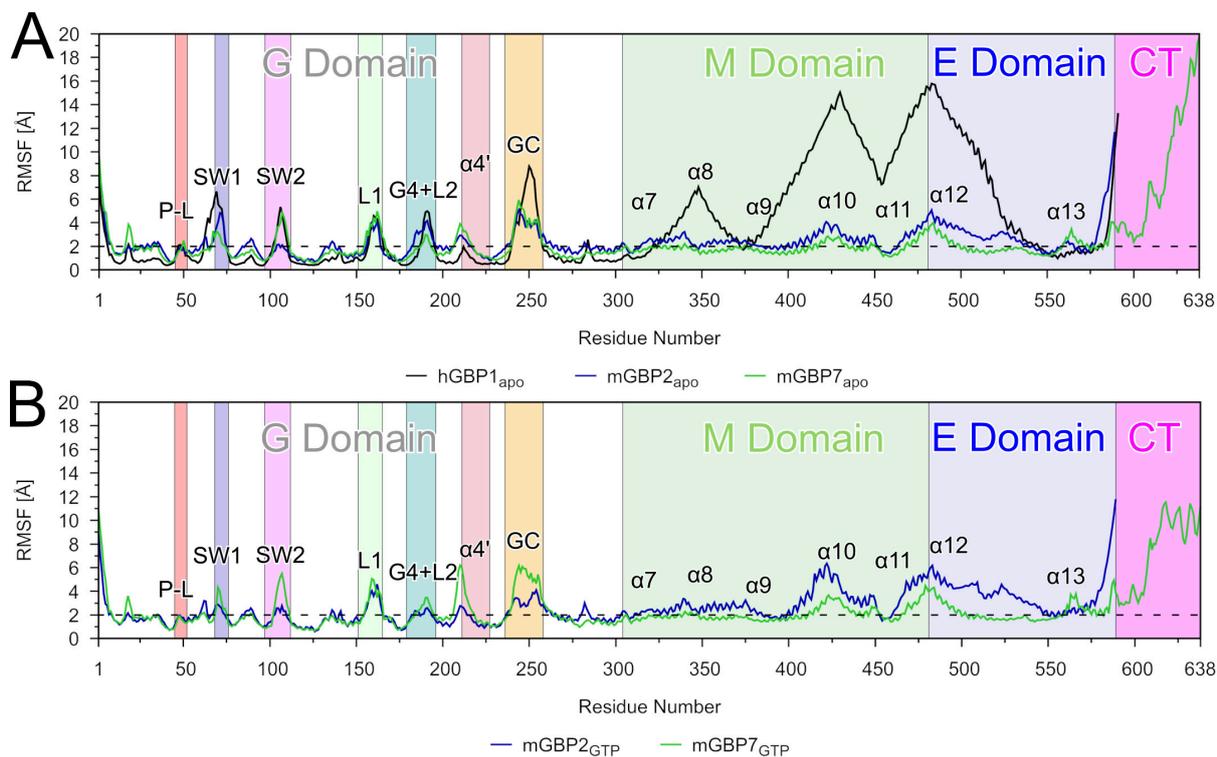
## 2 Supplementary information figures



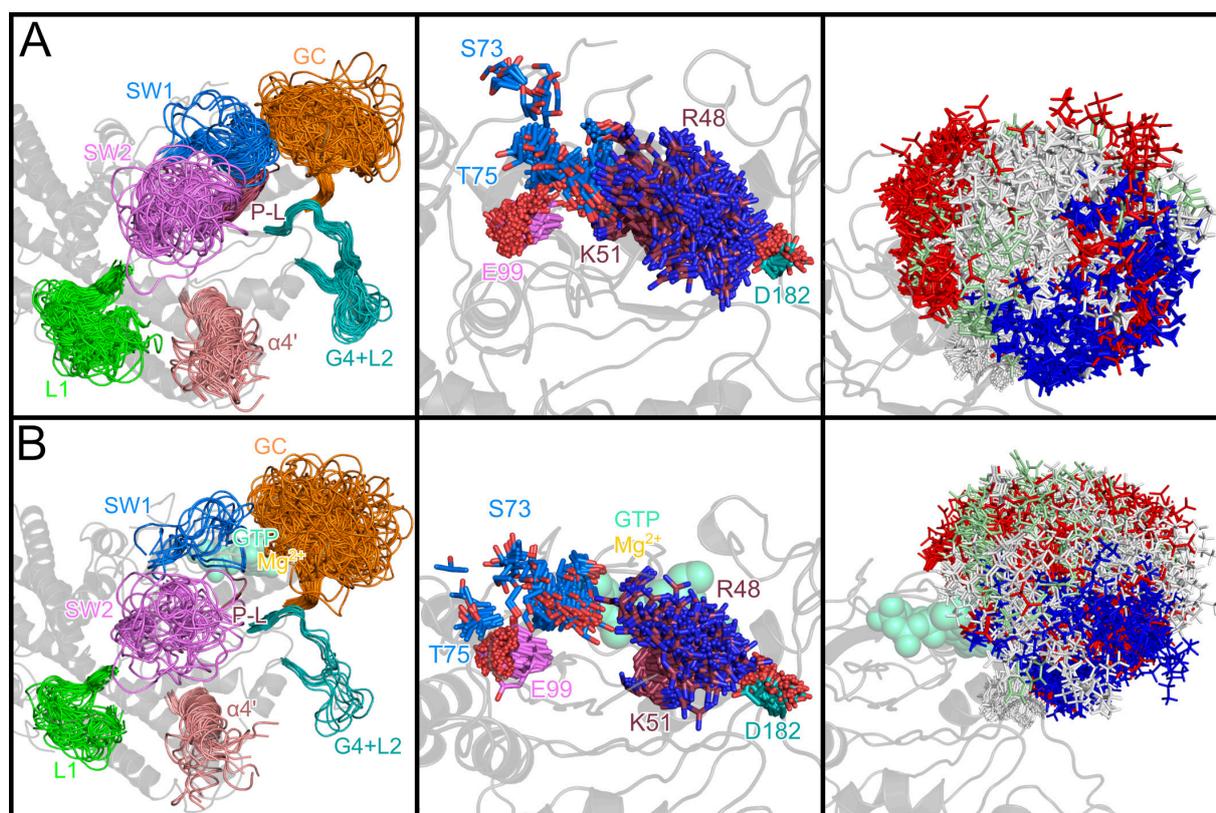
**Figure S1: Sequence alignment and characteristics of hGBP1, mGBP2, and mGBP7.** (A) In the sequence alignment, the residues are colored based on their physicochemical properties: apolar, gray; polar, green; aromatic, dark green; negatively charged, red; positively charged, blue; cysteine, yellow. The four conserved GTP-binding site motifs and other important structural elements are highlighted by black boxes and labeled. The key residues for GTP binding and hydrolysis are indicated by black arrows, while the green and blue arrows mark the beginning of the M and the E domain, respectively. The magenta arrows indicate the start of the CT tail in mGBP7. The sequence alignment was calculated with T-coffee<sup>24,25</sup> and the figure created with Jalview 2.<sup>26</sup> (B) The count of different residue types in the G motifs of hGBP1, mGBP2, and mGBP7.



