The effects of dimerization and membrane binding on protein dynamics: the cases of GABARAP and guanylate binding proteins

Inaugural-Dissertation
zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Xue Wang
aus China

Jülich, September 2022

aus dem Institut für Biologische Informationsprozesse: Strukturbiochemie (IBI-7)
des Forschungszentrum Jülich

Gedruckt mit der Genehmigung der

Mathematisch-Naturwissenschaftlichen Fakultät der

Heinrich-Heine-Universität Düsseldorf

Berichterstatter:

[1] Prof. Dr. Birgit Strodel

[2] Prof. Dr. Dieter Willbold

Tag der mündlichen Prüfung: 26.10.2022

Statement of authorship

I, Xue Wang, hereby certify that the works shown here are original and the result of my investigation to the best of my knowledge. The ideas and work of others were fully acknowledged and referenced in my paper. The thesis does not contain results that have been published elsewhere, nor does it contain materials that have been used to obtain a degree or certificate from this or other institutions or universities. I mentioned my contributions where the results have collaborated with others.

Düsseldorf, September 2022

Publications

Manuscripts that are part of the thesis:

X. Wang, W. Schumann, and B. Strodel. Effects of dimerization and membrane binding on GABARAP dynamics: a simulation study. (In preparation)

J. Loschwitz, X. Wang, and B. Strodel. Domain motions, dimerization, and membrane interactions of the murine guanylate binding protein 2. bioRxiv preprint. https://doi.org/10.1101/2022.04.27.489784.

L. Legewie, J. Loschwitz, N. Steffens, M. Prescher, X. Wang, S.H.J. Smits, L. Schmitt, B. Strodel, D. Degrandi, K. Pfeffer. Biochemical and structural characterization of murine GBP7, a guanylate binding protein with an elongated C-terminal tail. Biochem. J. 2019, 476, 3161-3182.

Publications beyond the scope of this thesis:

B.J. Taiwo, O.O. Olubiyi, X. Wang, F.A. Fisusi, A. Ganiyu, F.R. van Heerden, B. Strodel. Schistosomiasis: Snail-vector control, molecular modelling and dynamic studies of bioactive N-acetylglycoside saponins from Tetrapleura tetraptera. Comput. Biol. Chem., 77: 363-372 (2018)

D. Petrovic, X. Wang, B. Strodel. How accurately do force fields represent protein side chain ensembles? Proteins: Struct. Func. Bioinf., 86: 935-944 (2018)

Contents

Acknowledgments

When I decided to do a Ph.D. years ago, I didn’t know many things about a Ph.D. life and the difficulties I must face. But I’m not alone, many people give me so much help and support on the way to the goal. They deserve my sincere appreciation.

First and foremost, I want to give my sincerest appreciation to my supervisor, Prof. Birgit Strodel. She helps me develop and realize my ideas, supports me academically, and gives me so much patience, support and help when I feel down and depressed. It has been a great pleasure working with you.

Thanks Dr. Oliver Schillinger, Dr. Bogdan Barz, Dr. Michael Owen and Dr. Dusan Petrovic. They giving me so much help when I arrived in Germany and started my Ph.D. work in Forschungszentrum Jülich.

Sincerely thanks to my collaborators/colleagues: Jennifer, Wibke, Irina, Larissa, Dr. Philipp Neudecker and Prof. Klaus Pfeffer. It’s so nice to work together on these interesting projects. I will miss my dear colleagues from Birgit’s group: Jide, Jennifer, Wibke, Hebah, Anna, Suman, Maryam, Feng and Lara. They are all very warm and kind, maintain a cheerful atmosphere in the office. Special thanks to Jide for his valuable suggestion on my thesis and special thanks to Hebah and Ms. Stobbe for the help that they gave to me on the thesis printing and submitting.

I would like to express my sincere gratitude to project SFB1208 which funded by Deutsche Forschungsgemeinschaft, the great lectures every week and the progress reports every year from the project members really widen my knowledge and inspired me so much.

I would like to acknowledge the financial support from China Scholarship Council, SFB project and Institute of Biological Information Processing (IBI)-Structural Biochemistry (IBI-7). The computing time for this thesis was granted on the supercomputer JURECA at Forschungszentrum Jülich.

Last but not least, I own my deepest gratitude to my parents and sister, they have given me the greatest support and help during my education. Even though I have already left Germany, the years I have stayed will always be in my mind.

Abstract

As an important part of the biological systems, proteins usually reply to specific cellular responses by binding to other biomolecules (lipids, carbohydrates, nucleic acids, and proteins) and act synergistically. To study the actions in more detail, in silico experiments were introduced years ago and form the basis of the current thesis. Apart from acting as monomers, proteins can self-associate to form dimers or higher-order oligomers. This homodimerization phenomenon is important for regulation and catalysis. Moreover, in addition to functioning in the cytoplasm, some proteins can bind to lipid membranes, as monomers or oligomers, or even act together with other proteins to function as membrane channels, promote membrane elongation/fusion, or damage the membrane. The exercise of these functions is closely related to the dynamics of the proteins. Two such proteins or protein classes are the gamma-aminobutyric acid receptor-associated protein (GABARAP) and guanylate binding proteins (GBPs), which are proteins known to oligomerize and bind to lipid membranes. GABARAP and two of the GBPs, namely the murine GBP2 and GBP7, are studied here using various in silico methods. First, since not all protein crystal structures have been resolved by experimental means, homology modeling methods were used to construct the unknown protein structures. Dimer structures of GABARAP and mGBP2 were also predicted based on the information available from previous experimental research. The structural dynamics and stabilities, including their monomer and dimer form, were then investigated using molecular dynamics simulations. Finally, the interplay between the proteins and lipid membranes following their membrane binding was studied, to gain more detailed information on the diversified roles that they play in the cell. The most relevant findings of this thesis work are: i) GABARAP seems to be a very stable protein with a limited degree of conformational flexibility, while mGBP2 and mGBP7 are very flexible exhibiting a large-scale hinge motion; ii) the dimer models that were constructed for GABARAP based on crystallographic data are not stable in solution, requiring more work to determine the GABARAP dimer structure; iii) GABARAP and mGBP2 bind very stably to lipid membranes via their lipid anchors while only few amino-acid residues add to further protein-membrane interactions, and iv) membrane binding of mGBP7 is enabled via its C-terminal tail involving 49 amino-acid residues which was confirmed by experiments that were motivated by the current simulation results.

Abbreviations

Molecular dynamics	MD
Gamma-aminobutyric acid receptor-associated protein	GABARAP
Autophagy-related proteins	ATG
Light-chain 3	LC3
Guanosine triphosphate binding protein	GBP
Guanosine triphosphate	GTP
Guanosine diphosphate	GDP
Guanosine diphosphate	GMP
Glycosylphosphatidylinositol-linked proteins	GPI
Interferon-γ	IFNγ
Parasitophorous vacuole	PV
Coarse-grained	CG
Simple point charge	SPC
Extended simple point charge	SPCE
Three-site transferable intermolecular potential water molecules	TIP3PEW
Replica exchange molecular dynamics simulations	REMD
Hamiltonian replica exchange molecular dynamics simulations	H-REMD
Temperature replica exchange molecular dynamics simulations	T-REMD
Nuclear magnetic resonance	NMR
Cryogenic electron microscopy	cryo-EM
Phosphatidylcholine	PC
Phosphatidylserine	PS
Phosphatidylethanolamine	PE
Phosphatidylglycerol	PG
Sphingomyeline	SM

Root mean square deviation	RMSD
Root mean square fluctuation	RMSF
All-atom molecular dynamics	AAMD
C-terminal	CT
Periodic boundary conditions	PBC
Molecular mechanics	MM

1 Introduction

1.1 Importance of homodimerization

Proteins rarely act standalone in biological systems but bind to other biomolecules to reply to specific cellular responses. These biomolecules could be lipids, nucleic acids, proteins etc. The oligomeric proteins are typically dimers, trimers, tetramers, pentamers, hexamers, heptamers, and octamers (Figure 1). Dimers, which are proteins made up of two subunits take up the highest proportion among them. Except for binding to other proteins, a large number of proteins self-associate to form dimers or higher-order oligomers, both in relative isolation and within protein interaction networks and cascades [1][2]. Of all oligomeric proteins, there is a surprisingly high number of proteins made up of two subunits. The protein subunit interaction (either homodimer or heterodimer) is an important phenomenon in regulation and catalysis. And more notably, most of these proteins are made up of two identical subunits (Figure 1). Thousands of such interactions are theoretically possible in a combinatorial manner.

Dimerization and oligomerization can confer several different structural and functional advantages to proteins, including improved stability, control over the accessibility and specificity of active sites, and increased complexity. However, unwanted oligomerization of proteins can also lead to the formation of pathogenic structures. Important features such as surface hydrophobicity, internal empty or water-filled cavities and the distribution of amino acid residues have often been found to play a significant role in the protein folding and aggregation processes [3][4].

In 1965, Monod, Wyman and Changeux first provided the possible advantages of being homodimeric in their paper which describe the model of enzymes' allosteric transitions [5]. Koshland further confirmed that hypothesis by experiment on the folding of mixtures of different oligomers in vitro [6]. Interactions at the subunit binding site are usually highly specific due to the evolutionary selection. Studies on oligomers showed that the separation of dimeric protein subunits may affect the conformation of their monomer structure. This means that proteins' quaternary interactions can affect the structure of each monomer [7][8]. Except the advantages mentioned above, there're also other advantages such as genetic saving (size identically oligomeric protein and monomer protein, the oligomeric protein will save more genes); functional gain (improvement in the catalytic action); structural advantage (oligomeric process could induce subtle conformational changes at the active site to activate the monomeric subunits) [9].

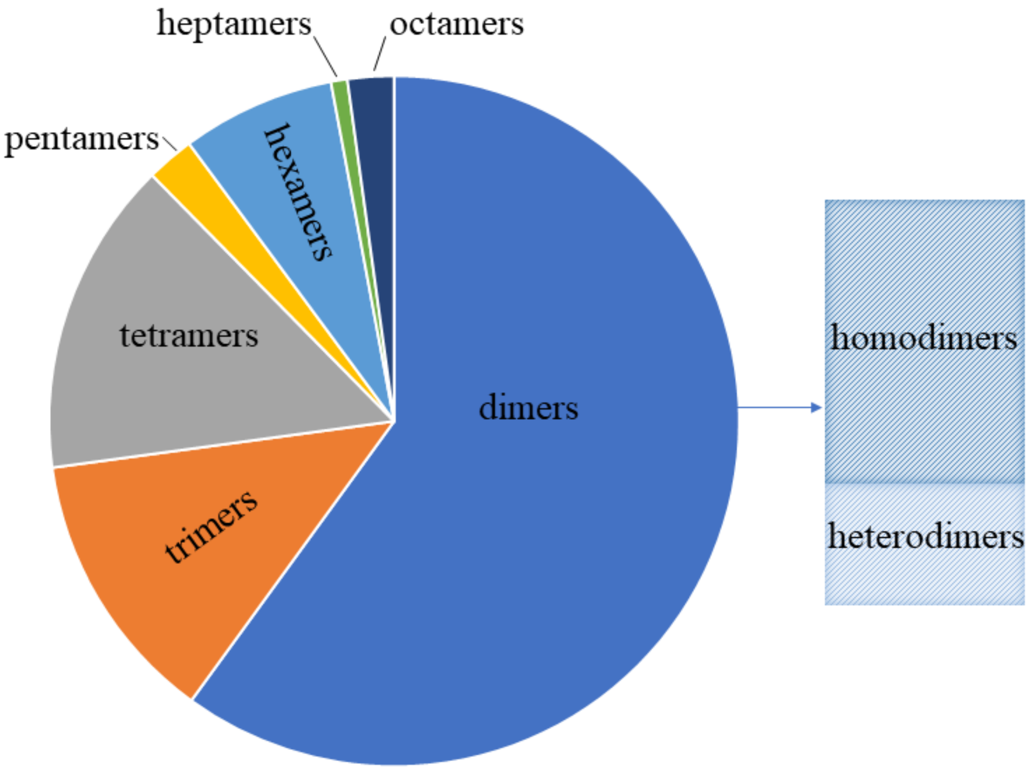


Figure 1. Percentage distribution of oligomeric proteins (dimers, trimers, tetramers, pentamers, hexamers, heptamers and octamers). The data were collected from the Protein Data Bank (PDB) [10].

1.2 Importance of membrane binding in general

Membrane protein refers to the general term for proteins that can bind or integrate into the membrane of cells or organelles. More than half of the proteins in cells can bind to the membrane in different forms. According to the different binding strength and position of the membrane, membrane proteins can be divided into three categories: integral membrane proteins, peripheral membrane proteins and lipid-anchored proteins.

Integral membrane proteins are also called intrinsic membrane proteins. These proteins are embedded in the bilayer of phospholipids, which are always bound to the membrane and can often cross the membrane multiple times to form a transmembrane channel.

Intrinsic membrane proteins can also be subdivided according to their relationship with bimolecular membranes: integral bitopic proteins, integral polytopic proteins and integral monotopic proteins (Figure 2). Integral bitopic and polytopic proteins, also known as transmembrane proteins (as the name implies), are proteins that span both ends of the membrane. The transmembrane region is often α-helical, and some of the transmembrane proteins form a barrel-packed transmembrane channel in β-folds. Integral monotopic proteins only bind to the membrane from one direction (outside or inside the membrane). Some of the membrane proteins are only partially inserted into the membrane, i.e., they do not cross the membrane.

Peripheral membrane proteins (extrinsic membrane proteins) are proteins that can temporarily bind to membranes or internal membrane proteins, mainly through hydrophobic, electrostatic and other non-covalent interactions. This binding can be done by adding polar reagents, such as high pH or high salt.

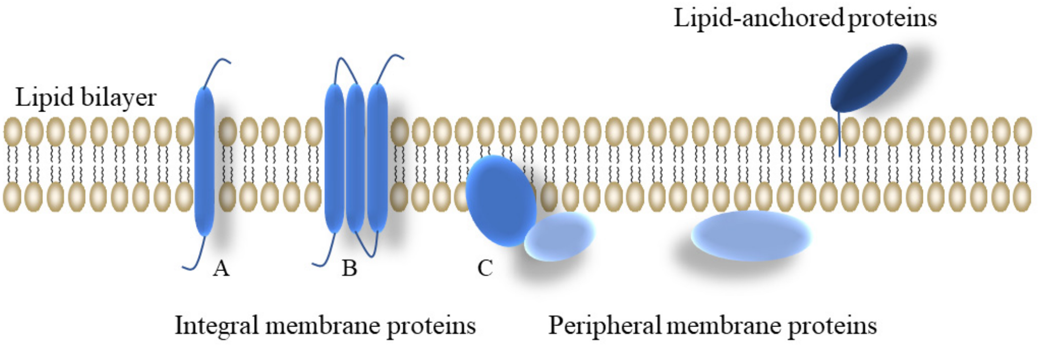


Figure 2. Structural classification of membrane proteins. Integral, peripheral and lipid anchored membrane proteins were colored by blue, light blue, dark blue, respectively.

Lipid-anchored proteins are membrane proteins that are covalently bonded to lipids and inserted into the membrane. They can be found both on the inner and outer surface of the bilayer membrane. There are three main types of lipid-anchored proteins: prenylated proteins, fatty acylated proteins and glycosylphosphatidylinositol-linked (GPI-linked) proteins (GPI) [11].

Prenylated proteins can be covalently attached to a hydrophobic isoprenoid group, such as the farnesyl or geranylgeranyl group. These proteins contain a prenylation motif called “CaaX box”, which is the most common prenylation site in proteins. The “C” in the CaaX box represents the cysteine that is becoming prenylated, “a” means any aliphatic amino acid and the X determines the type of prenylation that will occur. The protein will be farnesylated via the farnesyltransferase enzyme when “X” is alanine, methionine, serine or glutamine. However, when “X” is a leucine then the protein will be geranylgeranylated via the geranylgeranyl transferase enzyme [12][13]. Farnesyl or geranylgeranyl is covalently attached to the protein via thioether linkages at the cysteine residue of the CaaX box which is near the C-terminal of the protein.

Fatty acylated proteins are post-translationally modified proteins where fatty acids become covalently attached to certain amino acid residues. There are two kinds of fatty acids: saturated myristic acid (14 carbon atoms) and palmitic acid (16 carbon atoms), one or both of them can covalently bind to proteins [14][15].

The GPI anchor is attached to a protein by an amide bond to the C-terminal carboxyl group of the protein [16].

Different types of lipid-anchored proteins have different functions, they are found in a wide range of organisms, from bacteria and yeast to humans. The role of lipid anchored proteins is not well understood. However, it is believed that these proteins may play a role in cellular function such as membrane trafficking and vesicle formation. Now more lipid modified proteins are being identified and more information of the effects of lipid attachment on protein structure and function will be deduced [17].

1.3 Lipid-anchored membrane proteins: GABARAP and GBPs

1.3.1 GABARAP

With regard to GABARAP, which belongs to the autophagy-related protein 8 (ATG8) family, the phospholipid conjugation occurs during the autophagy process in the presence of ATG3, ATG5 and ATG7 (Figure 3). Autophagy was first mentioned by Christian de Duve in 1963 [18]. The name originates from Greek, with "auto" meaning “self” and "phagy" meaning “to eat”, hence it refers to the process of cells digesting and decomposing their own substances. Autophagy is the major intracellular degradation system. When the cell is in nutrient deficiency, hypoxia, or harbors damaged organelles, cytoplasmic materials are delivered to and degraded in the lysosome to produce decomposition products such as amino acids and monosaccharides for reuse by cells. It plays a key role in maintaining cell homeostasis and helping cells to respond to various stresses and other physiological processes. Therefore, the dysfunction of autophagy is closely related to many major human diseases including cancer, neurodegenerative diseases and metabolic diseases [19][20].

The ATG8 family members of humans and other mammals can be divided into the GABARAP subfamily (including GABARAP, GABARAPL1, GABARAPL2) and the light chain 3 (LC3) subfamily (including LC3A, LC3B, LC3C). GABARAP promotes the autophagosome formation and is essential for autophagosome-lysosome fusion, while the LC3 subfamily plays a less prominent role in these processes. GABARAP subfamily members were identified as primary contributors during the starvation autophagy process. Therefore, the study of the dynamics of GABARAP and GABARAP anchored to membrane could help us better understand how this protein works [21][22].

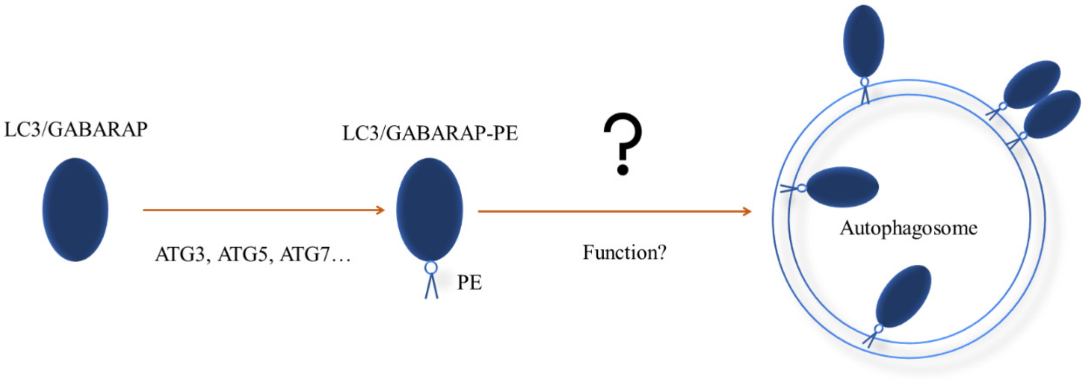


Figure 3. GABARAP binding to autophagosome membranes.

1.3.2 GBPs

Interferon-γ (IFNγ) is an immunomodulatory cytokine that is released by host cells to confront the infection with intracellular bacteria and parasites to activate the transcription of the interferon-stimulated genes [23]. Guanosine triphosphate (GTP) binding proteins (GBPs) belong to the most abundant proteins that are induced by IFNγ [24][25][26]. GBPs belong to dynamins and dynamin-like protein subfamilies, they contain a conserved GTPase-domain that can bind and hydrolyze GTP via GDP to GMP [27][28].