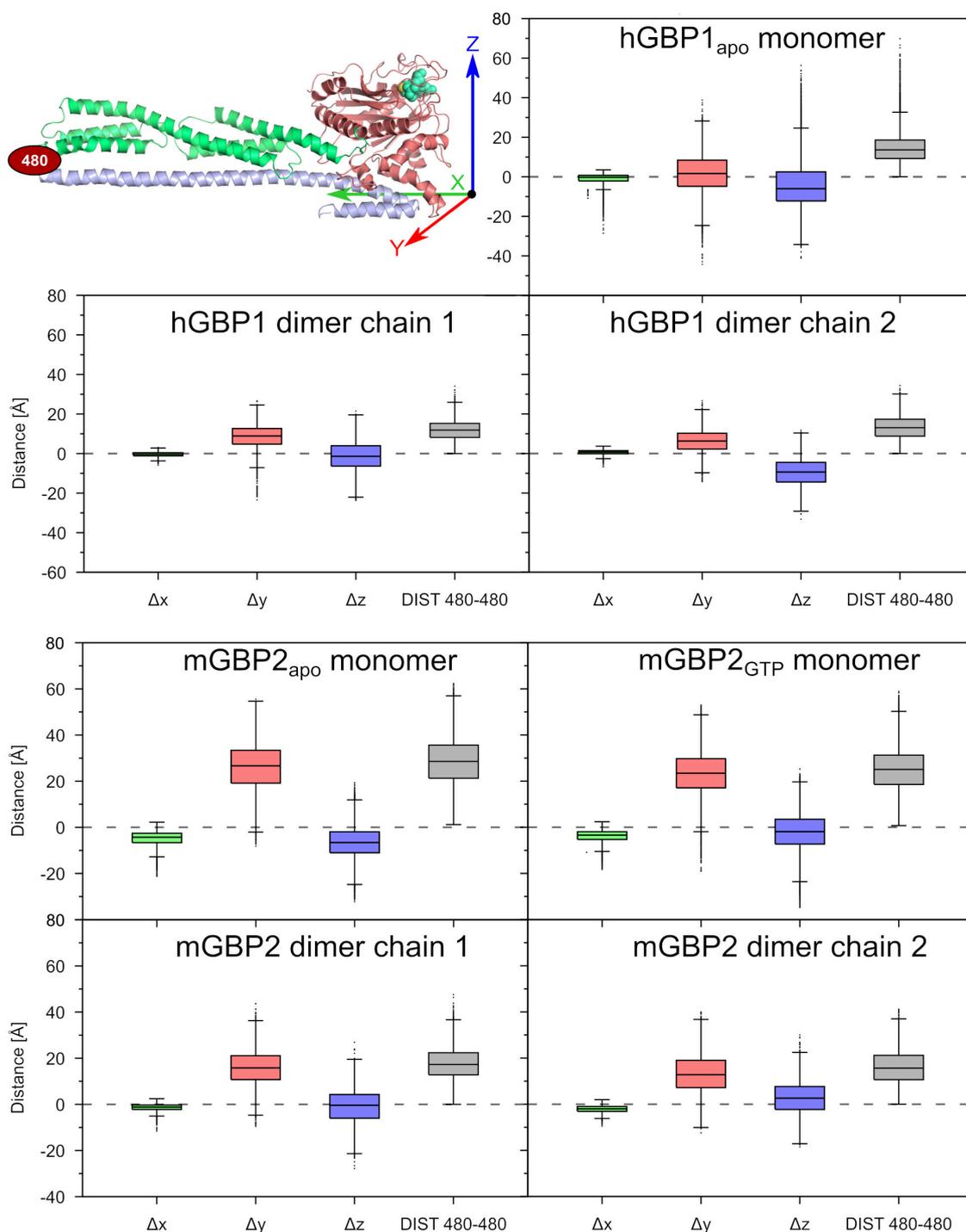
**Figure S2: Electrostatic potential surface (EPS) of hGBP1, mGBP2, and mGBP7 and the distribution of charged residues in the E domain of mGBP2.** (A) The EPS was calculated with the APBS webserver and was illustrated with PyMOL and the APBS tool 2.1. The EPS is shown between  $-5$  (red) and  $+5$  kT/e (blue) for the whole proteins and as zoom for their GTP binding sites. (B) The residues of  $\alpha 12/13$  of mGBP2 are colored based on their physicochemical properties: apolar and aromatic, white; polar, green; negatively charged, red; positively charged, blue.



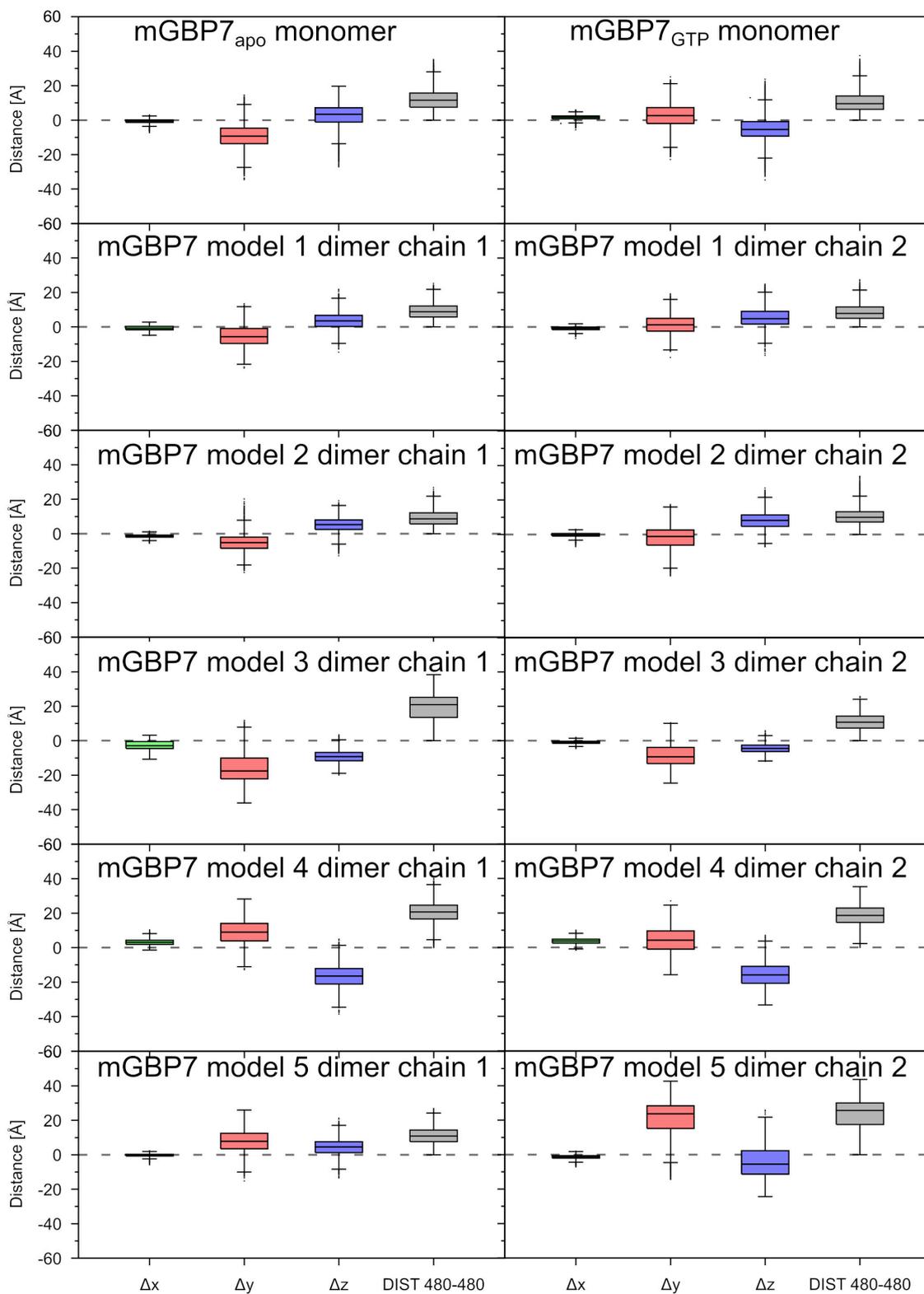
**Figure S3: Fluctuations of the GBP residues during HREMD simulations of the protein monomers.** The fluctuations are quantified by the RMSF of the  $C_{\alpha}$  atoms of hGBP1 (black), mGBP2 (blue), and mGBP7 (green) in **(A)** the apo-state and **(B)** the GTP-bound state. All motifs and loops of the G domains as well as helices in the M/E domain are labeled, and the background of the plots is colored to indicate the different structural parts of the GBPs (where the same colors as in Fig. 1C/D were used). The horizontal dashed line at 2 Å is to identify flexible residues with RMSF values exceeding that value.



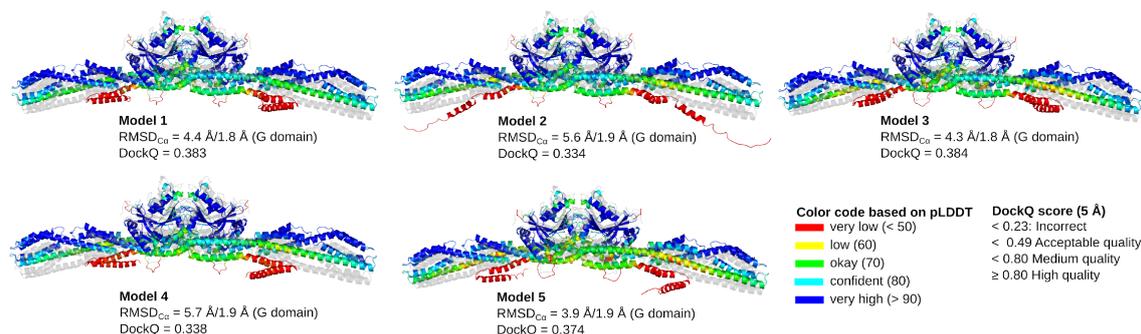
**Figure S4: Conformational clusters of G domain structural elements and selected residues of (A) mGBP7<sub>apo</sub> and (B) mGBP7<sub>GTP</sub> monomers.** (Left) Representative structures of the clusters of the G motifs and loops that were determined using a 2.0 Å RMSD cutoff applied to the fitted target replica of the respective HREMD simulation. The colors of the loops are: P-L in red, SW1 in blue, SW2 in magenta, L1 in green, G4+L2 in turquoise, the guanine cap in orange, and  $\alpha 4'$  in salmon. (Middle) The different conformations of key residues for GTP binding and hydrolysis, as sampled during the HREMD simulations, are shown as sticks for the side chains. The oxygen and nitrogen atoms of these side chains are colored in red and blue, respectively, while all other side-chain atoms are shown with the same color as chosen in (A) for the structural element they belong to. (Right) The different conformations of all side chains of the guanine cap are shown as lines and colored according to their residue type (white, apolar; green, polar; blue, positively charged; red, negatively charged). In all panels, the homology model of mGBP7 is shown as a gray cartoon. In mGBP7<sub>GTP</sub>, GTP and Mg<sup>2+</sup> are shown in green and orange.



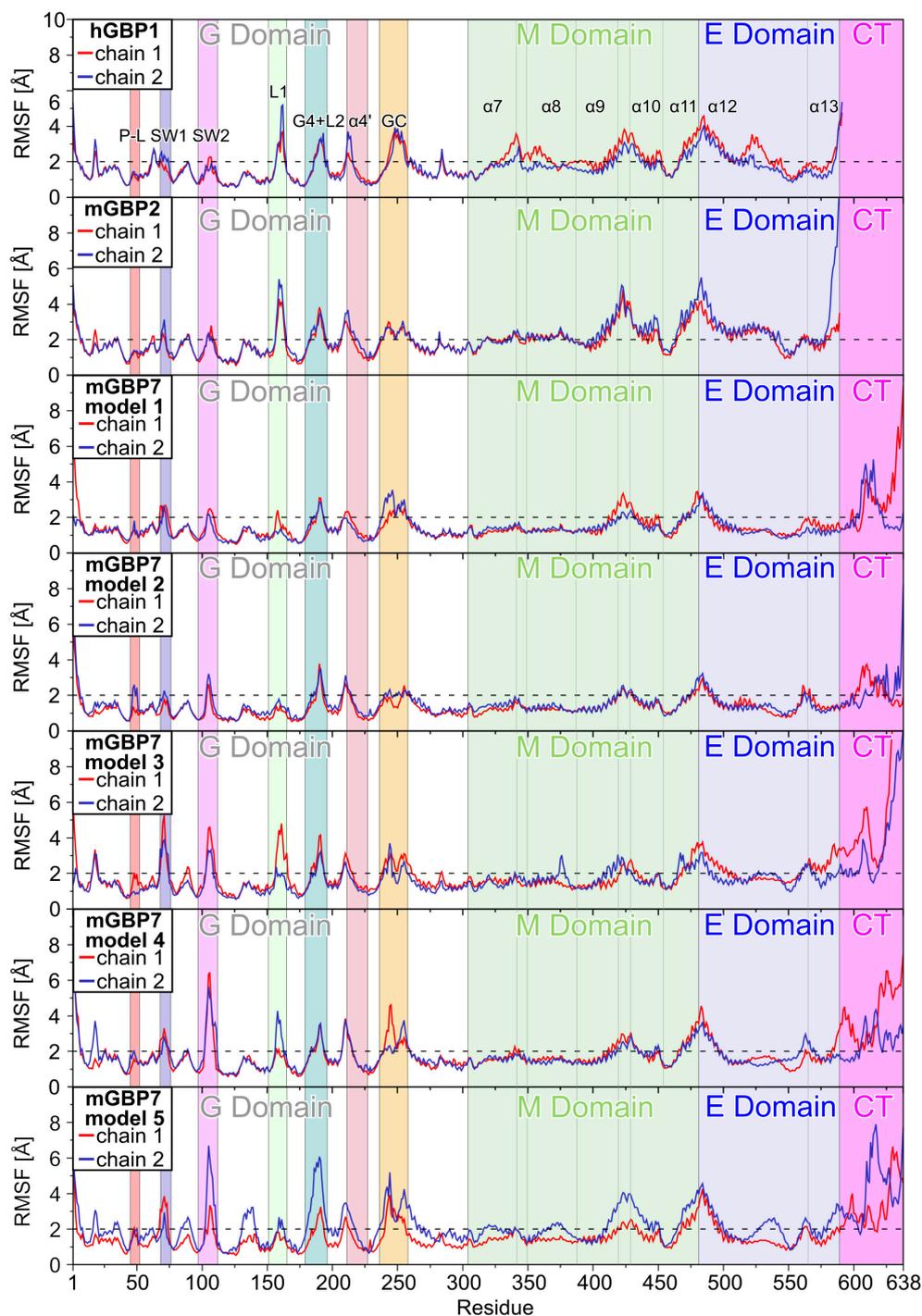
**Figure S5: Boxplots of  $\Delta x$ ,  $\Delta y$ , and  $\Delta z$  and  $d_{480}$  of residue 480 for the hGBP1 and mGBP2 systems.** The orientation of the coordinate system used for these calculations is shown in the top left panel. The colors of the boxes correspond to the colors of the coordinate axes:  $\Delta x$ , green;  $\Delta y$ , red;  $\Delta z$ , blue. The  $d_{480}$  boxes are shown in gray. The monomers and individual chains of the dimers were aligned to the hGBP1 or mGBP2 reference structure, illustratively shown for hGBP1 in the top left panel.