Toxoplasma gondii (*T. gondii*) is a parasitic protozoan that causes toxoplasmosis and is capable of infecting most warm-blooded animals. After invasion into host cells, the tachyzoite form of T. gondii replicates in a specialized vacuole, called the parasitophorous vacuole (PV), which protects the parasite from the host cell lysosomal degradation pathway and serves as a critical transport interface between the parasite and the host cell cytoplasm. Nonetheless, host cells have their own way of finding hidden parasites. For some of the murine GBPs (mGBPs) it was demonstrated that they are highly upregulated upon *Toxoplasma gondii* infection. The genes of the mGBP family mGbp1, mGbp2, mGbp3, mGbp5, and mGbp7 were located on clusters on chromosomes 3, while mGbp4, mGbp6, mGbp8, mGbp9, mGbp10, and mGbp11 were located on clusters on chromosomes 5. Research showed that among these genes, mGBP2 and mGBP7 are the most abundantly expressed proteins after stimulation of different type of murine cells with IFNγ, localize at the PV membrane and then destroy it, and are required for survival of infection with *T. gondii* [25][29].

The presence of the CaaX motif in GBPs enables the lipidation reactions. The motif is not common to all GBPs, only human GBP1, 2, 5 and murine GBP1, 2, 5 contain it (Figure 4). Interestingly, mGBP2 and mGBP5 co-localize with mGBP1 in *T. gondii*-infected cells [30]. The position X of the CaaX motif determines whether a farnesyl (C15, X = Met, Ser, Ala, Gln or Asn, hGBP1 with X = Ser) or geranylgeranyl (C20, X = Leu, Val or Phe, mGBP2 with X = Leu) isoprenoid is linked to the cysteine via a thioether. For complete modification, the last three amino acids (aaX) are cleaved off after prenylation and the modified cysteine becomes methylated. For hGBP1, it was already demonstrated that the GTPase cycle and farnesyl anchor attachment stimulate the polymerization and membrane binding of hGBP1[31]. As mGBP2 contains the CaaX motif as well, it is very interesting to investigate the dimerization of mGBP2 and the dynamics of its binding to the membrane.

Unlike mGBP2 and hGBP1, mGBP7 do not contain the CaaX motif, instead it features a much longer C-terminal sequence. As it has been revealed that both mGBP2 and mGBP7 can localize at the PV membrane, it will be interesting to see the difference of their binding patterns [32]. Before researching this, the primary thing is to obtain the structure of mGBP2 and mGBP7, as to date only for hGBP1 the crystal structure was determined [28].

Based on that hGBP1 structure and by employing the homology modeling method, it was straightforward to determine the structure of mGBP2. For mGBP7, on the other hand, the extra sequence on its C-terminal needed some special consideration, as will be shown in this thesis work.

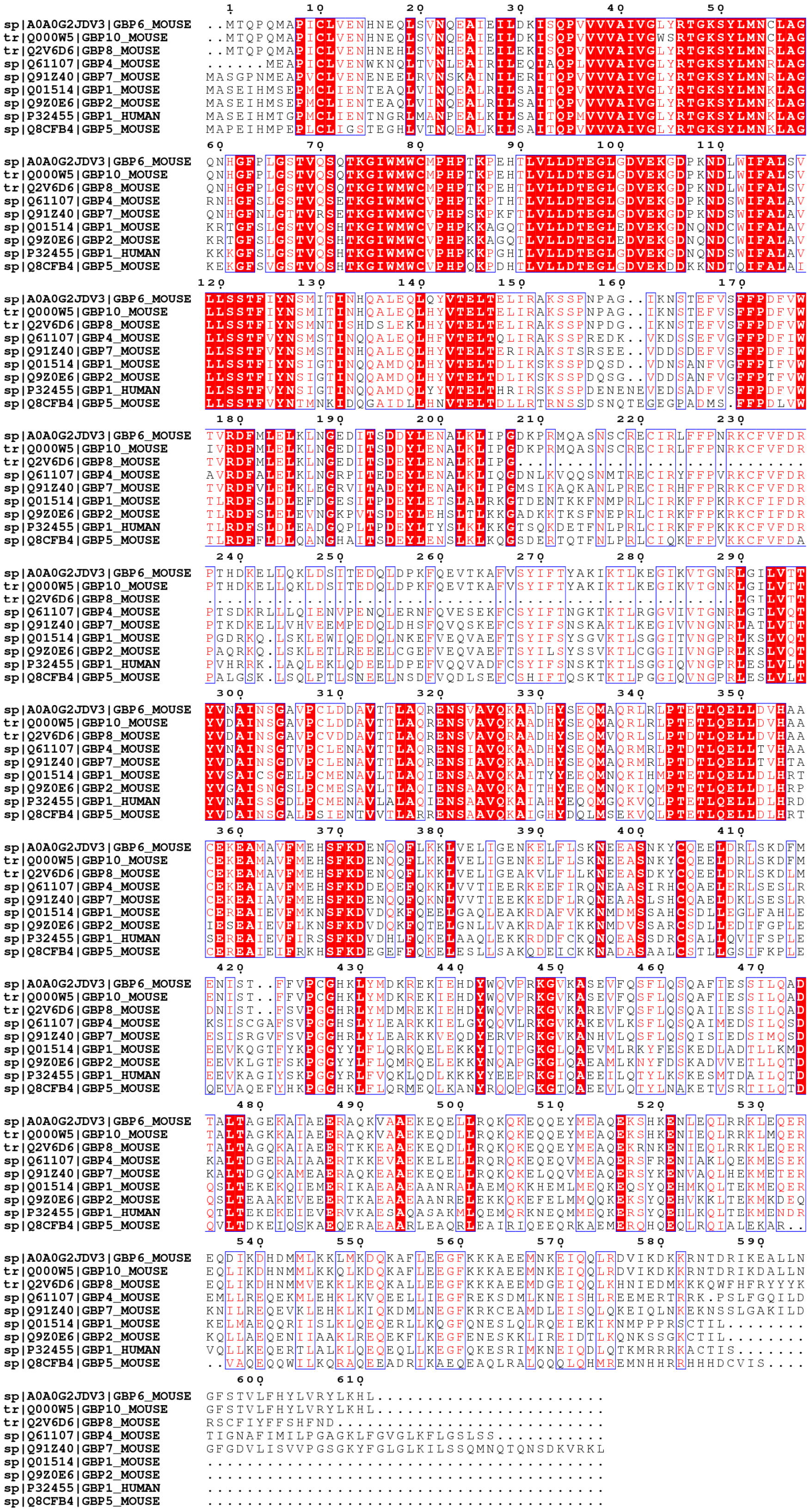


Figure 4. Sequence alignment results of hGBP1 and mGBP1, 2, 4, 5, 6, 7, 8, 10. Sequences were downloaded from UniProt [33], alignment was done using ClustalW, and the figure was created with ESPript3.0 [34].

2 Aims of this thesis

Both GABARAP and GBPs are membrane-binding proteins and play essential roles in biological processes via binding to membranes. However, their functions are totally different. GABARAP is involved in membrane fusion and elongation while GBPs destroy the membrane. For this reason, the binding mode of these proteins to membranes is worth studying. For example, GABARAP and GBPs might bind to the membrane not only as monomer model, but also as an oligomer or even larger multimer. Therefore, the prediction of aggregate models for these proteins and the resulting effects on the membrane are the focus of the current thesis work. To this end, multiple computational methods, in particular molecular dynamics (MD) simulations, combined with experimental data whenever available, are applied to study the dynamics following the dimerization and membrane binding of GABARAP, mGBP2 and mGBP7. The mutual effects between protein dynamics and the membranes are investigated in detail. In addition, residues that contribute the most to the membrane-binding process are identified by determining the binding energy between the proteins and membranes and its decomposition on a per-residue level. The results of the simulations provoked new experiments, as discussed in the thesis.

3 Materials and methods

3.1 Preparation of molecular models

To perform molecular dynamics simulations, the first and most basic task is to obtain starting structures of the molecules to be simulated. For thousands of proteins, the structure can be found in the Protein Data Bank. Most of these structures (≈88%) were determined by X-ray crystallography, about 7% by nuclear magnetic resonance (NMR) spectroscopy, and the remaining structures were resolved with cryogenic electron microscopy (cryo-EM), a technique of increasing importance in protein structure elucidation. However, some protein structures are missing, which in particular applied to membrane proteins as it is more difficult to determine their structures as compared to proteins in solution. Computational prediction of the 3D structure of proteins has been a long-standing challenge over the past 50 years, where a major breakthrough by AlphaFold2 was observed in 2021 [35]. However, before AlphaFold became available and for other systems than proteins, the following modeling methods are available to address this problem.

3.1.1 Homology modeling

The theoretical basis of homology modeling is that proteins from different sources or different biological functions may have similar structures if their sequences are similar. Therefore, homology modeling is a method that predicts the structure of a target protein using a protein with a known structure as a template [36][37]. The process of constructing and predicting protein structure by homology modeling can be broken down into the following steps: template selection, sequence alignment with the target template, construction of the model, and model assessment. Many homology modeling tools, many of them easy to handle, are available, such as SWISS-MODEL [38][39][40], MODELLER [41]，PARMODEL [42]，HHpred [43]，MODBASE [44][45], and PROTEUS [46]. We use SWISS-MODEL to build the protein structures in our work. The missing loops which cannot be created by homology modeling were constructed using Modloop [47][48], which is a web server for automated modeling of loops in protein structures, relies on the loop modeling routine in MODELLER. It predicts the loop conformations by satisfaction of spatial restraints, without relying on a database of known protein structures.

3.1.2 Membrane construction methods

When talking about biological membranes, what we know is that they are very complex entities that contain a variety of biological molecules. These biological molecules usually are lipids, proteins, and sugar moieties. The different functions of membranes are related to the proportions of the biological molecules and their spatial distribution. These differentiated properties depend largely on the cellular origin of the membrane [49] and may even change in time. Therefore, one of the most important tasks for membrane simulations is the construction of an appropriate membrane system. Molecular modeling of membrane is usually simplified, only focused on simulating small bilayer systems that contain just a few lipid moieties. Nonetheless, these systems are considered to be representative models of some aspects of real biological membranes. Therefore, the membranes in a simulation project are generally simple and contain only a small number of different lipid types. However, in biophysical experimental studies involving lipid membranes, the situation is usually the same, i.e., simplified membrane models are used to mimic biological membranes.

To simplify and avoid the flaws in membrane building, many new tools and approaches have been developed during the last few years. Nevertheless, building a membrane for an MD simulation still remains a challenging task. Accompanied with the increased computing power, the complexity and size of the simulated membranes were also augmented. For example, the membrane simulation systems can consist of many lipid moieties, involve more than one bilayer, and usually contain proteins [50][51] or polymers [52]. However, these complex systems are usually not easy to be set up with manual protocols.

To build a membrane for a certain project, the first thing to do is select the membrane composition. Ideally, the information of the membrane composition usually obtained from the experimental data of the systems that are to be mimicked, which including the required lipid components and their desired proportions [49]. However, mimicking the composition of biological membranes is usually not possible, because in most cases it is not known. The second challenge is to determine the so-called force field parameters of lipids, which is a time-consuming process, because not all possible lipids have already been parameterized for all force fields. Therefore, compromises have to be made in the membrane building step, the membrane composition in simulations is usually not fully consistent with the desired membrane composition. Besides, the choice of the force field is often limited to a few options, and the tools that can be used to construct and simulate the membrane are limited as well.

To help researchers easily build membranes based their own projects, new tools have recently started popping up. The common tools can be divided into two categories: web services (CHARMM-GUI [53], MemBuilder [54], LipidBuilder [55]) and distributed applications (for example, Packmol [56] and insane [57]). The lipids are randomly distributed to improve their lateral mixing in all these tools. The advantage of web-based applications is the user-friendly interface, after setting up the parameters step by step, all files needed to perform simulations, including coordinate files, topologies and simulation parameters were generated. However, their disadvantage is the generally limited possibility to make custom changes during the membrane setup, such as adding new lipid types. Moreover, the output can only be used for a limited number of MD protocols and force fields. As not all users are able to overcome these limitations, the web services can only be used if it meets all the requirements of a given project. For example, the newly released LipidBuilder partly solved the lipid-type limitation problem by adding a lipid builder application, but the output files are only compatible with the CHARMM force field and the NAMD software. Membuilder only supplies files in the formats of GROMACS. However, a number of GROMOS (43a1, 43A1-S3, and 53a6) and CHARMM (27 and 36) force fields as well as the Slipids force field are supported. CHARMM-GUI [58] nowadays provides input files for multiple MD packages, including GROMACS, NAMD, Amber, OpenMM and CHARMM with CHARMM36, CHARMM36m and AMBER as atomic force field options [59][60]. The structures of the membrane systems provided by CHARMM-GUI can also be used with other force fields by simply altering the atom ordering if it differs from the CHARMM-GUI output file.

In our research, the membranes were built by CHARMM-GUI, which provides membrane structures with a lateral density already close to its equilibrium value. When the desired lipid does not exist in a construction tool, the tool will be useless as it is impossible to reliably parameterize lipids automatedly for now. However, it is believed that in the near future, a general membrane builder tool will be developed and would be free of all the numerous limitations listed above.

3.1.3 Insertion of membrane proteins

The first simulations of membrane protein systems relied on assembling a membrane around a protein [61][62]. For preparing membrane protein simulations, the next step is to insert the proteins into the membrane, which is also a hard task. Nowadays, there are multiple protein embedding approaches available that can in principle be used. However, the approach chosen should meet multiple criteria, like compatibility with the lipids, protein models and force fields to be used. Just like for the membrane building approaches, there are also web-based tools available for protein embedding. These tools provide simple ways to perform protein embedding tasks. One can also perform MD simulations, in particular or some enhanced MD simulation methods, to fulfil this task. Third, some methods are based on independent software designed specifically for protein embedding purposes.

3.1.3.1 Protein or membrane expand/compress method

In this section, we will introduce two majorly used tools, InflateGRO [63] and g_membed [64]. Using a common scaling factor, the InflateGRO method resorts to lateral lipid translation within the membrane plane to first expand a bilayer, then delete lipids within a cut-off around the protein, followed by a series of alternating steps of compression and energy minimization to bring the system back to its natural dimensions. The usage of InflateGRO is described in detail by J.A. Lemkul in the GROMACS tutorial 2: KALP15 in DPPC [65].

Whereas the previous approach changes the membrane dimensions, g_membed takes the opposite approach, subjecting the protein to compression and expansion steps. The starting point of this approach is a fully solvated membrane system containing the protein in an optimized orientation. The protein is then narrowed by decreasing its width in the *xy*-plane, followed by the removal of overlapping molecules. Then the protein is grown back to its full size in a short MD simulation, thereby pushing the lipids away to optimally accommodate the protein in the membrane.

3.1.3.2 Protein insertion during membrane generation

Except generating a pre-equilibrated membrane-only system, CHARMM-GUI also allows to embed a protein into a membrane during the building process. In practice, the usage of the CHARMM-GUI membrane builder is organized in six steps. Step one is to upload and read the protein coordinates. Step two involves choosing the orientation of the protein. In step three the system size, water-layer thickness, lipid composition, ion type and concentration, as well as the system building method are being selected. In step four, based on the system size and other settings determined in the previous step, the individual pieces such as the lipid bilayer around the protein, additional water molecules and ions are built or provided. All the pieces generated from the former step are assembled together in step five. In the last step, the final input files for further equilibration MD simulations are generated [66].

For the membrane-anchored proteins GABARAP and GBPs studied in this work, the two methods mentioned in 3.1.3.1 and 3.1.3.2 are not suitable. For this, a third method introduced in 3.1.3.3, is applied.

3.1.3.3 Delete lipids overlapping with the protein

In addition to the methods mentioned above, another straightforward approach for the insertion of a protein into a membrane is to delete overlapping lipids. By removing lipids from a bilayer, a void is created for a protein to be placed into. Tools like genbox/solvate from the GROMACS package can be employed to achieve this goal. Researchers can also manually remove the lipids which collide with the protein. However, this method usually perturbs the lipid composition and the symmetry of the membrane. To avoid large disturbances, Shen et al. [67] described a gentler approach: delete a few lipids to create an initial hole first, then the hole cylindrically expanded by a repulsive potential. After that, a long equilibration MD simulation is required as the created hole does usually not match the shape of the protein. This method was therefore developed further by taking the protein shape into account when creating the hole. Approaches following this idea are, for example, implemented in make_hole from GROMACS and GRIFFIN [68]. The advantage of these two methodologies is their universality to both atomistic and coarse-grained models as well as independence from the used MD package. The disadvantage is that these approaches are not very user-friendly. In the meanwhile, as these methodologies require the removal of lipids, a more careful equilibration will be needed after generating the structure.

In our projects, GABARAP and GBPs are lipid-anchored membrane proteins, therefore, it was sufficient to remove only one lipid molecule, which was deleted in the upper membrane leaflet. The resulting hole was then occupied by the lipid anchor. Further short simulations were conducted to remove possible atom clashes and obtain equilibrated systems for further simulations.

3.2 Force fields for biomolecular simulations

To describe the potential energy of a system of particles (usually molecules and atoms), mathematical functions were generated. The form of the mathematical functions and the accompanying parameters sets, that are derived from both experimental work and high-level quantum mechanical calculations, form together a force field. The potential energy (U) of the system is a sum of the potential energy of bonded (U_{bonded}) and non-bonded ($U_{nonbonded}$) interactions of all the atoms (eq 3.1-3.3).

$$U = U_{bonded} + U_{nonbonded} \tag{3.1}$$

$$U_{bonded} = U_{bond} + U_{angle} + U_{dihedral} \tag{3.2}$$

$$U_{nonbonded} = U_{LJ} + U_{coulomb} \tag{3.3}$$

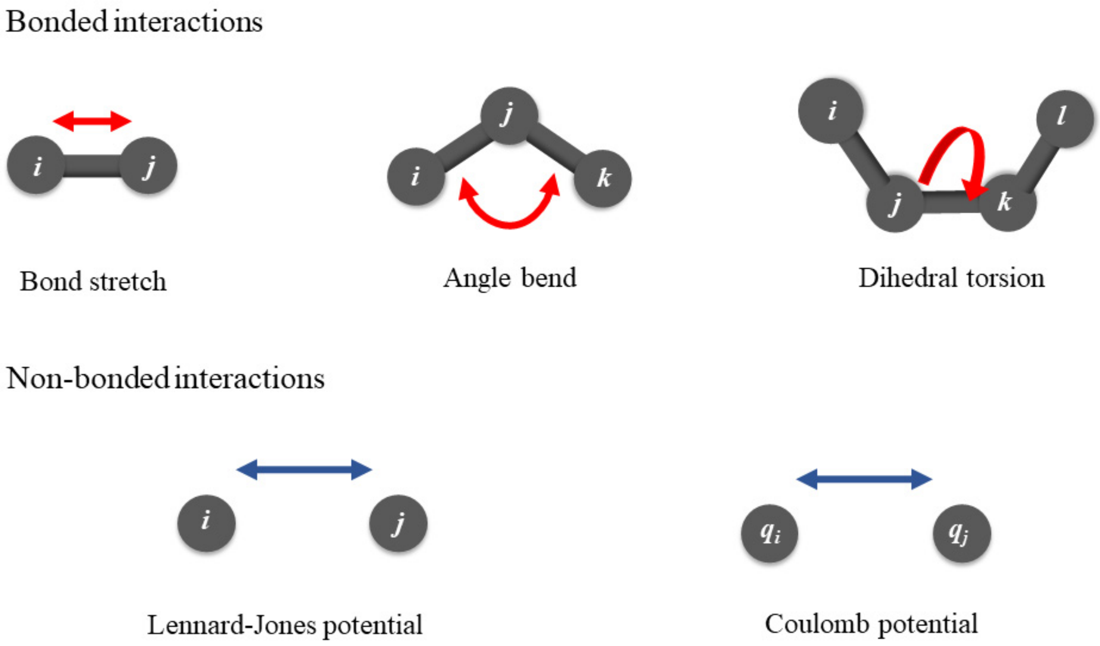


Figure 5. Bonded and non-bonded interactions used to describe interactions between atoms in a force field.