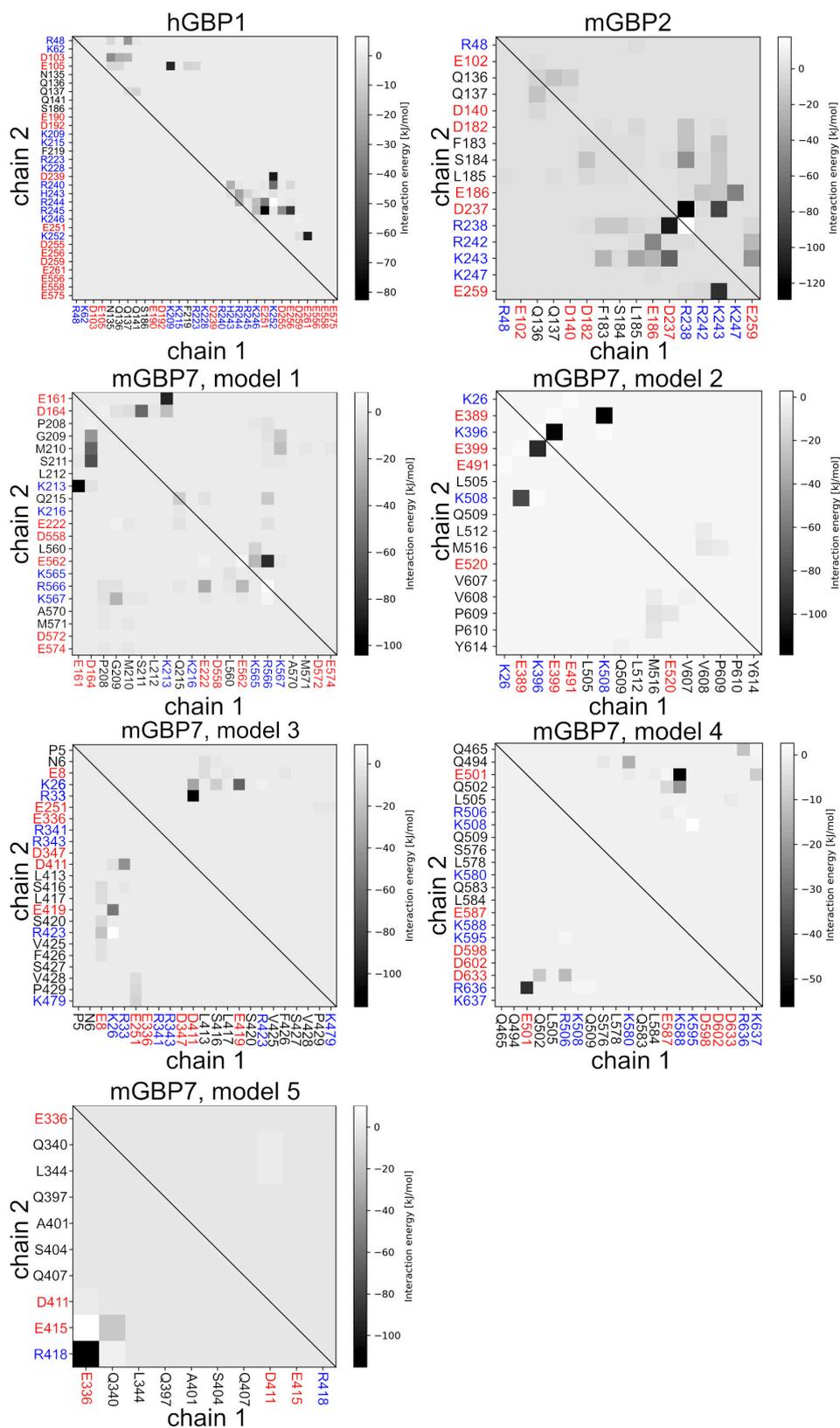
**Figure S6: Boxplots of  $\Delta x$ ,  $\Delta y$ , and  $\Delta z$  and  $d_{480}$  of residue 480 for the mGBP7 systems. See Fig. S5 for further details.**



**Figure S7: The five most likely mGBP7 dimer models as predicted by AlphaFold-Multimer.** The dimer models are colored based on a per-residue estimate of the prediction’s confidence (called pLDDT) on a scale from 0–100. Regions with pLDDT > 90 are shown in blue and expected to be modeled to high accuracy. Regions with pLDDT between 70 and 90 are expected to be modeled well (a generally good backbone prediction), while the predictions for regions with pLDDT between 50 and 70 are of low confidence and should be treated with caution. The 3D coordinates of regions with pLDDT < 50 should not be interpreted. Such low pLDDT values are a strong predictor of disorder or that the region in question is only structured as part of a complex. Most parts of the G and M domains in the AlphaFold-Multimer models are predicted with high confidence, while it is lower for  $\alpha$ 12/13 of the E domain and very low for the CT tail. The deviation from the G domain dimer model of mGBP2 (shown as gray cartoon) is provided in terms of the RMSD of the C $\alpha$  atoms and the DockQ score.<sup>27</sup>



**Figure S8: Fluctuations of the GBP residues during MD simulations of the protein dimers.** The fluctuations are quantified by the RMSF of the  $C_{\alpha}$  atoms of the two chains per dimer (shown in red and blue) after individually aligning the chains. All motifs and loops of the G domains as well as helices in the M/E domain are labeled, and the background of the plots is colored to indicate the different structural parts of the GBPs (where the same colors as in Fig. 1C/D were used). The horizontal dashed line at 2 Å is to identify flexible residues with RMSF values exceeding that value.



**Figure S9: interaction energies between the two chains of the GBP dimers.** These energies were calculated from the 100 ns MD simulations of the dimer models. Only the energies considerably different from zero are shown. The energies are given in kJ/mol, according to the color scale on the right per plot. It should be noted that the energetic scales differ between the plots, as the color black was chosen for the minimal energy encountered per system. The residue labels are given on the axes, with red and blue for negatively and positively charged residues, respectively.

### 3 Supplementary information tables

**Table S1: Flexibility of the motifs and loops of the G domain and helix  $\alpha 4'$ .** This analysis based on conformational clustering applied to the  $C_\alpha$  atoms after aligning the target replica of the respective HREMD simulation on the  $\beta$ -sheets of G domain and using a 2.5 Å RMSD cutoff for assigning cluster membership.

	hGBP1 <sub>apo</sub>			RMSD [Å] <sup>b</sup>			
	clusters	population [%] <sup>a</sup>					
P-L	6	99.7				6.3/5.9	
SW1	93	61.7				14.8/13.4	
SW2	59	73.7				13.1/11.3	
L1	36	79.0				9.2/8.2	
G4+L2	55	64.4				12.4/11.8	
$\alpha 4'$	2	100				4.6/2.5	
GC	197	34.4				20.4/19.9	
	mGBP2 <sub>apo</sub>			mGBP2 <sub>GTP</sub>			%change <sup>c</sup>
	clusters	population [%] <sup>a</sup>	RMSD [Å] <sup>b</sup>	clusters	population [%] <sup>a</sup>	RMSD [Å] <sup>b</sup>	
P-L	1	100	4.4/-	1	100	1.9/-	0
SW1	23	82.4	9.9/9.9	1	100	3.0/-	-95.7
SW2	3	100	5.3/4.9	1	100	4.4/-	-66.7
L1	23	89.6	11.2/10.8	17	92.2	9.9/9.4	-26.1
G4+L2	12	95.9	9.4/7.7	3	100	5.4/4.5	-75.0
$\alpha 4'$	2	100	5.4/3.4	3	100	6.1/5.2	+33.3
GC	29	66.1	10.7/8.9	3	100	6.6/5.3	-89.7
	mGBP7 <sub>apo</sub>			mGBP7 <sub>GTP</sub>			%change <sup>c</sup>
	clusters	population [%] <sup>a</sup>	RMSD [Å] <sup>b</sup>	clusters	population [%] <sup>a</sup>	RMSD [Å] <sup>b</sup>	
P-L	21	82.4	20.5/20.2	1	100	4.3/-	-95.2
SW1	45	68.4	29.8/27.5	14	89.7	9.8/8.5	-68.9
SW2	68	44.0	35.1/34.6	32	78.1	12.6/10.9	-52.9
L1	100	54.4	35.6/34.5	30	77.3	11.4/9.6	-70.0
G4+L2	36	67.8	24.9/23.7	6	99.6	8.3/6.4	-83.3
$\alpha 4'$	63	58.2	36.0/34.4	17	92.3	8.8/8.5	-73.1
GC	111	51.7	23.7/22.9	96	58.5	18.2/17.1	-13.5

<sup>a</sup> Percentage of the structures which are cumulatively represented by the first three clusters.