The bond stretch potential energy can be calculated using equation 3.4, where R is the distance between particles i and j (Figure 5), R_{bond} is the equilibrium bond length, K_{bond} is the stretching force constant.

$$U_{bond}(R) = \frac{1}{2}K_{bond}(R - R_{bond})^2 \qquad 3.4$$

U_{angle}, which represent the angle bend potential energy, describes the change in the angle between three consecutive atoms that are forming two bonds. It typically represented by a harmonic function below:

$$U_{angle}(\theta) = \frac{1}{2}K_{angle}[\cos(\theta) - \cos(\theta_0)]^2 \qquad 3.5$$

In equation 3.5, K_{angle} is the bending force constant, θ is the angle between atoms particles i, j and k while θ_0 the equilibrium value of the angle.

The common functional form for the dihedral angle potential $U_{dihedral}$ is shown below. It includes V_n as the barrier height and the integer n, which determines the periodicity of the potential, that is, the number of minima. Most common values are $n = 1$ or 2. For example, when the function is periodic by 360º, $n = 1$; for $n = 2$ the function is periodic by 180º.

$$U_{dihedral}(\varphi) = \frac{1}{2} V_n [1 + \cos{(n\varphi - \varphi_0)}] \tag{3.6}$$

A shifted Lennard-Jones (LJ) 12-6 potential energy function (eq. 3.7) and a Coulombic potential energy function (eq. 3.8) are used to describe the nonbonded interactions:

$$U_{LJ} = 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r} \right)^{12} - \left(\frac{\sigma_{ij}}{r} \right)^{6} \right] \tag{3.7}$$

$$U_{coulomb}(r) = \frac{q_i q_j}{4\pi\epsilon_0 \epsilon_r r} \tag{3.8}$$

Here, σ_{ij} denotes the closest possible distance between two particles i and j and ϵ_{ij} denotes the strength of their interaction. Different interaction levels correspond to different interaction strength values ϵ.

Careful selection of an appropriate force field is essential for every biomolecular simulation. When choosing the level of force field (for example, atomic force field or coarse-grained force field), the most important considerations are the time and length scales associated with the target project and the chemical precision that is required. Another factor that influences the choice of force field is its compatibility with available simulation software. Since not all molecules are parameterized in all force fields, one should choose a force field that includes the required molecular parameters or provides tools for parameterizing them.

The available lipid fields rarely cover all possible head groups and tail types. It is worth noting that certain head groups (such as phosphatidylcholine) and tails (such as palmitic acid or oleic acid) are usually parameterized first and are included in almost every lipid force field. However, for less ubiquitous lipid types, the parameters are not readily available, limiting the membrane composition that can be modeled. Moreover, the force fields of lipids, proteins and molecules that are outside the lipid and protein family (such as sugars, nucleotides or some small molecules) must be consistent and compatible with each other if these molecules are simulated together, which is the case when simulating membrane proteins. Therefore, the choice of a lipid force field limits the protein force field choice and vice versa. For a more detailed review and comparison of lipid and protein force fields, see resources [69][70][71]. It is worth noting that, as far as we know, when describing the combined membrane-protein system, there is no comprehensive comparison of the force field performance. In the following sections, some common force fields are introduced, which can be divided into different categories: atomistic force fields and coarse-grained force fields, according to the level of detail they provide. All-atom force fields provide parameters for every type of atoms in a system, including hydrogen atoms. Coarse-grained force fields are much cruder than all-atom force fields, providing an increased computational efficiency and thus allowing them to apply in long-time simulations.

3.2.1 Atomistic force fields

As the computing power is ever increasing and the accuracy provided by fully atomistic force fields has considerably improved, they are more and more used in membrane protein simulations. There are many all-atom force fields, such as from the AMBER, CHARMM and OPLS force field families, which are commonly used in biological simulations.

AMBER force fields encompass various versions, many of them are very popular, such as ff03 and ff99SB-ILDN. In particular ff99SB-ILDN [72] gaining wide spread acceptance. Version ff03 was used from 2003 onwards, and to some extent it still used today, yet ff99SB-ILDN is usually preferred. Based on ff99SB-ILDN, a refined force field, ff14SB was developed in 2015, which features improved amino-acid dependent properties, such as helical propensities and reproduces the differences in amino-acid specific Ramachandran maps using amino-acid specific CMAP corrections [73]. Nowadays, the ff19SB and ff14SB force fields are AMBER's primary protein model.

To overcome several limitations of the original AMBER lipid force field, multiple AMBER compatible lipid parameter sets, like Lipid11 [74], Lipid14 [75] and SLipids [76][77][78][79][80] were generated in consequence. Lipid11 was first reported in 2012, then the General AMBER Force Field (GAFF) lipid parameters [81] were subsequently combined with the development of Lipid11 [50] form the Lipid14 parameter set [75]. Lipid14 contains parameters for various lipid types and cholesterol [82]. The SLipids parameter set is available for saturated and unsaturated PC (phosphatidylcholine), PS (phosphatidylserine), PE (phosphatidylethanolamine), PG (phosphatidylglycerol), SM (sphingomyeline) lipids and cholesterol. They are supposed to work together with the protein force fields AMBER ff99SB, ff99SB-ILDN and ff03.

AMBER also supports the GLYCAM carbohydrate force field [83]. Except lipids, small molecules like drugs also often need to be included in protein membrane simulations. Several methods can be used to parameterize these molecules for the GAFF [84] force field [85][86][87][88]. In addition, a database containing many GAFF parameterized molecules (and also parameterized for the OPLS-AA and CHARMM force fields) stored in GROMACS format is available online [89].

Other force fields that are commonly used in membrane simulations are from the CHARMM force field family. In particular, the CHARMM36 lipid force field includes a great number of the most common cell lipid types [59][90], and the CHARMM36 protein force field [91] has also been employed extensively in membrane protein simulations. CHARMM also includes parameters for carbohydrates [92] and nucleic acids [93]. In order to apply general force fields to an arbitrary drug-like molecule, functionality for assignment of atom types, parameters, and charges is required. For example, the MATCH server could be used for automated atom typing [94] exists for CHARMM and the CHARMM General Force Field (CGenFF) [95]. It also provides ways to automate the force field parameter generation [96][97]. Like other force fields mentioned above, the parameterized molecules for the CGenFF in the GROMACS format are also available [89].

3.2.2. Water models of atomic force field

There are different water models that can be used in MD simulations, for example: TIP3P [98], TIP4P [99], TIP5P [100][101], simple point charge (SPC) model [102], the extended simple point charge (SPCE) [103], SPC/S (SPC like) [104], the three-site transferable intermolecular potential water molecules (TIP3PEW) [105], and two models of the four-site transferable intermolecular potential water molecules (TIP4PEW [106] and TIP4P-2005 [107]). The different water models have different properties and advantages.

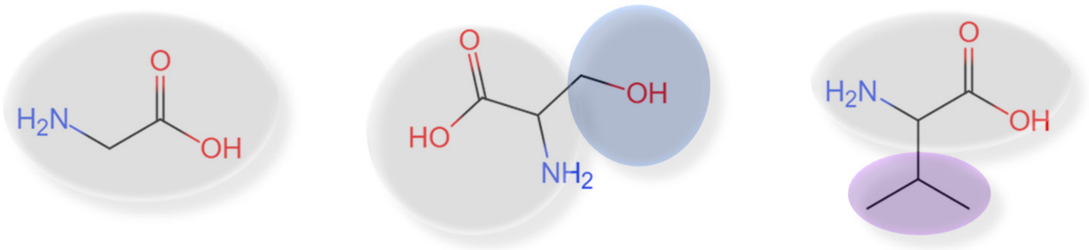
The research from Vega et al. showed that TIP5P and TIP4P water have similar qualities, while TIP3P water has a poorer performance. However, the advantage of TIP3P water model is that it is simpler than the other two models, and thus requires less computational cost, which is of importance for larger simulations. This is one of the reasons why it is more often used in simulations of biological systems [108]. The other reason is that many of the protein and lipid force fields were parameterized for usage with TIP3P or SPC, and the performance of these force fields may therefore depend on the usage of these water models. TIP4P and SPC water are better for reproducing experimental data compared to SPCE and TIP5P water models [109]. Studies of self-diffusion coefficients and radial distribution functions on water models indicate that the SPC/E water model produces the best bulk water dynamics and structure [110]. As each water model has its advantages and disadvantages, there is no agreement yet on which water model is the best one. However, when performing an all-atom simulation, one should be aware of the limitations of the chosen water model.

3.2.3 Coarse-grained force fields

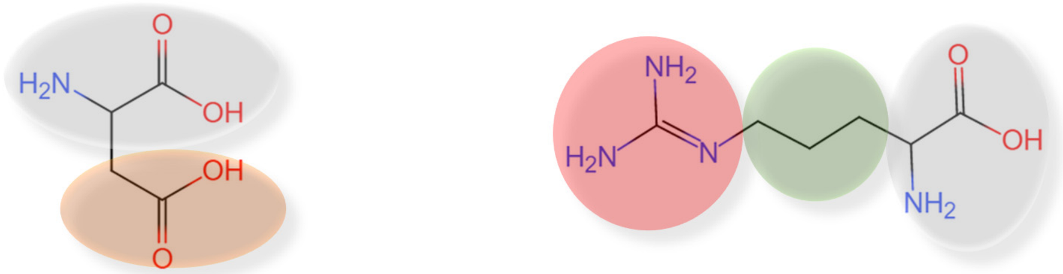
Simulations of large-size systems are computationally expensive. When increasing the size of a simulation system, the required computational cost for obtaining equilibrated results rises substantially [111]. To nonetheless being able to capture the events of biological processes' relevant time scales, which usually occur in the range of microseconds to seconds or even longer, coarse-grained (CG) models provide an alternative. It is a powerful tool for the study of large biological systems on time scales of up to milliseconds. They are, for example, applied to the study of biological complexes like ribosomes, cytoskeletal filaments and membrane protein systems [112][113].

There are a lots of coarse-grained models available, such as: MARTINI [114], UNRES (United Residue) [115], CABS (C-Alpha, Beta and Side Chain) [116], PRIMO (Protein Intermediate Model) [117][118], OPEP (Optimized Potential for Efficient Structure Prediction) [119] and PRIME model [120] etc. One of the most popular coarse-grained models for studying membrane proteins is MARTINI.

The MARTINI force field contains parameters for lipids [121], proteins [122][114], carbohydrates [123], nucleic acids [124] and polarizable water model [125]. The model is based on a four-to-one mapping, where four heavy atoms and their related hydrogen atoms are treated as an interaction site. This reduces the diverse set of molecules, as only four CG bead types are left: polar (P), nonpolar (N), apolar (C), and charged (Q) (Figure 6), which further simplified the parametrization and optimization strategy. Each bead type contains different subtypes, which are distinguished by their hydrogen-bonding capabilities: donor (d), acceptor (a), both (da), and by their degree of polarity (1 to 5, low polarity to high polarity).



Glycine Serine Valine



Aspartic Acid Arginine

Figure 6. Coarse-grained representation of the Martini model to amino acids, colored by bead type (purple is apolar, blue and green are intermediate, gray and orange are polar and red represents charged particles)

3.3 Simulation methods

From the end of the 1950s, people have begun to carry out computational experiments. At the beginning, the objects of these computational experiments were only small model systems, such as a collection of several hard spheres, which are also called ‘hard spheres model’ [126][127]. Alder and Wainwright [128], in the meanwhile, had foreseen that such computer simulations would become an important tool to reveal the relationship between macroscopic experimental phenomenon and its microscopic properties. Since the macroscopic and microscopic scales are completely different, to better reflect the corresponding macroscopic experimental phenomena by the microscopic simulation system, it is necessary to apply periodic boundary conditions on the simulation system to avoid edge effects that do not exist in practice. In principle, the complete theoretical study of any molecular system needs to solve the time-dependent Schrödinger equation.

However, more attention was often paid to the trajectories of the atomic nucleus in actual works. These trajectories can be obtained by applying the Born-Oppenheimer approximation and solving the equation of motion of classical mechanics [129].

3.3.1 Molecular dynamics simulation

The first MD simulation of a protein was performed based on classical molecular mechanics (MM) in the late 1970s [130]. The foundation of the computational classical or molecular mechanics (MM) method is Newton's second law:

$$F = ma \tag{3.9}$$

In equation 3.9, a represents the acceleration that an object undergoes, F is the force acting on that object which equals m (the mass of the object) times a. In an MD simulation, the behavior of a system is time-dependent. The Newton's second law equation motion can also be written as:

$$-\frac{\mathrm{d}U}{\mathrm{d}r_i} = m\frac{\mathrm{d}^2 r_i}{\mathrm{d}t^2} \tag{3.10}$$

U denotes the potential energy (as defined in eq 3.1–3.3) of the system under study with a certain conformation as represented by r_i, r_i which contains all atomic coordinates of the system. An MD simulation starts at time t_0 with an initial set of system specifications: the initial positions (i.e., the atomic coordinates which are usually obtained from a crystal structure or a homology model), the initial velocities (calculated using the Maxwell-Boltzmann distribution) and the potential energy. Then, by solving eq 3.9 for a short time interval Δt (usually 1-2 femtoseconds), new atomic positions and velocities are generated. This step is repeated until the envisaged simulation time has been reached, thereby generating the MD trajectory.

3.3.1.1 Numerical integration of Newton's equations

The numerical integration of the equations of motions in an MD can be accomplished using several methods. For example, after Alder and Wainwright performed the first molecular dynamics simulation experiments, the French physicist Verlet proposed a set of integration algorithms for Newtonian equations of motion, which are called Verlet's algorithm [131]. Besides that, there are other algorithms like the Leap-frog algorithm [132], Beeman algorithm [133] and Gear algorithm [134]. One of the most widely integration algorithms used in MD simulations and the default MD integrator in GROMACS is the Leap-frog algorithm.

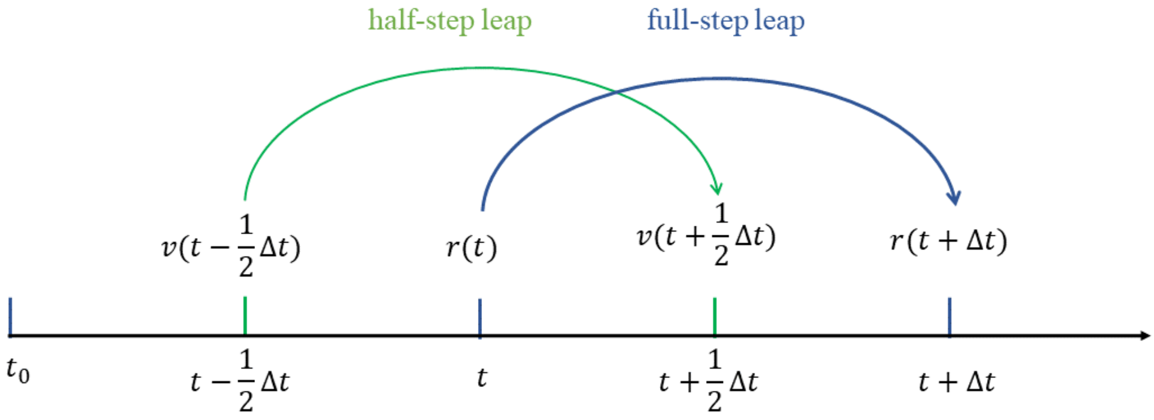


Figure 7. The Leap-frog integration method in which the new values of the particle velocities and positions are calculated at different time steps.

The Leap-frog algorithm uses positions r at time t and velocities v at time $t-\frac{1}{2}\Delta t$, it updates positions and velocities v using the forces $F(t)$ determined by the positions at time t using these relations:

$$v\left(t+\frac{1}{2}\Delta t\right)=v\left(t-\frac{1}{2}\Delta t\right)+\frac{\Delta t}{m}F(t) \tag{3.11}$$

$$r(t+\Delta t)=r(t)+\Delta t v\left(t+\frac{1}{2}\Delta t\right) \tag{3.12}$$

The velocities are updated at the half steps, while the positions and accelerations are updated at the full steps (Figure 7). This is where the name of the Leap-frog algorithm originates from.

3.3.1.2 Periodic boundary conditions

When running MD simulations, the goal is to mimic the structure of large systems in a realistic manner, which requires large-enough simulation boxes that place the particles far from the boundary of the simulation box in order to avoid unphysical artifacts at the boundary, where it would interface the vacuum. However, as a matter of fact, it is impossible to simulate an infinite system. Consequently, to minimize the effect of the boundaries, periodic boundary conditions (PBC) methods were introduced. PBC makes it possible to approximate an infinite system by using a unit cell and the unit cell is surrounded by translated copies of itself in all directions. If a molecule diffuses across the boundary of the simulation box it reappears on the opposite side. This unit cell is usually referred to as periodic box in an MD simulation. It should be mentioned that the usage of PBC requires the Coulomb interactions in equation 3.8 to be calculated with the particle mesh Ewald (PME) [135] method to avoid truncation errors which would appear if one simply cut the electrostatic interactions in an infinite system.

3.3.1.3 Temperature and pressure control

In MD simulations, quantities that we prefer to calculate are actually from a constant temperature ensemble, which can also be called a canonical ensemble. To simulate constant temperature, algorithms like the Berendsen [136] and Nosé-Hoover [137][138] thermostats can be used.

The Berendsen algorithm mimics weak coupling with first-order kinetics to an external heat bath with given temperature T_0. This algorithm allows the deviation of the system temperature from T_0 to be corrected according to the following equation:

$$\frac{\mathrm{d}T}{\mathrm{d}t} = \frac{T_0 - T}{\tau} \tag{3.13}$$

The temperature deviation decays exponentially with a time constant τ. This method is extremely efficient for relaxing a system to the target temperature.

When trying to relax a system to a target temperature, the Berendsen weak-coupling algorithm is very effective. After the system has reached equilibrium, the weak-coupling scheme is less useful. Instead, the Nosé-Hoover temperature coupling method is better suited here. In this method, the system Hamiltonian is extended by introducing a thermal reservoir and a friction term ξ in the equations of motion. The new equations of motion for the particles are:

$$\frac{\mathrm{d}^2 r_i}{\mathrm{d}t^2} = \frac{F_i}{m_i} - \frac{p_\xi}{Q}\frac{\mathrm{d}r_i}{\mathrm{d}t} \tag{3.14}$$

$$Q = \frac{\tau_T^2 T_0}{4\pi^2} \tag{3.15}$$

Q is a constant that determines the strength of the coupling, usually called the mass parameter of the reservoir, and p_ξ is the momentum of the reservoir. The equation of motion for the heat bath parameter ξ is:

$$\frac{dp_\xi}{dt} = (T - T_0) \tag{3.16}$$

Similar to temperature coupling, the system can also be coupled to a pressure bath. Berendsen, Parrinello-Rahman [139][140] and Martyna-Tuckerman-Tobias-Klein (MTTK) [141] pressure control methods are available in GROMACS. The Parrinello-Rahman and Berendsen barostats can be combined with multiple temperature coupling methods while MTTK can only be used together with Nosé-Hoover temperature control.

In the Berendsen barostat, the system is weakly coupled to an external bath that affects the pressure change. To this end, similar to the temperature coupling, an extra term is added to the equations of motion (equation 3.17) where τ_p is the time constant for the coupling.

$$\frac{dP}{\mathrm{d}t} = \frac{P_0 - P}{\tau_p} \tag{3.17}$$

When it comes to Parrinello-Rahman pressure coupling, it is similar in idea to the Nosé-Hoover temperature coupling method and follows the equation below:

$$\frac{\mathrm{d}^2 \mathrm{b}}{\mathrm{d}t^2} = V\mathrm{W}^{-1}\mathrm{b}'^{-1}(P - P_0) \tag{3.18}$$

Here, the box vectors are represented by the matrix b, V denotes the volume of the box, W determines the strength of the coupling, and P and P_0 represent the current and reference pressures, respectively.

3.3.1.4 Workflow of MD simulations

Nowadays, simulations have proven their value by predicting and optimizing the structure of peptide, proteins, and other biomolecules. They also helped to discover the functional mechanisms of proteins and other biomolecules and to uncover the molecular basis of diseases. The most widely used biomolecular simulation method is currently based on MD simulations, including all-atom and coarse-grained MD simulations as well as different enhancement techniques, such as replica exchange molecular dynamics simulations. Main method used in this research are all-atom MD simulations. Several tools or software can be used for running all-atom MD simulations, such as AMBER, GROMACS, NAMD, LAMMPS, YASARA etc. We used GROMACS in our research. A typical MD simulation experiment consists of several steps, which are summarized in Figure 8. One can see that, before a simulation can be started, an initial structure of the molecule(s) to be simulated is required. The structure of proteins can be determined experimentally, either by X-ray crystallography, NMR spectroscopy, or cryo-EM. These experimentally solved protein structures are collected in and distributed by the Protein Data Bank (PDB) [10].

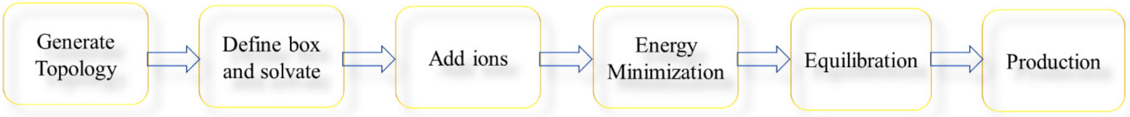


Figure 8. The workflow of an all-atom MD simulation.

3.3.2 Coarse grained molecular dynamics simulation

CG models are widely used for large-scale systems or long-time simulations, such as the study of the dynamics of multimers, protein-protein and protein-membrane interactions. As mentioned in section 3.2.3, there are various kinds of CG force fields, with the most commonly used one for membrane-protein system being the MARTINI force field. The workflow of a MARTINI CG simulation is similar to that of an all-atom MD simulation, as shown in Figure 9. The only difference is the topology setting part.

Figure 9. The workflow of a coarse-grained MD simulation using Martini.

3.3.3 Replica exchange molecular dynamics simulation

Biomolecules often exhibit many local free energy minima, some of the local free energy minima can be explored by conventional MD simulations, while others are separated by high free energy barriers and therefore not accessible on the time scale of an MD simulation [142][143]. To overcome this limitation, different enhanced sampling techniques were developed. One of them is the replica exchange molecular dynamics (REMD) simulation method which combines MD simulations with the Monte Carlo algorithm [144].

A detailed description of the REMD algorithm can be found in Okamoto and his colleagues' paper who first introduced this technique in biomolecular research [145]. The process of REMD simulations is shown in Figure 10. In a nutshell, multiple replicas of the same system are simulated in parallel at different temperatures (T-REMD) or at the same temperature but with different potential functions, called Hamiltonians (H-REMD). Regular attempts are made to swap between neighboring replicas with probabilities of accepting these exchange attempts defined by the Metropolis criterion. For T-REMD, where the simulation conditions differ in the temperature T, the exchange probability (p) is shown in equation 3.19. Here, E_i and E_j represent the potential energy of replica i and j, respectively, while Ti and T_j denote the corresponding temperatures, and k_B is the Boltzmann constant.

$$p = \min\left(1, e^{(E_i - E_j) * \left(\frac{1}{k_B T_i} - \frac{1}{k_B T_j}\right)}\right) \qquad 3.19$$

Compared to the T-REMD method, the H-REMD method is more efficient and requires fewer replicas. The temperature is kept the same in all replicas, only the Hamiltonian parameters differ for selected parts of the systems. To this end, the system is divided into separate regions: hot regions where the potential function is modified and cold regions with unmodified force field parameters. This approach allows a better sampling in the hot regions (for example: proteins), as the force field parameters are changed such that energy barriers become lower and can thus be better crossed, while the cold regions (for example: the solvent) are not affected. The Metropolis criterion in that case is:

$$p = \min\left(1, e^{\frac{E_i(r_i) - E_i(r_j) + E_j(r_j) - E_j(r_i)}{k_B T}}\right) \qquad 3.20$$

The scaling factor $\lambda \leqslant 1$ is used to modify the energy functions, with different λ values being applied in the different replicas. In the hot region, the atom charges are scaled by $\sqrt{\lambda}$, the Lennard-Jones parameters and proper dihedral potentials are scaled by λ , and the dihedral angles that cross the hot and cold regions are scaled by $\sqrt{\lambda}$. This translates to effective temperatures $\frac{T}{\lambda}$ in the hot region, while the temperature is unaffected in the cold region. The effective temperature at the interface of the hot and cold region is $\frac{T}{\sqrt{\lambda}}$.

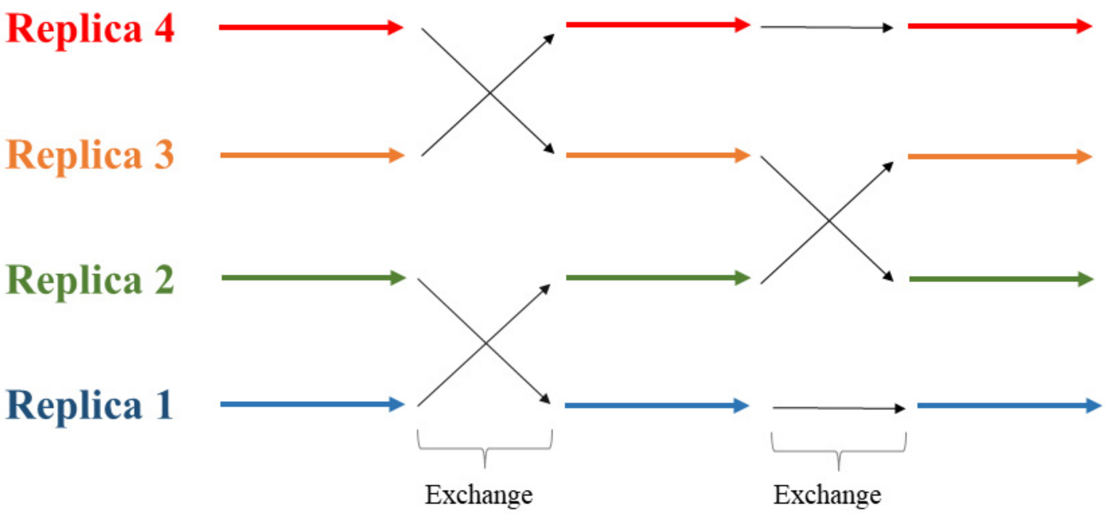


Figure 10. replica exchange md simulation process.

3.4 Analysis methods

Here, some of the analysis methods applied to the various MD simulations performed in this thesis work are provided. In section 4, where the results of the work are presented, more details on the system-specific analysis methods are given.

3.4.1 RMSD

A very common analysis method applied to MD simulations is to assess the degree of the structural flexibility via the root mean square deviation (RMSD). It measures the average displacement distance of the atoms of a selected time frame of the MD trajectory compared to a reference time frame, for which commonly the MD start structure is used. It is defined as following:

$$RMSD = \frac{1}{N}\sqrt{\sum_{i=1}^{N}\delta_i^2} \qquad 3.21$$

Here, N means the number of atoms of the selected conformation, i enumerates the N atoms, and δ_i is the distance between the positions of atom i in the structure under consideration and in the reference structure. Usually, the backbone atoms or C_α atoms are used to calculate the RMSD. The RMSD reflects the structural similarity of two conformations: the smaller, the more similar. It can also be used to measure the equilibration state of MD simulations.

3.4.2 RMSF

Another measure for the structural flexibility is the RMSF, which stands for root mean square fluctuation. It can be computed using each residue's C_α atom, backbone atoms, or heavy atoms. It represents the flexibility of an individual residue. In other words, the RMSF values help to find the most flexible and most stable residues during a simulation. It is defined as:

$$RMSF = \sqrt{\left\langle \left\| r_{i,t} - r_{ref} \right\|^2 \right\rangle} \qquad 3.22$$

Here, $r_{i,t}$ denotes the position vector of atom i at time t, while r_{ref} is the atom vector of a reference structure, and the angular brackets mean averaging over all frames of the selected trajectory. The reference structure usually is the average structure of a given trajectory, which is calculated in the first step of an RMSF analysis.

3.4.3 S^2 order parameter

The S^2 order parameters were calculated in this work to allow for a direct comparison between simulation and NMR spectroscopy results. These order parameters are a measure for the motions of N-H bonds in a protein and are thus related to the protein backbone flexibilities, which can be easily assessed in computationally feasible timescales. The change of the N-H bond vector orientations with time can be calculated from the MD trajectories. To obtain the internal protein motions, the overall protein rotations are removed first by aligning all trajectory frames to a reference structure (which usually is the starting structure of the MD simulation. The S^2 order parameter is then defined as

$$S^2 = \frac{1}{2}\left(3\sum_{i=1}^{3}\sum_{j=1}^{3}\langle \mu_i \mu_j \rangle^2 - 1\right) \qquad 3.23$$

where μ_i , μ_j denote the x, y, and z components of a given N−H bond vector scaled to unit length, while the angular brackets indicate averaging over time t. We used the software MOPS^2 to calculate the S^2 order parameters over MD trajectories, which was developed by Dr. Oliver Schillinger. The MOPS^2 is available at https://github.com/schilli/MOPS.

4 Results and discussion

The common aspect of GABARAP and some of the GBPs (for example, hGBP1 and mGBP2) is that they can conjugate to a lipid anchor and then be recruited by a lipid membrane. GABARAP mainly binds to the autophagosome membrane, promoting autophagosome formation and autophagosome-lysosome fusion, while GBPs bind to the parasite's vacuole membrane, which eventually kills the parasites by destructing the membrane that protects them. However, it is not clear how the proteins perform these functions. To shed light on that gap of knowledge, multiple molecular simulations, in cooperation with experimental procedures realized by our collaboration partners, were performed in this work. The overarching goals of these simulations were to provide details of the protein motions that may be of functional relevance, to understand how these motions affect the lipid membranes after binding of the proteins to the membranes and whether the protein dynamics is changed by the membranes, and what role protein dimerization plays in these processes.

In the first application, MD simulations with different force fields and water models were applied to study the dynamics of the GABARAP monomer. Special attention was put on the conformational dynamics of the α1-2 helices, as there was experimental evidence for an elevated flexibility of these helices. In order to rule out force field artefacts, different all-atom force fields were tested. Although the α1-2 helices showed indeed an increased flexibility compared to most other parts of the protein, they remained nicely folded and attached to the protein core. This finding did not alter when changing the force field, not even when using a force field that was shown to reproduce the conformational flexibility of intrinsically disordered proteins and also not when applying HREMD to GABARAP. To further test the flexibility of these two helices, we deliberately broke salt bridges by introducing mutations (E100A, D102A), as our MD simulations showed that α1-2 interact via stable salt bridges with the protein core. However, breaking these salt bridges confirmed the stability of both helices, as other, nearby charged residues replaced E100 or D102 in the salt bridges. The previously predicted dimeric models of GABARAP were also investigated in solution and bound to a membrane. However, these simulations revealed these dimer models to be unstable, questioning the validity of these models. The MD simulations of monomeric GABARAP anchored to a membrane allowed us to identify several residues which contribute the most to the membrane-binding process, in addition to the lipid anchor itself. This information was confirmed by comparing the S^2 order parameters calculated from our simulations to the experimental values obtained by our collaboration partners (Dr. Irina Apanasenko and Dr. Philipp Neudecker, IBI-7, Forschungszentrum Jülich).

The second application aimed at unraveling the conformational dynamics of mGBP2 in its monomeric form, predicting the dimer model and how dimerization affects the dynamics of mGBP2, and determining the effects of mGBP2 binding to the membrane on the protein dynamics and on the membrane properties. As a dynamin-like protein, mGBP2 plays various roles in the resistance against intracellular pathogens, which is controlled by GTP binding and hydrolysis. Here, we simulated mGBP2 with GTP being bound (holo state) and without it (apo state). As the mGBP2 structure was built based on the hGBP1 crystal structure using the homology modeling method, given their high sequence identity of 68%, the simulation of apo-mGBP2 allowed us to compare its structural dynamics to that of apo-hGBP1 which was analyzed in detail by Barz et al. in a previous study [146]. In fact, both proteins perform a large-scale hinge motion, which is only mildly reduced by the binding of GTP as the comparison of the apo and holo states of mGBP2 revealed. In fact, under all the conditions that were simulated, we observed similar hinge motions of mGBP2. However, the amplitude of the hinge motion became smaller in dimeric mGBP2 and following its membrane binding compared to monomeric mGBP2 in solution. It can be hypothesized that the hinge motions might be related to the destruction of the parasitophorous vacuole membrane if mGBP2 assemblies are bound to the membrane. This needs to be tested in future studies since with only one mGBP2 molecule bound to the membrane, no significant changes in the membrane properties were observed. Through decomposition on a per-residue basis of the mGBP2-membrane interaction energies, several residues that contribute the most to mGBP2 binding to the membrane were found. However, mutation experiments performed by our collaboration partners (Dr. Daniel Degrandi and Prof. Klaus Pfeffer, HHU Düsseldorf) did not find changes in the membrane binding of mGBP2 following mutating the suggested lysine residues to alanine. This observation confirms that the geranylgeranyl lipid anchor of mGBP2 is the most important group for the binding of this protein to lipid membranes.

The third study of this thesis work focused on mGBP7. As a member of the GBP family, unlike other GBPs, mGBP7 contains an unconventional and elongated C-terminus (Figure 4). Since mGBP7 is essential in the defense against *Toxoplasma gondii* infections, the aim was to biochemically characterize mGBP7 by the combined application of experimental and simulation methods. The experiments were accomplished by our collaboration partners (Dr. Larissa Legewie, Dr. Nora Steffens, Dr. Daniel Degrandi and Prof. Klaus Pfeffer, HHU Düsseldorf), while we performed the simulations of mGBP7. We started with a bioinformatics analysis that unraveled that mGBP7 possesses an elongated C-terminal (CT) tail, as opposed to the CaaX motif-containing hGBP1 and mGBP2, which has transmembrane characteristics. Structure predictions for mGBP7 including the CT tail were made using homology modeling, the stability of the structure tested in MD simulations, and the membrane-binding capabilities of the CT tail assessed with simulations including a lipid bilayer. Based on these simulation predictions, experiments were performed to test the importance of the CT tail for membrane binding. And indeed, confocal microscopy analyses revealed that the CT tail is required for recruitment of mGBP7 by the membrane of the parasitophorous vacuole of *Toxoplasma gondii*.

4.1 Manuscript I

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

Xue Wang, Wibke Schumann, and Birgit Strodel

X.W. contributed to this paper by performing all MD simulations and doing most of the simulation analysis. X.W. wrote the manuscript with inputs from the coauthors.

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

Xue Wang[1,2], Wibke Schumann[1,2], and Birgit Strodel[1,2]*

[1]Institute of Theoretical and Computational Chemistry, Heinrich Heine University Düsseldorf, 40225 Düsseldorf, Germany

[2]Institute of Biological Information Processing: Structural Biochemistry, Forschungszentrum Jülich, Jülich, Germany

*** Corresponding Author: Birgit Strodel**

b.strodel@fz-juelich.de

Abstract

Conformational dynamics are very important to understanding the function of a protein. GABARAP was thought to have high flexibilities on its a1 and a2 helices, these high flexibilities were also assumed to be related to its oligomerization mode. To figure this out, in this paper, we compare the dynamics of the GABARAP monomer, between our simulations with different force fields and water models. Hydrogen bonds were also broken by mutating the vital charged residues to Alanine. After that, multiple analyses were conducted, among them RMSF, hydrogen bonds and S2 order parameter. Based on previous study, a GABARAP dimer model was built and simulated both in solution and anchored to membrane. The dimer model was eventually proved not stable, whether in solution or anchored to membrane. However, the GABARAP monomer and dimer, both anchored to membrane systems, showed a similar interaction mode, residues Lys 2, Lys 38, Glu 7, Asn 82 and Glu 101 contributing significantly to the membrane binding.

1. Introduction

Autophagy is the non-damaging response of cells to continuous internal and external stimulation. It plays wide and very important roles in maintaining the balance of cell structure, metabolism and function [1] and is implicated in cancer [2], HIV/AIDS [3] and neuroplasticity [4]. In yeast, a single Atg8 protein triggers the formation of autophagosomes, which are double-membrane vesicles that subsequently fuse with lysosomes, thus degrading misfolded proteins and recycling cellular resources [5][6]. Atg8 proteins are conserved, ubiquitin-like proteins, that are divided into two subfamilies: light chain 3 (LC3) and gamma-aminobutyric acid receptor-associated protein (GABARAP). Unlike in yeast, six Atg8 orthologs participate in autophagic flux in human, named LC3A, LC3B, LC3C, GABARAP, GABARAPL1 and GABARAPL2 [7]. All of them are essential for autophagy by conjugating to autophagosomes and they can recruit two broad classes of molecules: autophagy receptors and autophagy adaptors. ATG8 also carry out functions such as autophagosome formation, autophagosome transport, and autophagosome fusion with the lysosome, by interacting with these receptors and adaptors (such as ULK1/2 [8], FYCO1 [9] and PLEKHM1 [10] et. al) on the autophagosomal membrane [11].

Studies show that during autophagosome formation, the C-terminal glycine of LC3/GABARAP is exposed through the action of the cysteine protease ATG4, by cleaving the C-terminal fragment off, generating the active form-I (LC3-I and GABARAP-I) variants. Afterwards, in the presence of ATG3, ATG5 and ATG7, LC3/GABARAP-I is conjugated to the membrane lipid phosphatidylethanolamine (PE) by a ubiquitin-like conjugation system, which also called form-II [12][13]. In 2000 and 2001, Paz et. al [14] and Knight et. al [15] first resolved the structure of GATE-16 and GABARAP respectively (PDBID: 1EO6 and 1GNU). The whole GABARAP structure includes a four-stranded β-sheet, with two α-helices adjacent to each face (Figure1). It contains 117 residues in total, which are divided into two domains: a small N-terminal subdomain (residues 1-26) and a C-terminal domain (CTD, residues 27-117). The N-terminal domain contains an α helix (H2), and a 3-10 helix (H1) that binds to tubulin. The four-stranded β-sheet in C-terminal domain is classified as a β-Grasp fold which is shared by the ubiquitin-like receptor binding region (residues 36-68) and membrane binding region (residues 68-116). During autophagosomal process, a LC3 interaction region (LIR) is common in all adaptors and receptors. This LIR motif binds to the GABAA-receptor binding region, this region is also called LIR docking site (LDS), which consists of two larger pockets known as hydrophobic pocket 1 (HP1) between α1 and β2 (V31 and L50) and hydrophobic pocket 2 (HP2) (Figure 1) between β2 and α3 (Y49, V51, F60 and L63) [16].

Above all, lots of experimental research has already been done on the Atg8 family. In 2015, Thukral et. al [17] used the MARTINI coarse-grain (CG) model to observe the spontaneous insertion events of lipid anchors, and the membrane-inserted state of LC3-PE. However, the oligomerization of GABARAP is still unknown. According to Coyle's [18] study, GABARAP's C-terminal domain adopts two conformations due to different crystallization conditions. GABARAP is monomeric in solution under a concentration of 100 μM, with a lower salt concentration, while the oligomeric conformation of GABARAP was crystallized under high salt conditions. The conformations of the monomeric and oligomeric form of GABARAP are quite similar, but they have a big difference on the first 10 residues. In the monomeric conformation, the first 10 residues point down toward the C-terminal domain surface, which is referred to as the closed conformation. Meanwhile in the oligomeric form, the first 10 residues are projected away from C-terminal domain surface, interacting with HP2 of an adjacent GABARAP molecule. The study showed a high flexibility of GABARAP in the N-terminal domain, which presumably is induced and stabilized via interactions with membranes or other proteins in *vivo*. In the same year, Jesper et. al [19] investigated the binding domain in GABARAP which mediating dimer formation by biochemical approaches. They identified GABARAP as a self-associating homodimer and the dimerization is not simultaneous with receptor binding, but rather an alternative. The interacting and dimerization domains are both located in the region from residues 41 to 51.