<sup>b</sup> The largest RMSD found between any two clusters.

<sup>c</sup> %change =  $-(1 - \text{\#clusters}(\text{with GTP})/\text{\#clusters}(\text{apo})) \cdot 100$  or

%change =  $+(1 - \text{\#clusters}(\text{apo})/\text{\#clusters}(\text{with GTP})) \cdot 100$



**Table S2:** Summary of crosslinks by DSSO and BS3 for the mGBP7 monomer and dimer.

State <sup>a</sup>	Residue pair <sup>b</sup>	Total <sup>c</sup>	model 1 <sup>d</sup>	model 2 <sup>d</sup>	model 3 <sup>d</sup>	model 4 <sup>d</sup>	model 5 <sup>d</sup>
<b>DSSO</b>							
M & D	106–216	5	both	intra	intra	intra	intra
M & D	495–440*	5	intra	intra	intra	intra	intra
M & D	580–588	12	intra	intra	intra	intra	intra
M & D	586–539	7	intra	intra	intra	intra	intra
M & D	588–539*	5	intra	intra	intra	intra	intra
M & D	595–524*	8	intra	both	intra	both	intra
(M) & D	588–155	3	intra	intra	intra	intra	intra
(M) & D	205–554	4	both	intra	intra	intra	intra
(M) & D	205–565	4	both	intra	intra	intra	intra
D	373–588	2	intra	intra	intra	intra	intra
D	613–510	2	intra	both	intra	both	intra
D	613–524	2	intra	both	intra	both	intra
D	524–588*	2	intra	intra	intra	both	intra
D	557–565	6	both	intra	intra	intra	intra
<b>BS3</b>							
M & D	588–539*	6	intra	intra	intra	intra	intra
M & D	92–89	18	intra	intra	intra	intra	intra
M & D	216–216	18	inter	no	no	no	no
M & D	495–440*	7	intra	intra	intra	intra	intra
M & D	611–510	5	intra	both	intra	both	intra
M & D	510–500	19	intra	both	intra	intra	intra
(M) & D	524–588*	4	intra	both	intra	both	intra
(M) & D	567–565	4	both	intra	intra	intra	intra
D	565–565	2	inter	no	no	no	no
D	595–524*	3	intra	intra	intra	both	intra
D	205–567	6	both	intra	intra	intra	intra
D	205–576	2	both	intra	intra	intra	intra
D	567–567	10	inter	no	no	no	no
D	567–576	2	both	intra	intra	intra	intra
M	98–89	2	—	—	—	—	—

<sup>a</sup> The crosslinks have been detected in the monomer (M) and/or dimer (D) band from crosslinked and size-separated mGBP7. Brackets are used if the linked peptide pair was found only in one of two experiments.

<sup>b</sup> Linked residues pairs found with both crosslinkers are marked by \*.

<sup>c</sup> The total number of crosslinks for this residue pair in the monomer and/or dimer.

<sup>d</sup> It is indicated which crosslinks agree with the respective dimerization model of mGBP7: within a monomer (intra), intermolecular (inter), both intra- and intramolecular (both), no crosslink possible (no).

**Table S3: Overall SAXS Data.**

<b>SAXS Device</b>	<b>Xenocs Xeuss 2.0 with Q-Xoom</b>
<b>Data collection parameters</b>	
Detector	PILATUS 3 R 300K windowless
Detector distance (m)	0.550
Beam size	0.8 mm x 0.8 mm
Wavelength (nm)	0.154
Sample environment	Low Noise Flow Cell, 1 mm $\phi$
s range (nm <sup>-1</sup> ) $\ddagger$	0.05 – 6.5
Exposure time per frame (s)	600 (6 frames)
<b>Sample</b>	<b>mGBP7</b>
Organism	Mus musculus (Mouse)
UniProt ID	Q91Z40 (1-638)
Mode of measurement	batch
Temperature ( $^{\circ}$ C)	10
Protein concentration (mg/ml)	1.66 – 6.85 (merged)
Buffer	50mM Tris pH 8.0, 5 mM MgCl, 2 mM DTT
<b>Structural parameters</b>	
I(0) from P(r)	0.01
$R_g$ (real-space from P(r)) (nm)	5.08
I(0) from Guinier fit	0.01
s-range for Guinier fit (nm <sup>-1</sup> )	0.099 – 0.240
$R_g$ (from Guinier fit) (nm)	5.35
points from Guinier fit	2 – 26
$D_{max}$ (nm)	17.00
POROD volume estimate (nm <sup>3</sup> )	176.37
<b>Molecular mass (kDa)</b>	
From I(0)	137.09
From Qp <sup>28</sup>	136.31
From MoW2 <sup>29</sup>	129.49
Bayesian Inference <sup>30</sup>	130.86
From POROD	110.23
From sequence	73.83 (monomer) 147.66 (dimer)
<b>Structure Evaluation</b>	
GASBOR fit $\chi_2$	1.14
SASREF fit $\chi_2$	1.27
Ambimeter score	2.601
Crysol fit $\chi_2$	1.89
<b>Software</b>	
ATSAS Software Version <sup>16</sup>	3.0.3
Primary data reduction	PRIMUS <sup>17</sup>
Data processing	GNOM <sup>19</sup>
<i>Ab initio</i> modeling	GASBOR <sup>20</sup>
Rigid body modeling	SASREF <sup>21</sup>
Superimposing	SUPCOMB <sup>22</sup>
Structure evaluation	AMBIMETER <sup>31</sup> / CRY SOL <sup>23</sup>
Model visualization	PyMOL <sup>32</sup>

$\ddagger$  s =  $4\pi \sin(\theta)/\lambda$ ,  $2\theta$  – scattering angle,  $\lambda$  – X-ray-wavelength

**Table S4:**  $R_g$  and  $D_{\max}$  of the mGBP7 dimer models in comparison to the SAXS data determined with CRY SOL.

Protein	$\chi^2$	$R_{g,\text{sim}}$ [nm]	$R_{g,\text{SAXS}}$ [nm]	$D_{\max,\text{sim}}$ [nm]	Envelope diameter [nm]
model 1	1.66	5.5	5.50	23.1	24.1
model 2	1.79	5.0	4.92	17.6	17.9
model 3	1.74	4.9	4.97	14.9	16.7
model 4	1.89	4.9	5.03	15.9	17.3
model 5	1.27	5.5	5.67	18.6	19.1

## References

- (1) Barz, B.; Loschwitz, J.; Strodel, B. Large-scale, dynamin-like motions of the human guanylate binding protein 1 revealed by multi-resolution simulations. *PLoS Comput. Biol.* **2019**, *15*, 1–29.
- (2) Loschwitz, J.; Steffens, N.; Wang, X.; Schäffler, M.; Pfeffer, K.; Degrandi, D.; Strodel, B. Domain motions, dimerization, and membrane interactions of the murine guanylate binding protein 2. *bioRxiv* **2022**,
- (3) Bussi, G. Hamiltonian replica exchange in GROMACS: a flexible implementation. *Mol. Phys.* **2014**, *112*, 379–384.
- (4) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New feathers for an old bird. *Comp. Phys. Commun.* **2014**, *185*, 604–613.
- (5) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.
- (6) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single-Crystals - a New Molecular-Dynamics Method. *Mol. Phys.* **1981**, *52*, 7182–7190.
- (7) T. Darden, D. Y.; Pedersen, L. Particle Mesh Ewald - an  $N \cdot \text{Log}(N)$  Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