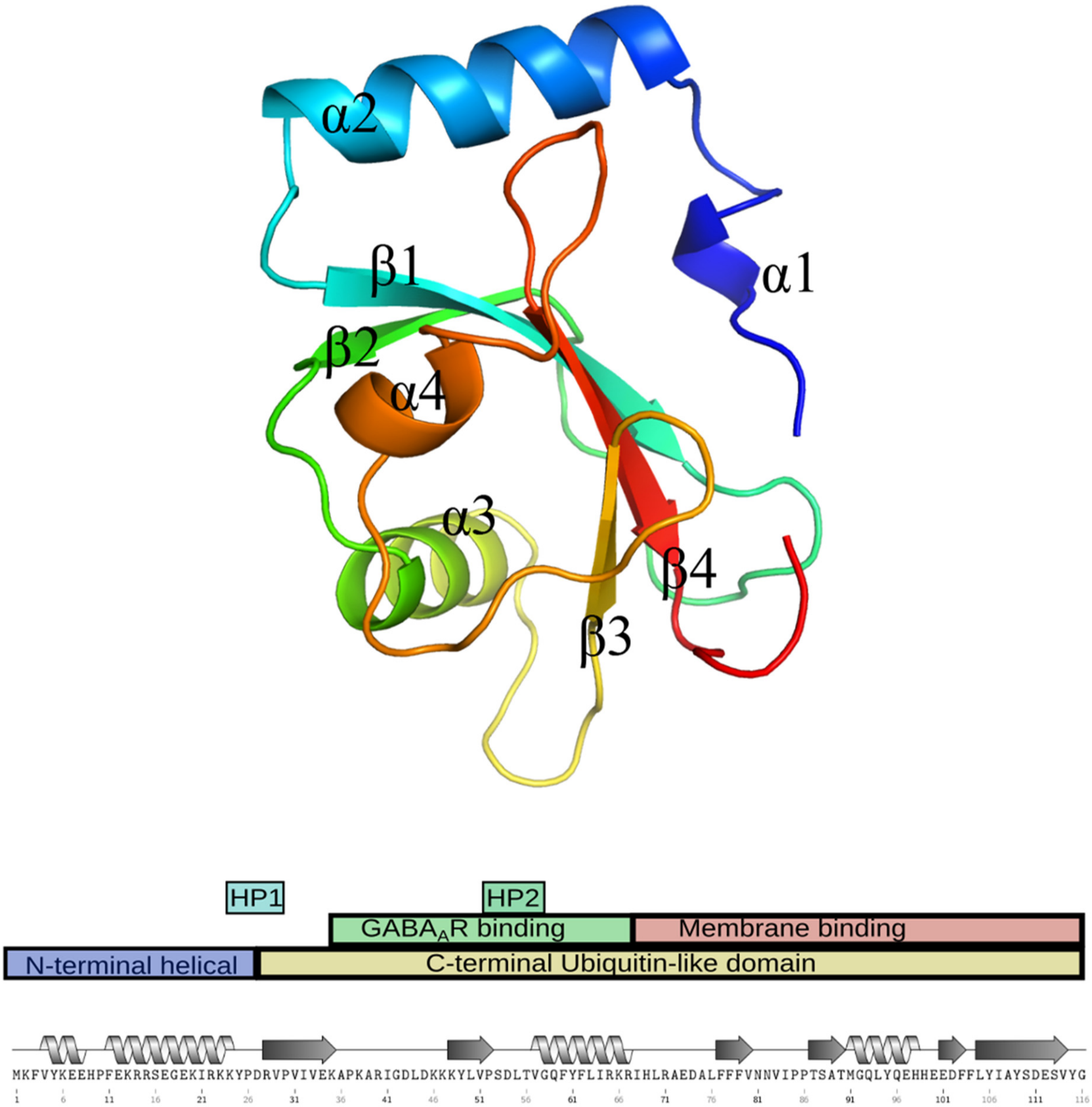


Figure 1. GABARAP shown in cartoon model, domains shown in different colors.

Even though a lot of studies have already been done on the dynamics and dimerization of GABARAPs, the oligomeric mode of GABARAP, based solely on the structure is still unclear. To end this, we built the dimeric model of GABARAP based on the information from Coyle [18], and the dynamics and stabilities of GABARAP monomers, dimers, and dimer binding to membrane were investigated by molecular dynamics simulations. Apart from that, the dynamics of GABARAP monomer were also studied as the flexibility of α1 helix is disputable.

2. Methods

As a valuable tool, MD simulations already have been widely used to study the structure and dynamics of proteins for decades [20]. Until today, the development and improvement of empirical force fields is still continuing, and a lot of popular force fields have been developed: CHARMM [21], AMBER [22], GROMOS [23], OPLS [24], and COMPASS [25]. The relative stabilities of different types of secondary structure and the accuracy of description of the structure and dynamics of globular proteins were improved in the CHARMM and Amber force fields. As studies showed that flexible folded proteins have different conformational space under different force fields, the water model is also crucial in molecular dynamics simulations [26][27][28], multiple simulations of GABARAP monomers were run with different force fields and water models in order to get a more precise understanding of GABARAP monomer dynamics. Specifically, Amberff03ws [29] with TIP4P-2005 [30], Amber99sb*-ildnp [31] with TIP3P, Charmm22 [32] with TIP3P and Charmm22 with TIP4P-D [33] water models were used (Table 1).

Table 1. Simulations and it's setting conditions.

Systems	Force field	Time (ns)	Water model	Temp (K)
GABARAP-WT	Amber ff03ws	1000	TIP4P-2005	298
GABARAP-WT	Charmm 22*	1000	charmm-modified TIP3P	298
GABARAP-WT	Charmm 22*	1000	TIP4P-D	298
GABARAP-E100A	Amber99sb*-ildnp	1000	TIP3P	298
GABARAP-D102A	Amber99sb*-ildnp	1000	TIP3P	298
GABARAP-E100A-D102A	Amber99sb*-ildnp	1000	TIP3P	298
GABARAP-G116C	Amber99sb*-ildnp	1000	TIP3P	298
GABARAP-monomer-mem	Amber99sb*-ildnp	1000	TIP3P	298
GABARAP-cis-dimer	Amber99sb*-ildnp	1000	TIP3P	313
GABARAP-trans-dimer	Amber99sb*-ildnp	1000	TIP3P	313
GABARAP-cis-dimer-mem	Amber99sb*-ildnp	1000	TIP3P	313

2.1 Molecular dynamics simulations

The MD simulations were performed using the program GROMACS [34]. We use the crystal structure of human GABARAP (PDB ID: 1GNU) [15] as the starting structure. PyMOL [35] was used for GABARAP mutation and visualisation. The settings and conditions of each simulation are showed in Table 1.

In our simulations, the periodic boundary conditions were used and long-range electrostatic interactions were treated with the Particle Mesh Ewald (PME) [36] summation using grid spacing of 0.16 nm. The cut-off distance was set to 1.0 nm and the van der Waals cut-off to 1.2 nm. The bond lengths were constrained by LINCS algorithm [37] and a time step of 2 fs was used. Coordinates were saved every 20 ps, each MD trajectory was run for 1 μs long.

For GABARAP monomeric simulations, the protein was placed in a dodecahedral water box, large enough to contain protein and at least 1.0 nm of solvent on all sides. The starting structures were subjected to energy minimization using the steepest descent method. The simulations were subjected to Nosé-Hoover T-coupling bath to maintain the exact temperature. The structures were then subjected to Parrinello-Rahman barostat for pressure coupling at 1 bar, before the 1 μs production runs were started [38].

For GABARAP anchored to membrane systems, the DMPC (1,2-dimyristoyl-sn-glycero-3-phosphocholine) membranes for monomer and dimer were both generated using Charmm-GUI [39]. The PE anchor was conjugated to Gly 116, while Leu 117 was deleted. To get the starting structure of each system, the protein was initially placed on the membrane with one lipid in bilayer replaced by the PE anchor, and then minimization of the systems in vacuum to remove clashes was performed. The PE anchored GABARAP simulation box contained one GABARAP protein and a membrane consisting of 153 (one lipid replaced by PE anchor) lipid molecules, with a box size 7 × 7 × 10 nm. Considering the size of dimer, the box size of the GABARAP dimer system was setting to 14 × 14 × 15 nm, with a membrane containing 611 lipids. TIP3P water model was used for simulations. To produce a neutral system, sodium and chloride were added by replacing water molecules all the system. Amber99SB*-ILDNP [31] force field and Slipids [40] force field were used for protein and lipids respectively. The PE parameters were derived from already existing DMPE (1,2-dimyristoyl-sn-glycero-3-phosphoethanolamine) lipid parameter.

Simulations were performed at 310 K, the methods of Nosé-Hoover and Berendsen [41] were used to maintain temperature in NVT and NPT ensembles, respectively.

2.2 Analysis protocols and methods

The S^2 order parameters from the MD trajectories were calculated using the Python software MOPS^2 (Molecular Order Parameters S^2), which was developed by Oliver Schillinger (https://github.com/schilli/MOPS). In our research, the last 600 ns of the MD trajectories were divided into 60 sub-trajectories of 10 ns length each. Then the S^2 order parameters were computed for each sub-trajectory and afterward averaged over all sub-trajectories [42].

The gmx_energy program in GROMACS packages was used to compute the Coulombic and Lennard-Jones interaction energies between each GABARAP residue and the DMPC membrane.

3. Results and Discussion

3.1 Dynamics of the monomer

3.1.1 Dynamics of α1 and α2 helices

While inspecting the trajectories and comparing the structures of frames during the simulations, we found that the N-terminal of GABARAP adopted various degrees of motions. To explore this, a cursory measurement was used to describe these motions, taking a2, b2 and b3 in C-terminal as two rigid parts, then three carbon atoms were chosen from residues G18, V33 and I107, respectively. The angle between carbon atoms of G18, V33 and I107 of the starting structure will be set as starting angle, the distance between carbon atoms of G18, V33 will be set as starting distance. When the simulation angle is greater or less than the starting angle, that means the a2 helix is swing right or left compared to its original position. The distance changes also reflect the a2 helix is “closer” or “farther” to the “rigid” C-terminal (Figure 2A). We can see that the α2 helix are more mobile before the 600 ns, but it eventually becomes stable at the end (Figure 2B, C, D).

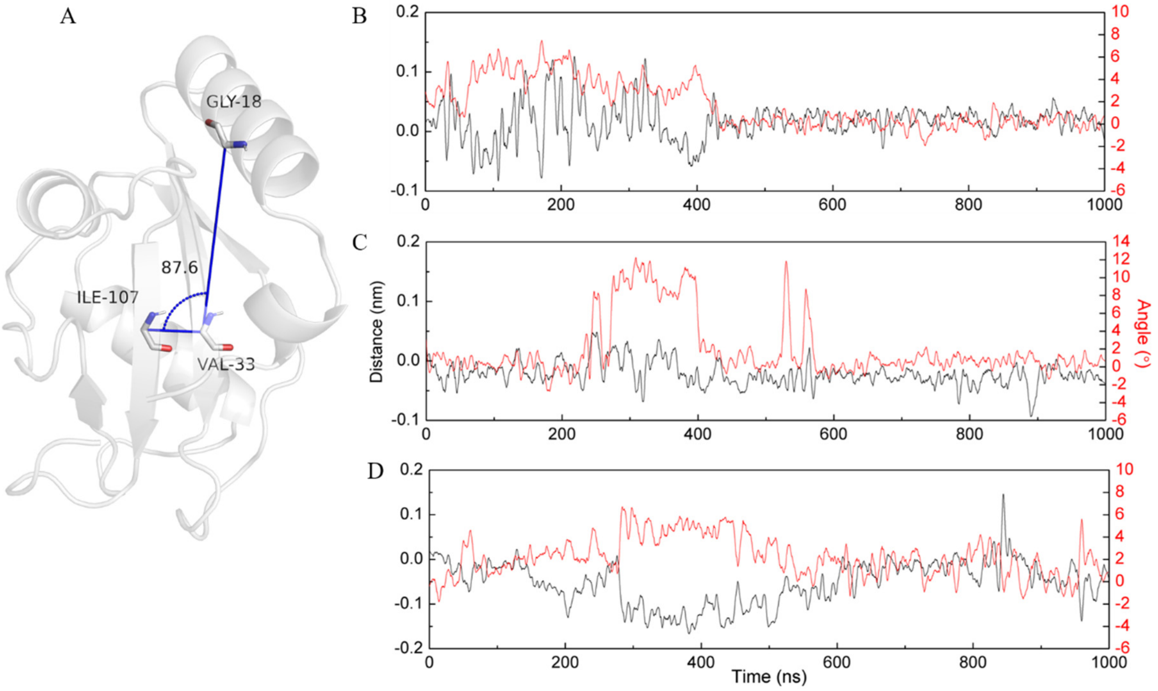


Figure 2. The angle between G18, V33 and I107 was shown in figure A. Angles and distances were red and black lines respectively in figure B (Amber03), C (C22), D (C22-TIP4P-D).

3.1.2 Root Mean Square Fluctuation results

Root Mean Square Fluctuation (RMSF) analysis reveals that, independent of force field, GABARAP is stable on the microsecond timescale. While Charmm22 with TIP4P-D (C22-TIP4P-D) increases the flexibility of the C-terminus, Amber03ws (Amber03 in Figure 3) has a positive influence on flexibility of the loop between residues 80-86. All force fields have in common that loop regions are flexible and β-strands especially stable, with Amber03 system showing more flexibility around β2 (Figure 3). The previously observed conformational flexibility of IR40/41 [43] can be confirmed, especially for the Amber force fields. While I41T is a mutation that causes endometrial cancer (COSMIC: https://cancer.sanger.ac.uk/cosmic), the exact change in GABARAP functionality is not known.

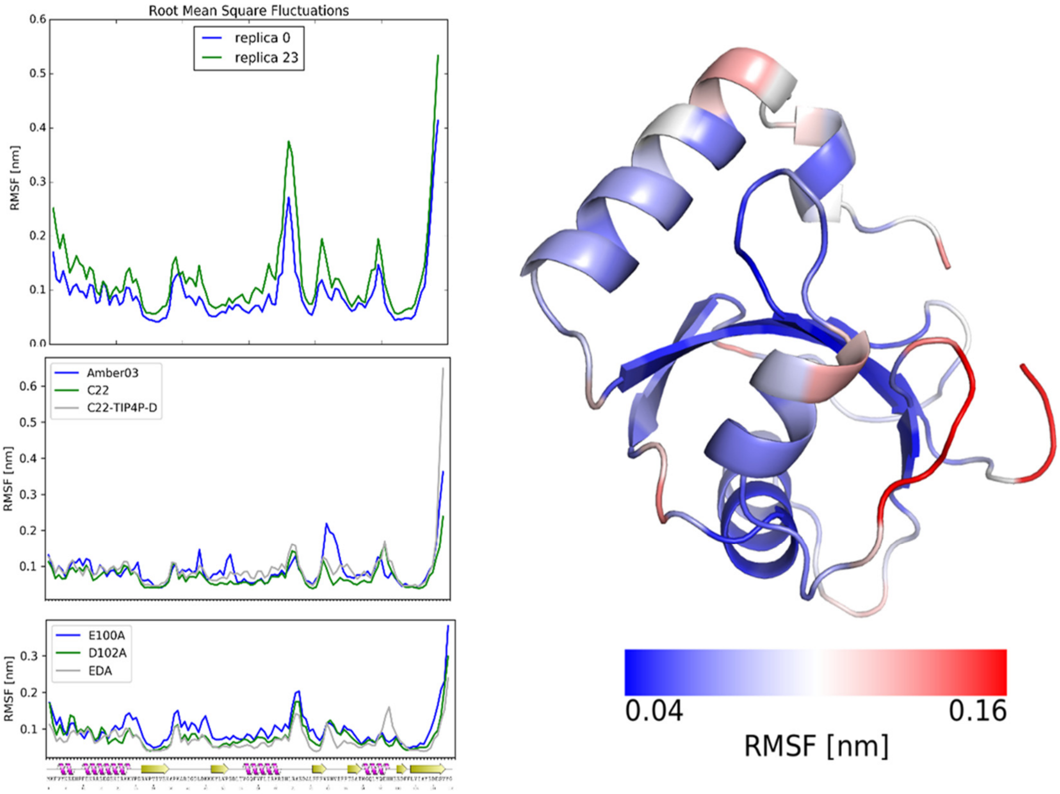


Figure 3. RMSF results of GABARAP monomer in solution, during HREMD (top), compared across different force fields (middle) and mutants (bottom). The RMSF of the CHARMM22 simulation is projected onto the structure (right).

3.1.3 Highly redundant salt bridges stabilize helices 1 and 2

The N-terminal helices α1 and α2, in the case of GABARAP, are acidic and basic, respectively, and capable of binding to the cytoskeleton. Therefore, these helices are worth further study. Under normal conditions we observe alpha1 tethered to the CTD by several salt bridges (K6-E100, R14-D102 most prominently), which prompted an *in silico* mutagenesis study. The wild type (WT), E100A, D102A and E100/D100A mutants were all simulated in Amber99sb*-ILDNP force field. These mutations had little effect on the RMSF and overall structure of the protein. The salt bridges, however, change in their constellation, rescuing the structure from a potentially nonfunctional conformation (Figure 4). The E100A mutation has little effect compared to the WT, but the D102A mutation has to be compensated for by a stronger R14-E100 salt bridge. In the double mutant, both K6 and R14 interact with E101 instead, but their contact frequency is lower, perhaps owed to geometrical restraints. These mutations are especially interesting in light of R15S, E100D, E101D being carcinogenic mutations with no known mechanism of action (https://cancer.sanger.ac.uk/cosmic).

These results are additionally supported by a REMD in which the structure only unfolds at 450 K, proving the extreme stability of these salt bridges and allowing for the necessary temperature raise while studying the membrane system. Furthermore, when separated from the CTD and simulated in isolation, helices 1 and 2 lose their secondary structure (results not shown).

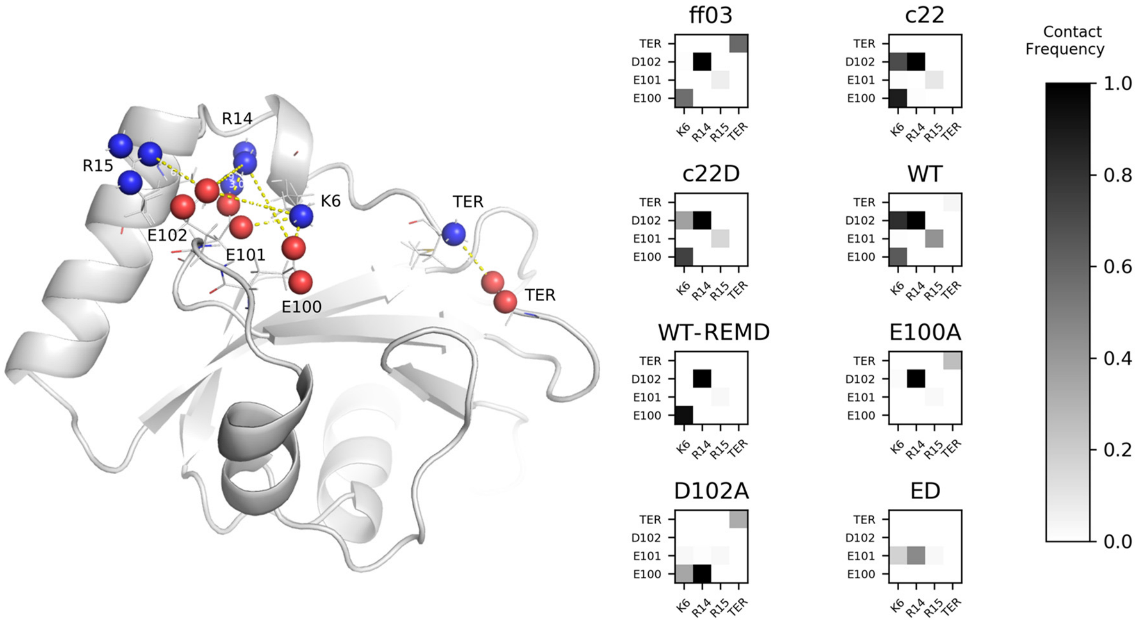


Figure 4. Salt bridges in all the monomer systems. Hydrogen bonds in all the monomer systems. Left: visualisation of residues in question on the structure. Right: Contact frequency of all pairs, compared across simulations. ff03: Amberff03ws with TIP4P-2005; WT: wild type of GABARAP with TIP3P; Charmm22 [32] with TIP3P and Charmm22 with TIP4P-D.

3.2 Dynamics of the dimer in solution

Coyle [18] first introduced the possible interaction form of GABARAP dimer, in the same year, Nymann et al. [19] identified that GABARAP interacting and dimerization domains are both situated in the region from residues 41 to 51. Furthermore, they suggest that GABARAP dimerization is not simultaneous with receptor binding but rather an alternative. Based on that, Dr. Oliver Weiergräber built the cis and trans dimer models of GABARAP, which is shown in Figure 5. The N-terminal residue M1 inserts into HP1 while V4 inserts into HP2. From Figure 5, it is apparent that these contacts alone cannot stabilize this dimeric binding mode, which means the dimer model we predicted cannot be stable in solution. Therefore, analysis will be shifted to its membrane-bound states.

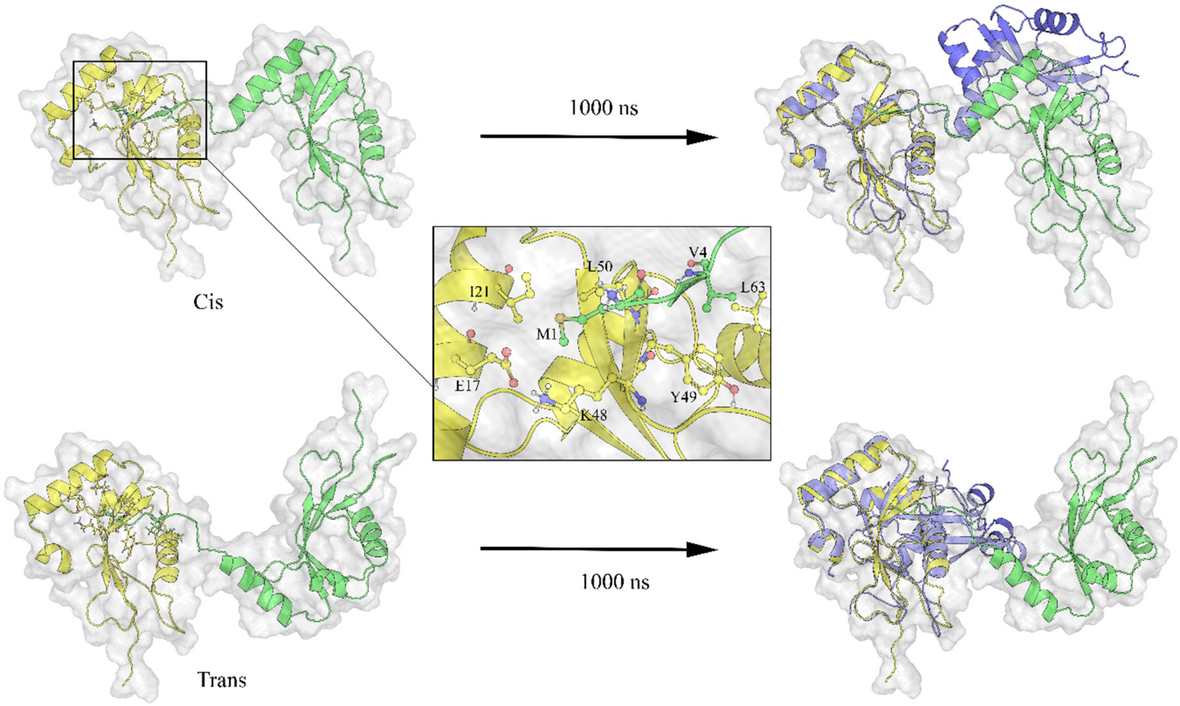


Figure 5. The starting structures and after 1 us simulation structures (colored marine) of GABARAP dimer in cis and trans states.

3.3 Dynamics of membrane-bound GABARAP monomer

When considering the S^2 backbone order parameter, the N- and C-terminus are most dynamic. By comparing monomer dynamics in solution and membrane-bound, we found GABARAP from different systems display variable flexibilities in different regions.

Significantly, residue I41, which located in the loop close to the C-terminal region, whose S^2 value is in the same range as the order parameters of the N and C-terminal residues (0.50 in WT system); this high flexibility could also be seen in RMSF results. The S^2 order parameters of the C-terminus in the GABARAP-G116C system are comparable lower to other systems (WT, GABARAP-MEM) mean flexibilities, these flexibilities might relate to the process of GABARAP changing from form-I to form-II. In addition, the regions from residue 66-77 and loop 84-89 should also not be overlooked, as they are located “upstream” and “downstream” of residues 81-83, which directly interacts with the membrane, and which might be affected by the flexibility of the two loops to some extent.

The Coulombic (Coul) interaction energy and Lennard-Jones (LJ) interaction energy between GABARAP monomer and the lipid bilayer were analyzed. From Figure 4B, the most outstanding energetically contributing (Coul + LJ) residues are Lys 2 (-82.27 kJ/mol) and Lys 38 (-88.94 kJ/mol). Apart from these two residues, other residues like Glu 7, Asn 82, Glu 101 also play important roles in GABARAP binding to the membrane, with a contribution energy of -49.07 kJ/mol, -62.68 and -59.88 kJ/mol, respectively. In summary, α1 and α2 helices, residues 81-82, 97-101 and C terminal loop are the major regions that contribute most the GABARAP and membrane interactions.

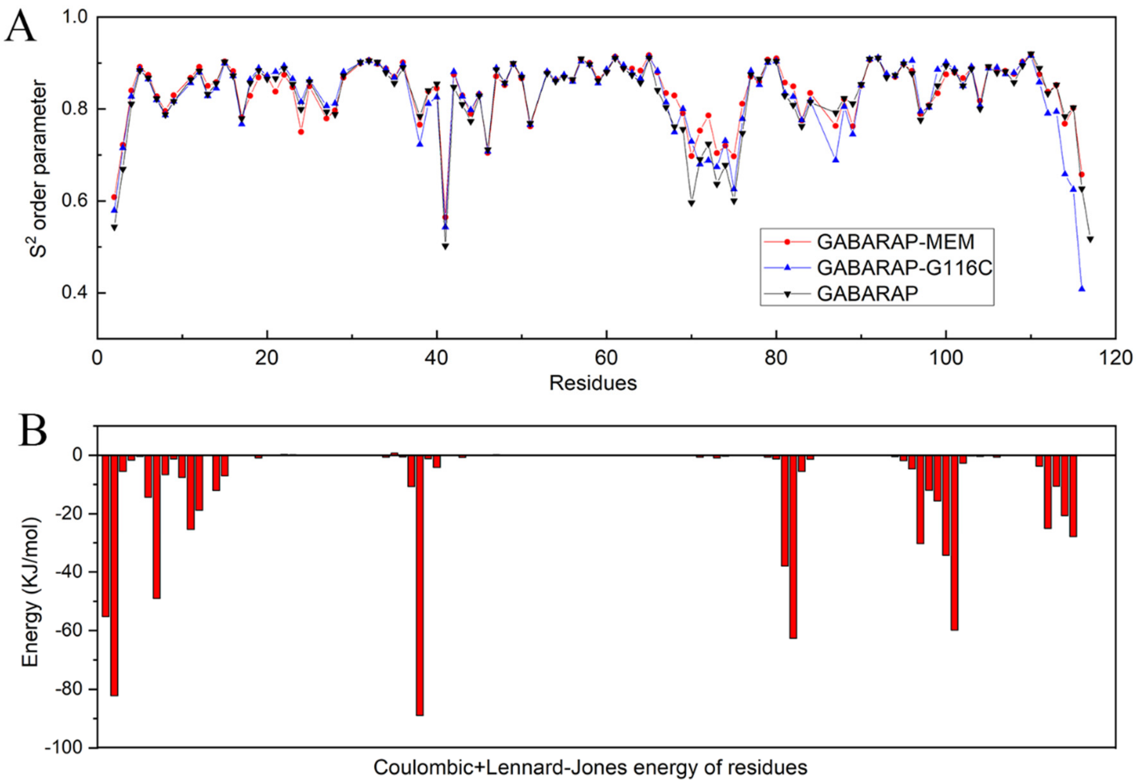


Figure 6. A. S2 order parameters for the N–H bond vectors. B. The sum of Coulombic and Lennard-Jones energies (short as energy) between GABARAP monomer and the bilayer of each residue.

3.4 Dynamics of membrane-bound GABARAP dimer

As the dimer model of GABARAP in solution is not stable, we assume that the interaction of GABARAP dimer with membrane will help for its stabilization. However, after 1us simulation, the dimer did not appear more stable than in solution, and the residue M1 was already extricating from the hydrophobic pocket 1 (Figure 7B). Although binding to membrane cannot stabilize the dimerization mode of GABARAP, from the simulation, we still can see a similar interaction pattern between GABARAP-monomer-MEM and GABARAP-dimer-MEM that Lys 2, Glu 7, Lys 38, Asn 82 and Glu 101 contribute significantly to GABARAP binding to membrane (Figure 7A). For the rest, Glu 97, Glu 100 and Tyr 115 in dimer anchored membrane system contribute more than that these residues in monomer system.

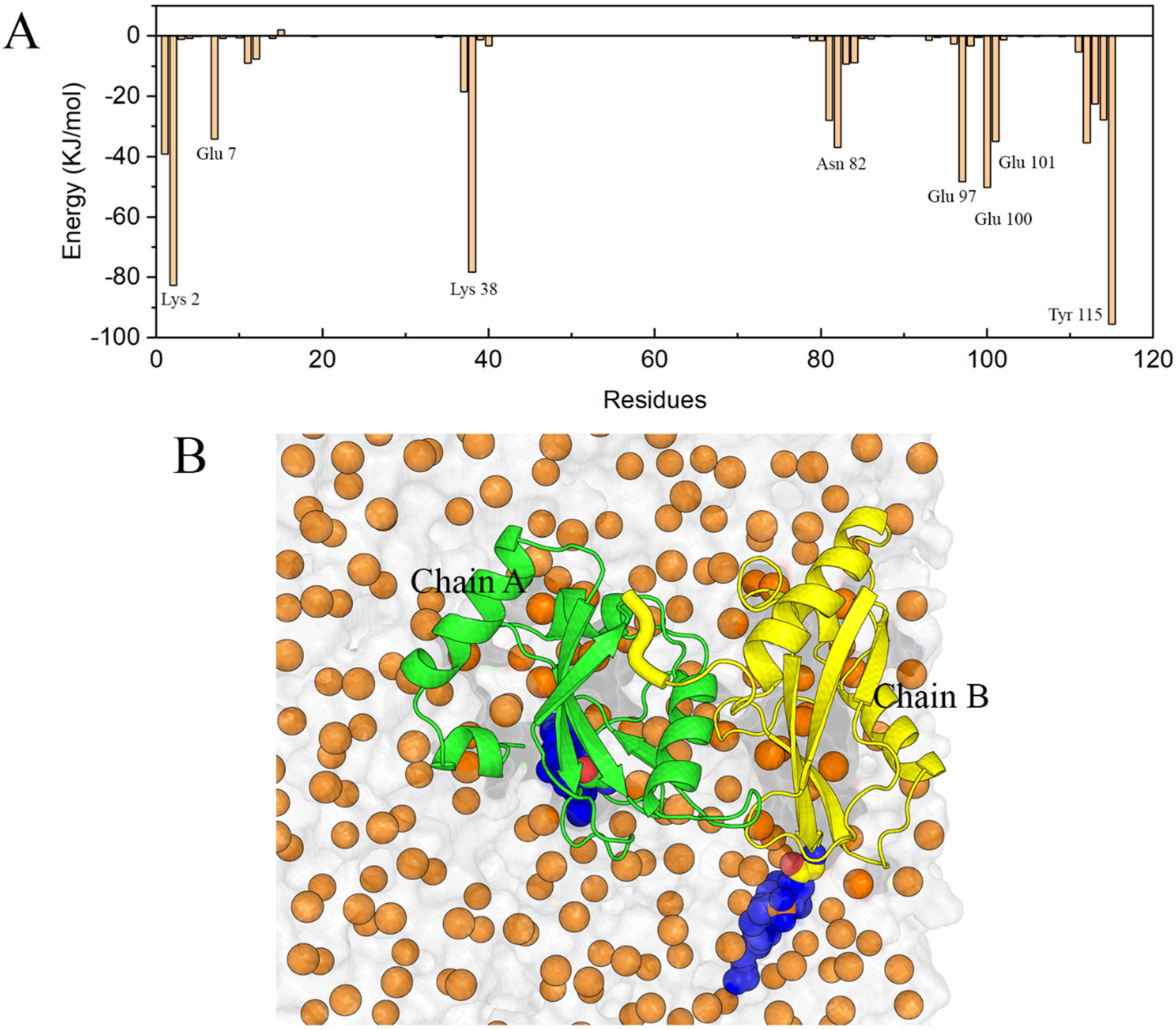


Figure 7. A. The sum of Coulombic and Lennard-Jones energy of each residue. Only results of the chain A of GABARAP dimer were showed here. B. GABARAP dimer bound to membrane after 1us simulation. Protein showed as cartoon and chain A and Chain B colored by green and yellow respectively.

4. Conclusion

The GABARAP monomer is stable in solution in our simulations, using different force fields and water models do not affect it very much. Mutation studies showed that, even if hydrogen bonding residues like E100 and D102 were mutated singly or simultaneously, other residues will compensate the lost interactions, which verified the stability of GABARAP from another side. The most likely dimerization mode predicted from previous study is not stable, whether in solution or anchored to membrane, that means other dimerization modes must be explored in the future. Residues Lys 2, Glu 7, Lys 38, Asn 82 and Glu 101 contribute most in GABARAP's membrane binding. Although the predicted dimerization mode in this paper is not stable, the similar interaction patterns of GABARAP monomer and this hypothetical dimer with membrane still provide a certain reference significance for the dimerization prediction work in the future.

Acknowledgements

We thank Dr. Oliver Weiergräber for helping us created the GABARAP dimer model. This study was supported by Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)-project number 267205415-CRC 1208. Xue Wang is supported by the China Scholarship Council.

References

[1] J. D. Rabinowitz and E. White, “Autophagy and Metabolism,” vol. 1029, no. December, pp. 1344–1349, 2010.

[2] C. W. Yun and S. H. Lee, “The Roles of Autophagy in Cancer.,” *Int. J. Mol. Sci.*, vol. 19, no. 11, Nov. 2018.

[3] C. Dinkins, M. Pilli, and J. H. Kehrl, “Roles of autophagy in HIV infection,” *Immunol. Cell Biol.*, vol. 93, no. 1, pp. 11–17, 2015.

[4] F. M. Menzies *et al.*, “Autophagy and Neurodegeneration: Pathogenic Mechanisms and Therapeutic Opportunities.,” *Neuron*, vol. 93, no. 5, pp. 1015–1034, Mar. 2017.

[5] J. H. Hurley and L. N. Young, “Mechanisms of autophagy initiation,” *Annu. Rev. Biochem.*, 2017.

[6] H. Nakatogawa, Y. Ichimura, and Y. Ohsumi, “Atg8, a Ubiquitin-like Protein Required for Autophagosome Formation, Mediates Membrane Tethering and Hemifusion,” *Cell*, vol. 130, no. 1, pp. 165–178, 2007.

[7] T. Shpilka, H. Weidberg, S. Pietrokovski, and Z. Elazar, “Atg8: an autophagy-related ubiquitin-like protein family.,” *Genome Biol.*, vol. 12, no. 7, p. 226, Jul. 2011.

[8] C. Kraft *et al.*, “Binding of the Atg1/ULK1 kinase to the ubiquitin-like protein Atg8 regulates autophagy,” *EMBO J.*, vol. 31, no. 18, pp. 3691–3703, Sep. 2012.

[9] S. Pankiv *et al.*, “FYCO1 is a Rab7 effector that binds to LC3 and PI3P to mediate microtubule plus end–directed vesicle transport,” *J. Cell Biol.*, vol. 188, no. 2, pp. 253–269, Jan. 2010.

[10] D. G. McEwan *et al.*, “PLEKHM1 regulates autophagosome-lysosome fusion through HOPS complex and LC3/GABARAP proteins.,” *Mol. Cell*, vol. 57, no. 1, pp. 39–54, Jan. 2015.

[11] V. V Rogov *et al.*, “Structural and functional analysis of the GABARAP interaction motif (GIM),” *EMBO Rep.*, vol. 18, no. 8, pp. 1382–1396, Aug. 2017.

[12] Y. Ichimura *et al.*, “A ubiquitin-like system mediates protein lipidation,” *Nature*, vol. 408, no. 6811, pp. 488–492, 2000.

[13] N. Mizushima and M. Komatsu, “Autophagy: renovation of cells and tissues.,” *Cell*, vol. 147, no. 4, pp. 728–741, Nov. 2011.

[14] Y. Paz, Z. Elazar, and D. Fass, “Structure of GATE-16, Membrane Transport Modulator and Mammalian Ortholog of Autophagocytosis Factor Aut7p *,” *J. Biol. Chem.*, vol. 275, no. 33, pp. 25445–25450, Aug. 2000.

[15] V. N. Bavro, M. Sola, A. Bracher, M. Kneussel, H. Betz, and W. Weissenhorn, “Crystal structure of the GABAA-receptor-associated protein, GABARAP,” *EMBO Rep.*, vol. 3, no. 2, pp. 183–189, Feb. 2002.

[16] A. Boeske *et al.*, “Direct binding to GABARAP family members is essential for HIV-1 Nef plasma membrane localization,” *Sci. Rep.*, 2017.

[17] L. Thukral, D. Sengupta, A. Ramkumar, D. Murthy, N. Agrawal, and R. S. Gokhale, “The Molecular Mechanism Underlying Recruitment and Insertion of Lipid-

Anchored LC3 Protein into Membranes," *Biophys. J.*, vol. 109, no. 10, pp. 2067–2078, 2015.

[18] J. E. Coyle, S. Qamar, K. R. Rajashankar, and D. B. Nikolov, "Structure of GABARAP in two conformations: Implications for GABAA receptor localization and tubulin binding," *Neuron*, vol. 33, no. 1, pp. 63–74, 2002.

[19] J. Nymann-Andersen, H. Wang, and R. W. Olsen, "Biochemical identification of the binding domain in the GABAA receptor-associated protein (GABARAP) mediating dimer formation," *Neuropharmacology*, vol. 43, no. 4, pp. 476–481, 2002.

[20] J. A. McCammon, B. R. Gelin, and M. Karplus, "Dynamics of folded proteins," *Nature*, vol. 267, no. 5612, pp. 585–590, 1977.

[21] A. D. MacKerell *et al.*, "All-atom empirical potential for molecular modeling and dynamics studies of proteins.," *J. Phys. Chem. B*, vol. 102, no. 18, pp. 3586–3616, Apr. 1998.

[22] W. D. Cornell *et al.*, "A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules," *J. Am. Chem. Soc.*, vol. 117, no. 19, pp. 5179–5197, May 1995.

[23] N. Schmid *et al.*, "Definition and testing of the GROMOS force-field versions 54A7 and 54B7.," *Eur. Biophys. J.*, vol. 40, no. 7, pp. 843–856, Jul. 2011.

[24] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, "Development and testing of a general amber force field," *J. Comput. Chem.*, vol. 25, no. 9, pp. 1157–1174, Jul. 2004.

[25] H. Sun, "COMPASS: An ab Initio Force-Field Optimized for Condensed-Phase ApplicationsOverview with Details on Alkane and Benzene Compounds," *J. Phys. Chem. B*, vol. 102, no. 38, pp. 7338–7364, Sep. 1998.