- (8) Essmann, U.; Perera, L.; Berkowitz, M. L. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.
- (9) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463–1472.
- (10) Feenstra, K. A.; Hess, B.; Berendsen, H. J. C. Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. *J. Comput. Chem.* **1999**, *20*, 786–798.
- (11) Legewie, L.; Loschwitz, J.; Steffens, N.; Prescher, M.; Wang, X.; Smits, S. H.; Schmitt, L.; Strodel, B.; Degrandi, D.; Pfeffer, K. Biochemical and structural characterization of murine GBP7, a guanylate binding protein with an elongated C-terminal tail. *Biochem. J.* **2019**, *476*, 3161–3182.
- (12) Hanahan, D. Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* **1983**, *166*, 557–580.
- (13) Brenig, K.; Grube, L.; Schwarzländer, M.; Köhrer, K.; Stühler, K.; Poschmann, G. The Proteomic Landscape of Cysteine Oxidation That Underpins Retinoic Acid-Induced Neuronal Differentiation. *J. Proteome Res.* **2020**, *19*, 1923–1940.
- (14) Prescher, N.; Hänsch, S.; Knobbe-Thomsen, C. B.; Stühler, K.; Poschmann, G. The migration behavior of human glioblastoma cells is influenced by the redox-sensitive human macrophage capping protein CAPG. *Free Radic. Biol. Med.* **2021**, *167*, 81–93.
- (15) Iacobucci, C.; Götze, M.; Ihling, C. H.; Piotrowski, C.; Arlt, C.; Schäfer, M.; Hage, C.; Schmidt, R.; Sinz, A. A cross-linking/mass spectrometry workflow based on MS-cleavable cross-linkers and the MeroX software for studying protein structures and protein-protein interactions. *Nature protocols* **2018**, *13*, 2864–2889.

- (16) Manalastas-Cantos, K.; Konarev, P. V.; Hajizadeh, N. R.; Kikhney, A. G.; Petoukhov, M. V.; Molodenskiy, D. S.; Panjkovich, A.; Mertens, H. D. T.; Gruzinov, A.; Borges, C.; Jeffries, C. M.; Svergun, D. I.; Franke, D. *ATSAS 3.0: expanded functionality and new tools for small-angle scattering data analysis. J. Appl. Crystallogr.* **2021**, *54*, 343–355.
- (17) Konarev, P. V.; Volkov, V. V.; Sokolova, A. V.; Koch, M. H. J.; Svergun, D. I. *PRIMUS: a Windows PC-based system for small-angle scattering data analysis. J. Appl. Crystallogr.* **2003**, *36*, 1277–1282.
- (18) Guinier, A. Studies on transformation of *Escherichia coli* with plasmids. *Ann. Phys.* **1939**, *12*, 161–237.
- (19) Svergun, D. I. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* **1992**, *25*, 495–503.
- (20) Svergun, D. I.; Petoukhov, M. V.; Koch, M. H. Determination of Domain Structure of Proteins from X-Ray Solution Scattering. *Biophys. J.* **2001**, *80*, 2946–2953.
- (21) Petoukhov, M. V.; Svergun, D. I. Global Rigid Body Modeling of Macromolecular Complexes against Small-Angle Scattering Data. *Biophys. J.* **2005**, *89*, 1237–1250.
- (22) Kozin, M. B.; Svergun, D. I. Automated matching of high- and low-resolution structural models. *J. Appl. Crystallogr.* **2001**, *34*, 33–41.
- (23) Svergun, D.; Barberato, C.; Koch, M. H. J. *CRY SOL* – a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *J. Appl. Crystallogr.* **1995**, *28*, 768–773.
- (24) Notredame, C.; Desmond, G.; Jaap, H. T-coffee: a novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **2000**, *302*, 205–217.

- (25) Madeira, F.; Park, Y. m.; Lee, J.; Buso, N.; Gur, T.; Madhusoodanan, N.; Basutkar, P.; Tivey, A.; Potter, S.; Finn, R.; Lopez, R. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* **2019**, *47*, W636–W641.
- (26) Waterhouse, A. M.; Procter, J. B.; Martin, D. M. A.; Clamp, M.; Barton, G. J. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **2009**, *25*, 1189–1191.
- (27) Basu, S.; Wallner, B. DockQ: A Quality Measure for Protein-Protein Docking Models. *PLoS ONE* **2016**, *11*, 1–9.
- (28) Porod, G. Die Röntgenkleinwinkelstreuung von dichtgepackten kolloiden Systemen. *Kolloid-Zeitschrift* **1952**, *125*, 51–57.
- (29) Fischer, H.; de Oliveira Neto, M.; Napolitano, H. B.; Polikarpov, I.; Craievich, A. F. Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale. *J. Appl. Crystallogr.* **2010**, *43*, 101–109.
- (30) Hajizadeh, N. R.; Franke, D.; Jeffries, C. M.; Svergun, D. I. Consensus Bayesian assessment of protein molecular mass from solution X-ray scattering data. *Sci. Rep.* **2018**, *8*, 1–13.
- (31) Petoukhov, M. V.; Svergun, D. I. Ambiguity assessment of small-angle scattering curves from monodisperse systems. *Acta Crystallogr. D* **2015**, *71*, 1051–1058.
- (32) PyMol, The PyMOL Molecular Graphics System, Version 2.0. 2015; Schrödinger, LLC.

## C.1 Further Manuscripts

**Effects of dimerization and membrane binding on GABARAP dynamics: a simulation study**

*Xue Wang, Wibke Schumann, and Birgit Strodel*  
in preparation

**Molecular dynamics simulations of protein aggregation: protocols for simulation setup and analysis with Markov state models and transition networks**

*Suman Samantray, Wibke Schumann, Alexander-Maurice Illig, Martín Carballo-Pacheco, Arghardwip Paul, Bogdan Barz, and Birgit Strodel*  
bioRxiv, <https://doi.org/10.1101/2020.04.25.060269>

## C.2 MD Parameters

### C.2.1 NVT Equilibration

Tab. C.1.: NVT equilibration parameter file

variable	value	comment
define	-DPOSRES	position restrain the protein
integrator	md	leap-frog integrator
nsteps	50000	0.02 ps * 50000 = 100 ps
dt	0.002	2 fs
nstcomm	100	Remove relative motion of protein/bilayer/solvent/ions
comm-mode	Linear	Remove relative motion
comm-grps	Protein Non-Protein	Remove relative motion
nstxout	1000	save coordinates every 2 ps
nstvout	1000	save velocities every 2 ps
nstenergy	1000	save energies every 2 ps
nstlog	1000	update log file every 2 ps
continuation	no	first dynamics run
constraint_algorithm	LINCS	holonomic constraints
constraints	all-bonds	all bonds (even heavy atom-H bonds) constrained
lines_iter	1	accuracy of LINCS
lincs_order	6	also related to accuracy
cutoff-scheme	Verlet	pair list with buffering
ns_type	grid	search neighboring grid cells
nstlist	20	10 fs
rlist	1.2	short-range neighborlist cutoff (in nm)
pbc	xyz	3-D PBC
coulombtype	PME	Particle Mesh Ewald for long-range electrostatics
pme_order	4	cubic interpolation
fourierspacing	0.16	grid spacing for FFT
rcoulomb	1.2	short-range electrostatic cutoff (in nm)
rvdw	1.2	short-range van der Waals cutoff (in nm)
DispCorr	EnerPres	account for cut-off vdW scheme
tcoupl	V-rescale	
tc-grps	Protein Non-Protein	

variable	value	comment
tau_t	0.1 0.1	time constant, in ps
ref_t	298 298	reference temperature, one for each group, in K
pcoupl	no	no pressure coupling in NVT
gen_vel	yes	assign velocities from Maxwell distribution
gen_temp	298	temperature for Maxwell distribution
gen_seed	-1	generate a random seed