[26] S. Rauscher, V. Gapsys, M. J. Gajda, M. Zweckstetter, B. L. De Groot, and H. Grubmüller, "Structural ensembles of intrinsically disordered proteins depend strongly on force field: A comparison to experiment," *J. Chem. Theory Comput.*, vol. 11, no. 11, pp. 5513–5524, 2015.

[27] A. E. Aliev and D. Courtier-Murias, "Correction to 'Experimental Verification of Force Fields for Molecular Dynamics Simulations Using Gly-Pro-Gly-Gly,'" *J. Phys. Chem. B*, vol. 116, no. 27, pp. 7996–7996, 2012.

[28] S. Boonstra, P. R. Onck, and E. Van Der Giessen, "CHARMM TIP3P Water Model Suppresses Peptide Folding by Solvating the Unfolded State," *J. Phys. Chem. B*, vol. 120, no. 15, pp. 3692–3698, 2016.

[29] R. B. Best, W. Zheng, and J. Mittal, "Balanced protein-water interactions improve properties of disordered proteins and non-specific protein association," *J. Chem. Theory Comput.*, vol. 10, no. 11, pp. 5113–5124, 2014.

[30] C. P. Lawrence and J. L. Skinner, "Flexible TIP4P model for molecular dynamics simulation of liquid water," *Chem. Phys. Lett.*, vol. 372, no. 5, pp. 842–847, 2003.

[31] A. E. Aliev, M. Kulke, H. S. Khaneja, V. Chudasama, T. D. Sheppard, and R. M. Lanigan, "Motional timescale predictions by molecular dynamics simulations: case study using proline and hydroxyproline sidechain dynamics.," *Proteins*, vol. 82, no. 2, pp. 195–215, Feb. 2014.

[32] M. Buck, S. Bouguet-Bonnet, R. W. Pastor, and A. D. MacKerell, “Importance of the CMAP correction to the CHARMM22 protein force field: dynamics of hen lysozyme,” *Biophys. J.*, vol. 90, no. 4, pp. L36-8, 2006.

[33] S. Piana, A. G. Donchev, P. Robustelli, and D. E. Shaw, “Water Dispersion Interactions Strongly Influence Simulated Structural Properties of Disordered Protein States,” *J. Phys. Chem. B*, vol. 119, no. 16, pp. 5113–5123, Apr. 2015.

[34] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen, “GROMACS: Fast, flexible, and free,” *J. Comput. Chem.*, vol. 26, no. 16, pp. 1701–1718, Dec. 2005.

[35] L. L. C. Schrödinger, “The PyMOL Molecular Graphics System, Version 1.8,” 2015.

[36] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, “A smooth particle mesh Ewald method,” *J. Chem. Phys.*, vol. 103, no. 19, pp. 8577–8593, Nov. 1995.

[37] B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije, “LINCS: A linear constraint solver for molecular simulations,” *J. Comput. Chem.*, vol. 18, no. 12, pp. 1463–1472, Sep. 1997.

[38] M. Parrinello and A. Rahman, “Polymorphic transitions in single crystals: A new molecular dynamics method,” *J. Appl. Phys.*, vol. 52, no. 12, pp. 7182–7190, Dec. 1981.

[39] J. Lee *et al.*, “CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field,” *J. Chem. Theory Comput.*, vol. 12, no. 1, pp. 405–413, Jan. 2016.

[40] J. P. M. Jämbeck and A. P. Lyubartsev, “Derivation and Systematic Validation of a Refined All-Atom Force Field for Phosphatidylcholine Lipids,” *J. Phys. Chem. B*, vol. 116, no. 10, pp. 3164–3179, Mar. 2012.

[41] P. H. Hünenberger, “Thermostat Algorithms for Molecular Dynamics Simulations BT - Advanced Computer Simulation: Approaches for Soft Matter Sciences I,” C. Dr. Holm and K. Prof. Dr. Kremer, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 105–149.

[42] P. Maragakis *et al.*, “Microsecond molecular dynamics simulation shows effect of slow loop dynamics on backbone amide order parameters of proteins,” *J. Phys. Chem. B*, vol. 112, no. 19, pp. 6155–6158, 2008.

[43] C. Möckel *et al.*, “Integrated nmr, fluorescence, and molecular dynamics benchmark study of protein mechanics and hydrodynamics,” *J. Phys. Chem. B*, vol. 123, no. 7, pp. 1453–1480, 2019.

4.2 Manuscript II

Domain motions, dimerization, and membrane interactions of the murine guanylate binding protein 2

Jennifer Loschwitz, Xue Wang, and Birgit Strodel

bioRxiv preprint. doi: https://doi.org/10.1101/2022.04.27.489784

X.W. contributed to this paper by performing and analyzing the MD simulations of the mGBP2 monomer anchored to lipid membranes.

Domain motions, dimerization, and membrane interactions of the murine guanylate binding protein 2

Jennifer Loschwitz[1,2], **Xue Wang**[1,2], **and Birgit Strodel**[1,2,*]

[1]Institute of Theoretical and Computational Chemistry, Heinrich Heine University Düsseldorf, 40225 Düsseldorf, Germany
[2]Institute of Biological Information Processing: Structural Biochemistry (IBI-7), Forschungszentrum Jülich, 52425 Jülich, Germany
[*]b.strodel@fz-juelich.de

ABSTRACT

Guanylate-binding proteins (GBPs) are a group of interferon-inducible GTPases that are essential components of cell-autonomous immunity against intracellular pathogens. They restrict the replication of intracellular pathogens by targeting and lysing the membrane of the pathogen-containing vacuole. The aim of our work is to provide a molecular-level understanding of how the GBPs destroy the parasitophorous vacuole membrane. To this end, we use molecular dynamics (MD) simulations to elucidate the domain motions of GBPs and the effects on these motions following GTP binding, GBP multimerization, and membrane binding, assuming that each of these steps plays important roles in the defense mechanism. Here, we present our results for the murine GBP2, one of the eleven GBPs present in mice, obtained from standard and replica exchange MD simulations with an accumulated simulation time of 30 μs. The main findings are that mGBP2 features a large-scale hinge motion in its M/E domain which is not hugely affected by either GTP binding, geranylgeranylation, homodimer formation, or lipid-membrane binding. Our prognosis is that this hinge motion has the potential to destroy the membrane following the formation of supramolecular mGBP2 complexes on the membrane surface.

1 Introduction

The superfamily of dynamin-like proteins includes the 65–73 kDa guanylate binding proteins (GBPs) that are involved in various structural changes for remodeling cellular membranes or causing fusion/fission events via guanosine-5'-triphosphate (GTP) binding and hydrolysis[1–4]. GBPs are induced by interferon-γ and they accumulate at membranes of intracellular pathogens and bacteria to initiate host cell survival responses[5–9], including in autoimmune and inflammatory diseases[4,10,11]. Various pathogens are targted by GBPs, including the parasitophorous vacuole membrane of the protozoan parasite *Toxoplasma gondii*[12–16]. To date, the process of how the GBPs bind to and destroy the target's membrane during the host response as well as the role of GTP binding in it are still poorly understood[12,13,17]. What is clear, however, is that for human guanylate binding proteins (hGBPs) GTP binding shifts the monomer/dimer equilibration toward the dimer and even to polymerization[8,18–20].

Until now, seven hGBPs (hGBP1 to hGBP7) and eleven murine GBPs (mGBP1 to mGBP11) are known[21–23]. In this study, we will concentrate on mGBP2, which is the homologue of hGBP1 with a sequence identity of 66.2% for the whole protein. The structure of hGBP1 has been resolved as nucleotide-free (PDB ID 1DG3, 592 residues)[24] and nucleotide-bound monomer (PDB ID: 1F5N)[25], while the structure of mGBP2 was not determined yet by experimental means, but can be modeled using its homology to hGBP1. Generally, GBPs have between 586 and 638 residues, and contain three domains. This is illustrated taking our example of interest, mGBP2, in Fig. 1**a**. The proteins have a GTPase (G), a middle (M) and an effector (E) domain. The N-terminal GTPase domain is generally large (residue 1–303 in mGBP2), binds GTP and hydrolyzes it to GMP via GDP as intermediate. The middle domain (residues 304–481) gives the elongated structure and finally, the effector domain (482–589) is responsible for membrane interaction.

The GTP binding site in the G domain involves four conserved sequence elements called G motifs (Fig. 1**a**, Supplementary Tab. S1). These are: the canonical G1 motif GXXXXGK(S/T), also called phosphate-binding loop (P-L), the switch1/G2 (SW1) motif, the phosphate- and Mg^{2+}-binding switch2/G3 (SW2) DXXG sequence, and the nucleotide-specificity providing G4 motif X(V/L)RD with the loop 2 (G4+L2)[12,23,26]. The phosphate-binding loop has an integrated arginine finger R48 and the residue K51 for GTP-hydrolysis, where the K51A mutant leads to a dysfunctional mGBP2[12]. Like hGBP1 and also mGBP7, mGBP2 can hydrolyze GTP to GMP via GDP by moving the β-phosphate of GDP into the γ-phosphate position for the hydrolysis[12,26–28]. This contrasts with other GTPases, such as Ras and other dynamin-like proteins, where the hydrolysis is limited to GTP. In hGBP1, the S73/T75 residues, which are part of the SW1 motif, form a hydrogen bond network, together with E99 in SW2 and with water molecules, to stabilize the Mg^{2+} ion affiliated with the γ-phosphate of GTP[28]. While E99 and S73 are not directly involved in the hydrolysis, replacing them with alanine reduces the enzymatic efficiency[29]. Finally, D182, which is part of the G4+L2 motif shall be specially mentioned, as it is responsible for the nucleotide preference for GTP in mGBP2[12] and in hGBP1[25]. Another key structural element of the G domain is the guanine cap involving residues 235–256. From the crystal structure of the G domain dimer of hGBP1 (PDB ID: 2BC9) it is known that the protein dimerization is mainly realized by the two guanine caps interacting with each other, involving a transition toward a closed conformation of this loop region[27]

The M domain of both hGBP1 and mGBP2 is a two-helix bundle (α7, α8 and α10, α11) connected with a long helix (α9). This bundle of helices gives the GBPs an elongated shape, at up to 120 Å long, and stabilizes the α12 of the E domain (Fig. 1**a**, Supplementary Tab. S1)[24]. The adjacent E domain with the long helix α12 and a short helix α13 is stretched back to the G domain along to the M domain, which is stabilized by contacts with both the M and the G domain. Of particular interest are salt bridges formed between α4' of the G domain and α13 as well as hydrophobic contacts between loop 1 of the G domain and the E domain (Fig. 1**a**). In mGBP2 these electrostatic contacts involve residues R225 and K226 of α4', E554 and E561 of α12, and E566 and E573 of α13. In hGBP1, the analagous interactions likely break during the conformational change following GTP binding[8,30,31].

When mice are infected, by the intracellular protozoan *T. gondii*, some of the eleven murine GBPs are highly expressed and they form dimers and larger multimers in a GTP-dependent manner[7,32]. It was demonstrated that hetero- and homo-multimers of mGBPs form vesicle-like structures in the cytosol that destabilize the membrane of the parasitophorous vacuole, in which the parasite hides from the host cell response[7,12,13,22]. This membrane consists of parts of the host plasma membrane. The only high-resolution structural information that is available for GBP dimers is the crystal structure of the G domain dimer of the hGBP1 (PDB ID: 2B92)[27]. Based on this structure we built models for the full-length dimers of hGBP1[31] and mGBP2 (Fig. 1**b**), which span a length of about ∼240 Å. The dimerization interface mainly involves the guanine cap regions of both proteins. Moreover, dimerization has a strong effect on the self-activation of GTP hydrolysis, in which two arginines that are part of the dimer interface, R240 and R244 play a key role[33]. Association with the membrane of the parasitophorous vacuole is most likely mediated by the C-terminal lipid anchor that mGBP2 features, as does hGBP1. Both proteins have a CaaX motif for post-translational modification by a farnesyl group (hGBP1) or a geranylgeranyl group (mGBP2)[13,21,30,34,35]. This anchor lies at the end of the E domain; in the case of mGBP2 it is attached to C586 (Fig. 1**c**).

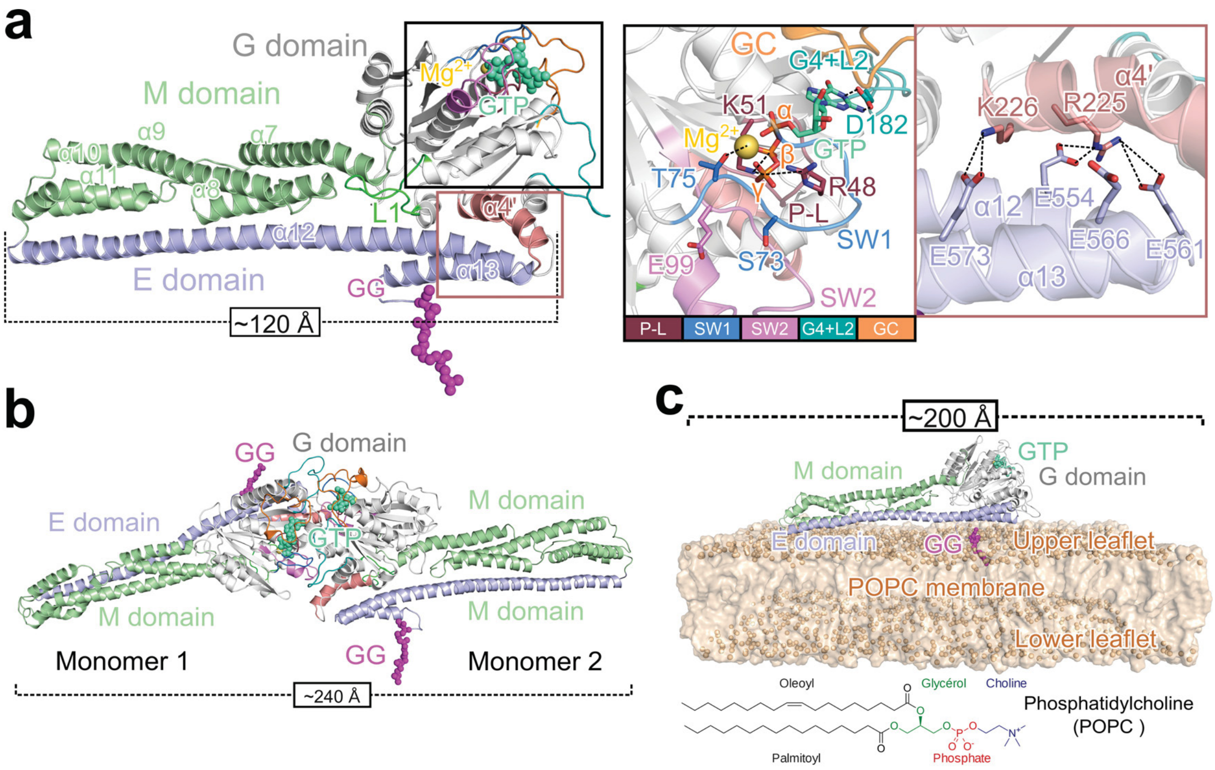


Figure 1. Model systems of mGBP2 investigated in this study. **(a)** The homology model of the entire mGBP2 molecule contains three different domains: (i) the G domain (gray cartoon) with GTP (green spheres) and Mg^{2+} (orange sphere); (ii) the M domain (green cartoon); (iii) the E domain (blue cartoon), and the geranylgeranyl lipid anchor (GG, magenta spheres). The black frame shows a zoomed view of the GTP binding site with the different G motifs and loops being highlighted (see color code at the bottom). Moreover, the residues important for GTP binding and hydrolysis are indicated too, along with their interaction (dashed black lines) with the different parts of GTP (α-, β- and γ-phosphate) or with Mg^{2+}. The brown frame shows the interactions between the α4' of the G domain and α12/13 of the E domain. The side chains of the important residues are shown as sticks and salt bridges are indicated by dashed black lines. **(b)** The model of the mGBP2 dimer, that includes both GTP and the geranylgeranyl group, is based on the crystal structure of the G domain dimer of hGBP1 (PDB ID: 2BC9)[27]. **(c)** The membrane-associated mGBP2 monomer model is anchored via the geranylgeranyl group to the POPC membrane (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine, brown surface, phosphor atoms highlighted as spheres). The chemical structure of POPC is given at the bottom.

The aim of this study is to analyze the conformational dynamics of mGBP2 and how the protein motions are affected by GTP binding, post-translational modification, membrane binding, and dimerization. To this end, we perform all-atom molecular dynamics (MD) simulations, employing the Hamiltonian replica exchange molecular dynamics (HREMD) technique to enhance the conformational sampling. To mimic the different protein states, we simulate the *apo*-protein that has neither GTP bound nor the lipid anchor attached (mGBP2$_{\text{apo}}$), the protein with GTP being bound (mGBP2$_{\text{GTP}}$), and the protein that has GTP bound and the geranylgeranyl anchor attached, which we denote as the *holo*-state (mGBP2$_{\text{holo}}$). As mGBP2 is the homologue of hGBP1, for which we previously uncovered a large-scale hinge motion involving the M and E domains[31], we pay particular attention to identifying similar motions in mGBP2 and how they are affected by GTP binding and the presence of the geranylgeranyl anchor, as this has not been studied before, also not for hGBP1. We further simulate and analyze the dynamics of the mGBP2 monomer, on a simple 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC), where the influence of GTP is being considered by modeling mGBP2 in the GTP-bound and GTP-free state. Finally, the dynamics of the mGBP2 dimer is explored. The simulations of these different systems give us fundamental insights into the dynamics of mGBP2, which is of relevance for its biological functions.

2 Results

2.1 Monomeric mGBP2 in solution

We started the simulations with the monomeric mGBP2 in its apo-state in solution. This serves as a reference to delineate the effects of GTP binding, geranylgeranylation, protein dimerization, and membrane binding by successively including these aspects in the simulations following thereafter. Because mGBP2 is a large protein with almost 600 residues, we performed enhanced sampling by running Hamilitonian replica exchange molecular dynamics (HREMD) simulations involving 16 replicas with 200 ns per replicate simulation (amounting to 3.2 μs cumulated simulation time) for each of the monomer systems to better sample their conformational space. For the analysis, we use the target replica simulated with an unmodified energy function. First, we unravel the overall motions and then investigate in more detail how the different regions of mGBP2 are affected by GTP binding and geranylgeranylation.

2.1.1 Overall flexibility

To evaluate the protein flexibility, we calculated the root mean square fluctuations (RMSF) of the C_α atoms. Fig. 2 shows that GTP and geranylgeranyl have only minor effects on the overall dynamics of mGBP2. For being better able to correlate the dynamics with the mGBP2 structure, we projected the RMSF values onto the structure from which the MD simulations were started (Supplementary Fig. S1**a–c**). We observed that mGBP2 exhibits very similar dynamics to hGBP1,[31] with the core of the G domain being the least flexible. The most flexible part is the tip of the M and E domain (around residue 480) and the neighboring helices $\alpha 10$ (residues 420–446) and $\alpha 12$ (residues 482–561). The dynamics of the M and E domain (M/E domain) derives from a hinge motion of both domains, which will be discussed below. However, the M/E domain of mGBP2 is stiffer than that of hGBP1, as the latter reached RMSF values of up to 16 Å, compared to only 6 Å in mGBP2.

2.1.2 Effects of GTP and geranylgeranyl binding on the G domain.

The dynamics of the G domain loops in mGBP2 without GTP are very similar to their dynamics in the GTP-free hGBP1[31]. In particular, the SW1 and guanine cap are very flexible with RMSF values above 5 Å, while GTP binding makes these loops more rigid (Fig. 2). To quantify the dynamics of the G domain, we clustered the conformations of each loop by using a cutoff of 2.5 Å with a fitted trajectory on the β-sheets of the G domain. The results of this analysis are summarized in Tab. 1 and shown in Fig. 3.