## C.2.2 NpT Equilibration

**Tab. C.2.:** NpT equilibration parameter file

variable	value	comment
define	-DPOSRES	position restrain the protein
integrator	md	leap-frog integrator
nsteps	500000	0.002 ps * 500000 = 1000 ps (1 ns)
dt	0.002	2 fs
nstcomm	1	
comm-mode	Linear	
comm-grps	Protein Non-Protein	
nstxout	5000	save coordinates every 10 ps
nstvout	5000	save velocities every 10 ps
nstenergy	5000	save energies every 10 ps
nstlog	5000	update log file every 10 ps
continuation	yes	Restarting after NVT
constraint_algorithm	LINCS	holonomic constraints
constraints	all-bonds	all bonds (even heavy atom-H bonds) constrained
lincs_iter	1	accuracy of LINCS
lincs_order	6	also related to accuracy
cutoff-scheme	Verlet	pair list with buffering
ns_type	grid	search neighboring grid cells
nstlist	20	10 fs
rlist	1.2	short-range neighborlist cutoff (in nm)
pbc	xyz	3-D periodic boundary conditions
coulombtype	PME	Particle Mesh Ewald for long-range electrostatics
pme_order	4	cubic interpolation
fourierspacing	0.16	grid spacing for FFT
rcoulomb	1.2	short-range electrostatic cutoff (in nm)
rvdw	1.2	short-range van der Waals cutoff (in nm)
DispCorr	EnerPres	account for cut-off vdW scheme
tcoupl	v-rescale	More accurate thermostat
tc-grps	Protein Non-Protein	two coupling groups
tau_t	0.1 0.1	time constant, in ps
ref_t	298 298	reference temperature, one for each group, in K
pcoupl	Berendsen	Pressure coupling on in NPT

variable	value	comment
pcoupltype	isotropic	uniform scaling of x-y box vectors, independent z
tau_p	1.0	time constant, in ps
ref_p	1.0	reference pressure, x-y, z (in bar)
compressibility	4.5e-5	isothermal compressibility, $bar^{-1}$
refcoord-scaling	all	Scale all coordinates when scaling the box volume
nstpcouple	10	Pressure coupling frequency
gen_vel	no	Velocity generation is off

## C.2.3 Production Run

**Tab. C.3.:** Production run parameter file

variable	value	comment
integrator	md	leap-frog integrator
nsteps	500000000	dt * nsteps = 1000 ns
dt	0.002	2 fs
nstcomm	100	
comm-mode	Linear	
comm-grps	Protein Non-Protein	
nstxout	0	save coordinates every 0 ps
nstvout	0	no trr is generated
nstxtcout	10000	xtc every 20 ps
nstenergy	1000	save energies every 2 ps
nstlog	1000	update log file every 2 ps
continuation	yes	Restarting after NPT
constraint_algorithm	LINCS	holonomic constraints
constraints	all-bonds	all bonds constrained
lincs_iter	1	accuracy of LINCS
lincs_order	6	for higher time step
cutoff-scheme	Verlet	pair list with buffering
ns_type	grid	search neighboring grid cells
nstlist	20	every 20 steps
rlist	1.2	SR neighborlist cutoff [nm]
pbc	xyz	Periodic boundary conditions
coulombtype	PME	Particle Mesh Ewald
pme_order	4	cubic interpolation
fourierspacing	0.16	grid spacing for FFT
rcoulomb	1.2	SR electrostatic cutoff [nm]
rvdw	1.2	SR vdW cutoff [nm]
DispCorr	EnerPres	account for cut-off vdW scheme
tcoupl	Nose-Hoover	
nhchainlength	1	for Nosé-Hoover chain
tc-grps	Protein Non-Protein	
tau_t	0.8 0.8	tau_t > 20*nsttcouple*dt
ref_t	300 300	temperature [K] for each group

variable	value	comment
nsttcouple	10	Frequency of temperature coupling
pcoupl	Parrinello-Rahman	Pressure coupling on in NpT
pcoupltype	isotropic	
tau_p	8.0	[ps] tau_p > 10 tau_t
ref_p	1.0	reference pressure, x-y, z [bar]
compressibility	4.5e-5	isothermal compressibility, bar <sup>-1</sup>
refcoord_scaling	all	
nstpcouple	10	Pressure coupling frequency
gen_vel	no	Velocity generation is off

## C.2.4 HREMD Parameter File

**Tab. C.4.:** HREMD production run parameter file

variable	value	comment
integrator	md	leap-frog integrator
nsteps	50000000	dt * nsteps = 100 ns
dt	0.002	2 fs
nstcomm	100	
comm-mode	Linear	
comm-grps	Protein Non-Protein	
nstxout	0	save coordinates every 0 ps
nstvout	0	no trr is generated
nstxtcout	10000	xtc every 20 ps
nstenergy	1000	save energies every 2 ps
nstlog	1000	update log file every 2 ps
continuation	yes	Restarting after NPT
constraint_algorithm	LINCS	holonomic constraints
constraints	all-bonds	all bonds constrained
lincs_iter	1	accuracy of LINCS
lincs_order	6	for higher time step
cutoff-scheme	Verlet	pair list with buffering
ns_type	grid	search neighboring grid cells
nstlist	20	every 20 steps
rlist	1.2	SR neighborlist cutoff [nm]
pbc	xyz	Periodic boundary conditions
coulombtype	PME	Particle Mesh Ewald
pme_order	4	cubic interpolation
fourierspacing	0.16	grid spacing for FFT
rcoulomb	1.2	SR electrostatic cutoff [nm]
rvdw	1.2	SR vdW cutoff [nm]
DispCorr	EnerPres	account for cut-off vdW scheme
tcoupl	v_rescale	best for Plumed
nhchainlength	1	for Nosé-Hoover chain
tc-grps	Protein Non-Protein	
tau_t	0.8 0.8	tau_t > 20*nsttcouple*dt
ref_t	300 300	temperature [K] for each group

variable	value	comment
nsttcouple	10	Frequency of temperature coupling
pcoupl	Parrinello-Rahman	Pressure coupling on in NpT
pcoupltype	isotropic	
tau_p	8.0	[ps] tau_p > 10 tau_t
ref_p	1.0	reference pressure, x-y, z [bar]
compressibility	4.5e-5	isothermal compressibility, bar <sup>-1</sup>
refcoord_scaling	all	
nstpcouple	10	Pressure coupling frequency
gen_vel	no	Velocity generation is off

## C.2.5 Pulling and USMD Parameter File

**Tab. C.5.:** pulling/USMD production run parameter file

variable	value	comment
define	-DPOSRES_G	
integrator	md	leapfrog algorithm
dt	0.004	4 fs with vsites
nsteps	25000000	100 ns
nstcomm	100	
comm-mode	Linear	
comm-grps	Protein Non-Protein	
nstxout	0	save coordinates every 0 ps
nstvout	0	save velocities every 0 ps
nstxtcout	5000	xtc compressed trajectory output every 20 ps
nstenergy	500	save energies every 2 ps
nstlog	500	update log file every 2 ps
continuation	yes	Restarting after NPT
constraint_algorithm	lincs	holonomic constraints
constraints	all-bonds	
lincs_iter	1	accuracy of LINCS
lincs_order	6	for higher time step
cutoff-scheme	Verlet	pair list with buffering
ns_type	grid	search neighboring grid cells
nstlist	20	every 20 steps
rlist	1.2	SR neighborlist cutoff [nm]
coulombtype	PME	Particle Mesh Ewald
pme_order	4	cubic interpolation
fourierspacing	0.16	grid spacing for FFT
rcoulomb	1.2	SR electrostatic cutoff [nm]
rvdw	1.2	SR vdW cutoff [nm]
DispCorr	EnerPres	account for cut-off vdW scheme
rcoulomb	1.2	
rvdw	1.2	
DispCorr	EnerPres	account for cut-off vdW scheme
tcoupl	v_rescale	best for Plumed

variable	value	comment
nhchainlength	1	for Nosé-Hoover chain
tc-grps	Protein Non-Protein	
tau_t	0.8 0.8	$\tau_t > 20 * \text{nsttcouple} * dt$
ref_t	300 300	temperature [K] for each group
nsttcouple	10	Frequency of temperature coupling
pcoupl	Parrinello-Rahman	Pressure coupling on in NpT
pcoupltype	isotropic	
tau_p	8.0	[ps] $\tau_p > 10 \tau_t$
ref_p	1.0	reference pressure, x-y, z [bar]
compressibility	4.5e-5	isothermal compressibility, $\text{bar}^{-1}$
refcoord_scaling	all	
nstpcouple	10	Pressure coupling frequency
gen_vel	no	Velocity generation is off
pbcs	xyz	
pull	yes	
pull_ncoords	1	only one reaction coordinate
pull_ngroups	2	two groups defining one reaction coordinate
pull_group1_name	group_A	
pull_group2_name	group_B	
pull_coord1_type	umbrella	harmonic potential
pull_coord1_geometry	distance	
pull_coord1_dim	Y Y Y	
pull_coord1_groups	1 2	
pull_coord1_start	yes	define initial COM distance $> 0$
pull_coord1_rate	0.0	restrain in place=0, pull= $> 0$ nm/ps
pull_coord1_k	3	varies from 3 to 6000 $\text{kJ mol}^{-1} \text{nm}^{-2}$ (USMD) or e.g. 1000 $\text{kJ mol}^{-1} \text{nm}^{-1}$ for pulling

## C.2.6 MARTNI Parameter Files

**Tab. C.6.:** MARTINI production run parameter file

variable	value	comment
integrator	md	
tinit	0.0	
dt	0.020	better energy conservation
nsteps	50000000	1 $\mu$ s
nstxout	5000	
nstvout	5000	
nstfout	5000	
nstlog	5000	
nstenergy	5000	
nstxout-compressed	5000	
compressed-x-precision	100	
cutoff-scheme	Verlet	
nstlist	20	
ns_type	grid	
pbc	xyz	
verlet-buffer-tolerance	0.005	
epsilon_r	15	
coulombtype	reaction-field	
rcoulomb	1.1	
vdw_type	cutoff	
vdw-modifier	Potential-shift-verlet	
rvdw	1.1	
tcoupl	v-rescale	
tc-grps	protein membrane solute	
tau_t	1.0 1.0 1.0	
ref_t	303 303 303	
Pcoupl	Parrinello-rahman	
Pcoupltype	semiisotropic	membrane bilayer
tau_p	12.0	
compressibility	3e-4 3e-4	
ref_p	1.0 1.0	
gen_vel	no	

variable	value	comment
refcoord_scaling	all	

```

1  1
2  71 302
3  3
4  1 70
5  303 358
6  661 693
7  1
8  359 660

```

**Listing C.1:** ELNEDIN file for BDLP

## C.3 Scripts

### C.3.1 Hinge Motion

```

1  # Example Usage: vmd -dispdev text -e angle.tcl -args
    ↪ md_protein.pdb md_protein.xtc ref.pdb
2  set input1 [lindex $argv 0]
3  set input2 [lindex $argv 1]
4  set ref [lindex $argv 2]
5
6  mol new $input1
7  animate delete beg 0 end 0
8  animate read xtc $input2 waitfor all
9  set nf [molinfo 0 get numframes]
10
11 mol new $ref
12 mol top 0
13
14 proc align { rmolid smolid2 seltext1 seltext2 } {
15     set ref [atomselect $rmolid $seltext1 frame 0]
16     set sel [atomselect $smolid2 $seltext2]
17     set all [atomselect $smolid2 all]
18     set n [molinfo $smolid2 get numframes]
19

```

```

20   for { set f 0 } { $f < $n } { incr f } {
21     $sel frame $f
22     $all frame $f
23     $all move [measure fit $sel $ref]
24   }
25   $ref delete
26   $all delete
27   $sel delete
28   return
29 }
30
31 align 1 0 "resid 1 to 300 and name CA" "resid 1 to 300 and name
    ↪ CA"
32
33 set fil1 [open "hinge1_tcl.dat" w]
34 set fil2 [open "dist_tcl.dat" w]
35 set fil3 [open "hinge2_tcl.dat" w]
36 set fil4 [open "distxyz_tcl.dat" w]
37
38 for {set j 0} {$j<$nf} {set j [expr $j+1]} {
39   set cnt 0
40   if { [ expr $j\%1000 == 0 ] } {
41     puts "frame $j"
42   }
43
44   # Go to a specific frame
45   animate goto $j
46   display update
47   set ang [measure angle [[atomselect 0 "name CA and (residue 4
    ↪ or residue 359 or residue 587) "] get index]]
48   puts $fil1 "$j $ang"
49   set ang2 [measure angle [[atomselect 0 "name CA and (residue
    ↪ 331 or residue 309 or residue 206) "] get index]]
50   puts $fil3 "$j $ang2"
51
52   set sel1 [atomselect 0 "name CA and resid 580" frame 0]
53   set coord1 [$sel1 get {x y z} ]
54   set sel2 [atomselect 0 "name CA and resid 580"]
55   set coord2 [$sel2 get {x y z} ]

```

```

56     set vec [vecsub [lindex $coord2 0] [lindex $coord1 0]]
57     set dist [veclength [vecsub [lindex $coord2 0] [lindex $coord1
    ↪ 0]]]
58     set distx [lindex $vec 0]
59     set disty [lindex $vec 1]
60     set distz [lindex $vec 2]
61     puts $fil4 "$j $distx $disty $distz $dist"
62     puts $fil2 "$j $dist"
63 }
64 close $fil1
65 close $fil2
66 close $fil3
67 close $fil4
68 exit

```

**Listing C.2:** Hinge Motion with tcl

## C.3.2 Plumed Pulling

```
1  g1: GROUP ATOMS=890
2  g2: GROUP ATOMS=1136
3  # drag this
4  dist1: DISTANCE ATOMS=890,11360
5  # the movingrestraint
6  restraint: ...
7  MOVINGRESTRAINT
8  ARG=dist1
9  AT0=0.5 STEP0=0 KAPPA0=0
10 AT1=0.8 STEP1=2000 KAPPA1=1000
11 AT2=1.0 STEP2=4000 KAPPA2=1000
12 AT3=1.2 STEP3=6000 KAPPA3=1000
13 #...
14 AT46=9.8 STEP46=220000 KAPPA46=1000
15 AT47=10.0 STEP47=240000 KAPPA47=1000
16 ...
17 # monitor the two variables and various restraint outputs
18 PRINT STRIDE=20 ARG=* FILE=COLVAR
```

**Listing C.3:** plumed pulling

## Colophon

This thesis was typeset with  $\text{\LaTeX}2_{\epsilon}$ . It uses the *Clean Thesis* style developed by Ricardo Langner. The design of the *Clean Thesis* style is inspired by user guide documents from Apple Inc.

Download the *Clean Thesis* style at <http://cleanthesis.der-ric.de/>.