The two most flexible loops are SW1 and the guanine cap, which can both switch between open and closed forms and adopt all kinds of intermediate states. In SW1, residue S73 that is important for GTP binding by stabilization of Mg^{2+} is flexible in the absence thereof and can also interact with R48, which is the arginine finger needed for the GTP hydrolysis. Without GTP being bound, R48 is very flexible too and adopts conformations covering the whole GTP binding site, thereby also interacting with the negatively charged residue E249 of the guanine cap and E105 of SW2. The flexibility of SW1 creates space for the the guanine cap in the GTP binding site, which can therefore adopt all kinds of conformations (Fig. 3**a**). In contrast, the residues K51 in the phosphate-binding loop, T75 in SW1, E99 in SW2, and D182 in the G4 motif/L2 loop are not very flexible despite the absence of GTP. These residues were also fixed in the GTP-free hGBP1, from which we concluded that the phosphate-binding loop and SW2 do not require GTP for adopting a closed conformation as they can be closed without GTP[31]. However, GTP binding is assumed to shift the equilibrium toward the closed states of SW1 and the guanine cap.

The current results obtained for mGBP2 in the presence of GTP confirm that this is indeed the case. In mGBP2$_{\mathrm{GTP}}$, SW1 adopts only a single conformation, which is the closed one that is stabilized by key interactions: R48, S73 and K51 interact with the γ-phosphate of GTP and T75 has contacts with Mg^{2+} (Fig. 3**b**). Residue D182 (in the G4 motif/L2 loop), on the other hand, switches between conformations, where it interacts with GTP, and orientations that are turned away from GTP for interacting with R238 in the guanine cap. Residue E99 (in SW2) is always pointed away from GTP. For hGBP1 it was shown that E99 instead forms a hydrogen-bond network involving two water molecules and S73[28].

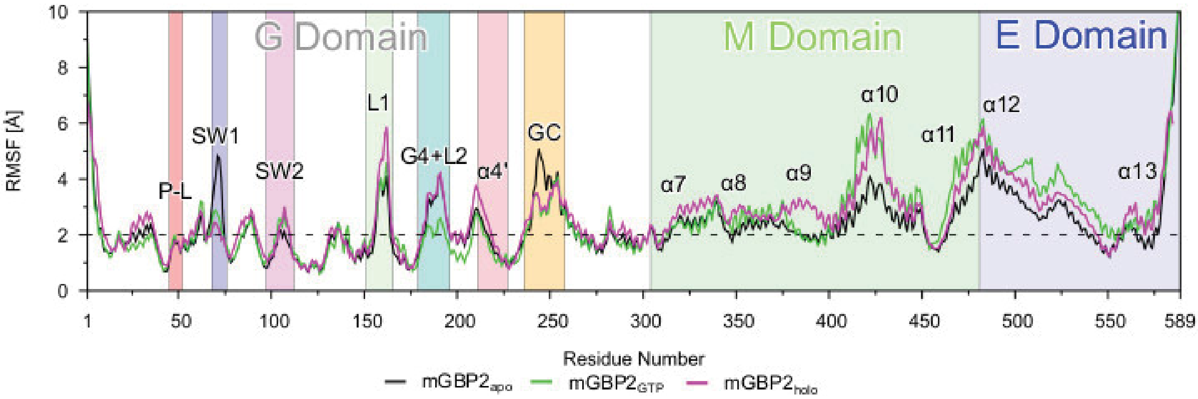


Figure 2. The root mean square fluctuations per residue of monomeric mGBP2 obtained from HREMD simulations of mGBP2$_{\mathrm{apo}}$ (black), mGBP2$_{\mathrm{GTP}}$ (green), and mGBP2$_{\mathrm{holo}}$ (magenta), as sampled in the target replica. All important motifs, loops, and helices are labeled and a background color added using the same colors as for the corresponding structural units as in Fig. 1. Abbreviation: P-L = phosphate-binding loop, GC = guanine cap.

Table 1. Flexibility of the G domain loops of monomeric mGBP2 determined by clustering of the conformations sampled in the HREMD target replica. Abbreviation: P-L = phosphate-binding loop, GC = guanine cap.

	mGBP2$_{\mathrm{apo}}$			**mGBP2$_{\mathrm{GTP}}$**			**mGBP2$_{\mathrm{holo}}$**		
(Motif-)Loop	**clusters**	**population[%]**[a]	**RMSD[Å]**[b]	**clusters**	**population[%]**[a]	**RMSD[Å]**[b]	**clusters**	**population[%]**[a]	**RMSD[Å]**[b]
P-L	1	100	–	1	100	–	1	100	–
SW1	23	82.4	9.9	1	100	–	2	100	3.6
SW2	3	100	4.9	1	100	–	3	100	4.9
L1	23	89.6	10.8	17	92.2	9.4	25	76.8	13.1
G4+L2	12	95.9	7.7	3	100	4.5	12	96.3	8.8
α4’	2	100	3.4	3	100	5.2	5	99.9	6.1
GC	29	66.1	8.9	3	100	5.3	8	91.7	6.4

[a] Percentage of the structures, which are cumulatively represented by the first three clusters.
[b] The largest RMSD found between two clusters.

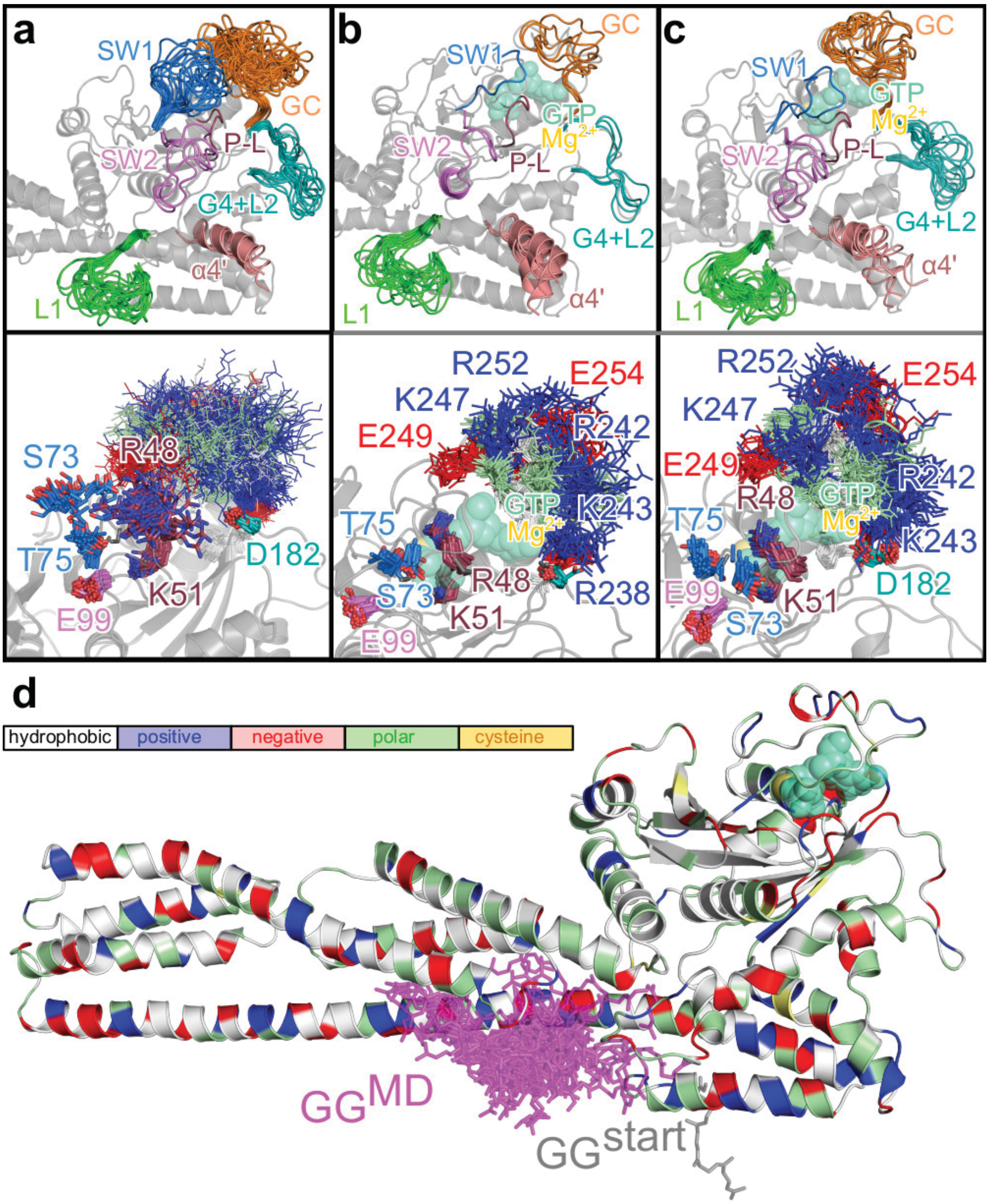


Figure 3. Conformational clusters of the G domain loops of (a) mGBP2$_{apo}$, (b) mGBP2$_{GTP}$, and (c) mGBP2$_{holo}$ as well as (d) for the geranylgeranyl anchor. (a-c) In the upper panels, each central cluster conformation is shown, using different colors for the different loops as indicated in the figures. For mGBP2$_{GTP}$ and mGBP2$_{holo}$, the GTP-binding pocket is indicated by showing GTP and Mg^{2+} (in transparent green and light orange). The helix α4' is also highlighted in light red. The homology model of mGBP2 is shown as reference in a gray cartoon representation. In the lower panels, the different conformations of key residues are shown, as sticks colored according to their residue type (white: apolar; green: polar; blue: positively charged; red: negatively charged), and labeled. Abbreviation: P-L = phosphate-binding loop, GC = guanine cap. **(d)** The central cluster conformations of the geranylgeranyl group (GG, magenta sticks) as obtained from the HREMD simulation of mGBP2$_{holo}$ are shown. The initial conformation of the geranylgeranyl group used to start the HREMD simulation is also depicted (gray sticks), while the homology model of the whole protein is presented as cartoon and colored based on residue type as explained in the legend box.

As a result of the stable interactions and conformations of S73 and R48, they are no longer available for interactions with residues of the guanine cap. This, in turn, causes the guanine cap to adopt an ordered conformation, which is stabilized by interactions among the guanine cap residues, in particular between the the positively charged residues R242, K243, K247, and R252 that form salt bridges with the negatively charged residues E249, E253, E254, E255 (Fig. 3**b**). When the geranylgeranyl group is added to the picture, leading to mGBP2$_{\text{holo}}$, most loops are slightly more flexible than those in mGBP2$_{\text{GTP}}$ (Fig. 3**c**). We will see shortly that this is because of interactions of the geranylgeranyl lipid anchor with the G domain. Nonetheless, despite this increased flexibility, the two conformational clusters identified for SW1 in mGBP2$_{\text{holo}}$ all belong to the closed state. The phosphate-binding loop containing the residues R48 and K51 for GTP hydrolysis is already quite stable without GTP being bound. In mGBP2$_{\text{GTP}}$ and mGBP2$_{\text{holo}}$ this loop becomes even more rigid (RMSF values under 2 Å). In both cases, only one conformational cluster was found for the loop. Similarly, SW2 and D182 from the G4 motif also rigidified in the presence of GTP. Thus, the interplay between the G domain and GTP as well as Mg^{2+} stabilize the loops surrounding the GTP binding site in a conformation optimal for the subsequent hydrolysis reaction. Different behavior between mGBP2$_{\text{GTP}}$ and mGBP2$_{\text{holo}}$ is observed for the G4 motif/L2 loop. Only in the case of mGBP2$_{\text{GTP}}$, does GTP binding cause this loop region to become stable; only three conformational clusters are being found here (Tab. 1). This stability results from a β-hairpin conformation that is formed in L2, which in turn includes three stable backbone hydrogen bonds between L187 and V194, V189 and K192, and S196 and L200 as well as a hydrogen bond between the side chain or backbone of T195 and the side chain of E198. This conformation is further stabilized by a hydrophobic cavity made of the residues of a short helix and the adjacent β-hairpin, consisting of F183, L187, V189, V194, Y199, and I235. In mGBP2$_{\text{holo}}$ and mGBP2$_{\text{apo}}$, on the other hand, this G4 motif/L2 loop is flexible with RMSF values between 3 and 4 Å leading to 12 conformational clusters in either case. Here, the β-hairpin is only intermittently adopted.

Next, we turn our attention to α4', as this helix was suggested to undergo conformational changes upon GTP binding and hydrolysis[30]. However, even in GTP-free mGBP2 we find α4' to be flexible. In fact, in each system, this helix prefers a position slightly different from the one predicted by homology modeling (Fig. 3). In the presence of GTP and the geranylgeranyl group, i.e., in mGBP2$_{\text{holo}}$, α4' is particularly flexible, which is visible from its RMSF values of up to 4 Å (Fig. 2). Moreover, it can move up to 18 Å (Supplementary Fig. S2**c**), causing its reversible unfolding between residues K212–Q219. This instability of α4' results from the geranylgeranyl group interacting with loop L1, which in turn affects α4'. In mGBP2$_{\text{holo}}$, loop L1 is the most flexible, as the number of loop L1 clusters is 25 in mGBP2$_{\text{holo}}$, 23 in mGBP2$_{\text{apo}}$, and 17 in mGBP2$_{\text{GTP}}$. Moreover in mGBP2$_{\text{holo}}$, the first three clusters represent only 77% of the L1 conformations, while over 95% of the structures are included in the first three L1 clusters in mGBP2$_{\text{apo}}$ and mGBP2$_{\text{GTP}}$.

Finally, we assess the effect of GTP on the guanine cap, which we identify as not as strong as for the other motifs/loops that are in direct contact with GTP. However, as already explained above, as a consequence of the other motifs/loops becoming more ordered upon GTP binding, this inflicts a certain degree of order in the guanine cap too. Fig. 3**b/c** shows that in mGBP2$_{\text{GTP}}$ and mGBP2$_{\text{holo}}$ the guanine cap adopts a conformation that is closing off the GTP binding site. The clustering analysis also reveals that this loop also becomes more rigid, as evidenced by the number of conformational clusters decreasing from 29 in mGBP2$_{\text{apo}}$ to 3 and 8 in mGBP2$_{\text{GTP}}$ and mGBP2$_{\text{holo}}$, respectively. The increased stability of the guanine cap following GTP binding should facilitate the dimerization of mGBP2, as the protein-protein interface in the dimer mainly involves the guanine caps of both proteins. Because GTP binding shifts the conformational equilibrium of the guanine cap already in the monomer towards the one in the dimer, it can be expected that this conformation shift facilitates protein-protein recognition for dimer formation. Indeed we do demonstrate this to be the case in the dimer analysis section below.

In summary, all loops of the G domain become more rigid upon GTP binding. Some of the conformational stiffening is being undone by geranylgeranylation, which especially applies to loops L1 and G4+L2 as well as helix α4', due to their direct or indirect interactions with the geranylgeranyl group. Overall, mGBP2$_{\text{GTP}}$ seems the most stable monomeric protein state compared to the two other states, mGBP2$_{\text{apo}}$ and mGBP2$_{\text{holo}}$.

2.1.3 Motions of the geranylgeranyl lipid anchor

To better understand the effects of the geranylgeranyl lipid anchor on the structural dynamics of mGBP2, we analyze its motions in detail. We attached the geranylgeranyl group at the C-terminal end of mGBP2 such that it pointed away from the protein (Fig. 1**a**). However, from this position it readily moved away, during the HREMD simulation of mGBP2$_{\text{holo}}$, in order to avoid being completely water-exposed as it has mostly hydrophobic moieties. The clustering analysis revealed the geranylgeranyl anchor to prefer positions close to helices α7, α9 and α12, whose hydrophobic residues attract the anchor (Fig. 3**d**). Interestingly, the geranylgeranyl anchor did not move towards the most hydrophobic part of mGBP2, which is a hydrophobic cave formed by the G domain helices α4' and α6 as well as L1 and the β-sheet behind them, along with the first half of α7 and the loop leading to α7. Nonetheless, the geranylgeranyl anchor does interact with the L1 loop, which, as a result, becomes more flexible (as discussed above). The interactions of the geranylgeranyl anchor with the α7, and also the helices α9, α10, and α13, which are all part of the M or E domain, cause an increase in the motions of these two domains, which will be analyzed in more detail below. The geranylgeranyl anchor is rather flexible, as supported by 79 conformational clusters representing only 39.4% of the geranylgeranyl conformations that we identified from applying conformational clustering with a 2.0 Å cutoff to the HREMD simulation of mGBP2$_{\text{holo}}$ (Tab. 1 and Fig. 3**d**). This flexibility derives from the conformational freedom of the last seven residues (K578–K585) of mGBP2 that are located between α13 and the geranylgeranyl anchor.

Taken together, these data indicate that the geranylgeranyl anchor is highly flexible and moves to a more hydrophobic cavity formed by residues of the M and E domain. However, it does not constantly bind to specific residues but it keeps its flexibility.

2.1.4 Effects of GTP binding and geranylgeranylation on the M/E domain

Next, we had a closer look at how GTP and the geranylgeranyl lipid anchor influence the M/E domain of mGBP2. Therefore, we applied principal component analysis (PCA) to the M/E domain motions. In the following we focus on the two eigenvectors PC1 and PC2 that represent the domain's two main motions. In order to allow for a direct comparability between the eigenvectors of the three monomeric systems under study, i.e., mGBP2$_{\text{apo}}$, mGBP2$_{\text{GTP}}$, and mGBP2$_{\text{holo}}$, we applied the PCA to the concatenated target-replica trajectories of the three HREMD simulations and then projected the individual trajectories onto the common eigenvectors. The resulting plots are shown in Fig. 4**a** and characteristics of these projections are summarized in Supplementary Tab. S2. In general, the distribution of the conformations along PC1 and PC2 is very similar in the three systems. The main difference between them is the amplitude of the motions. In the mGBP2$_{\text{apo}}$, the value range is between -12.1 nm^2 (PC1$_{\text{min}}$) and 8.5 nm^2 (PC1$_{\text{max}}$) for PC1 as well as -8.1 nm^2 (PC2$_{\text{min}}$) and 10.9 nm^2 (PC2$_{\text{max}}$) for PC2, which serve as reference. Upon GTP binding, the deflection along PC1 increased, with PC1$_{\text{max}}$ reaching 12.7 nm^2 (ΔPC1$_{\text{max}} = 4.2$ nm^2), whereas all other values lie in the same area as for mGBP2$_{\text{apo}}$. When the geranylgeranyl anchor is further added to the protein, this main motion became even more pronounced: the mGBP2$_{\text{holo}}$ has an almost two-fold higher PC1$_{\text{max}}$ value of 16.9 nm^2 (ΔPC1$_{\text{max}} = 8.4$ nm^2) compared to mGBP2$_{\text{apo}}$. On the other hand, the motion along PC2 is somewhat reduced in this protein, as PC2$_{\text{max}}$ only reaches a value of 9.4 nm^2 (ΔPC2$_{\text{max}} = -1.5$ nm^2). In order to depict the two main motions of the proteins, in Fig. 4**b** the mGBP2 conformations corresponding to the minimal and maximal PC1/PC2 values are shown. One can see that the motion along PC1 stems from a hinge movement of the M/E domain, similar to the one that we had already identified for hGBP1[31]. The motion along PC2, on the other hand, is a twist movement of the M/E domain. Hence, addition of GTP and geranylgeranylation have a particular enhancing effect on the hinge motion. This is not only confirmed by the larger RMSF values observed for the M/E domain of the mGBP2$_{\text{GTP}}$ and mGBP2$_{\text{holo}}$ compared to the mGBP2$_{\text{apo}}$ (Supplementary Fig. S1**a–c**), but are also witnessed by the maximal deviations of the tip of the M/E domain from the starting structure of the HREMD simulations (~ 33 Å for mGBP2$_{\text{apo}}$ vs. > 42 Å for mGBP2$_{\text{GTP}}$ and mGBP2$_{\text{holo}}$).

To further quantify the motions of the M/E domain, we calculated the change in Cartesian coordinates (Δx, Δy, Δz) of residue L480, which defines the tip of the M/E domain. Fig. 5**a** shows the reference conformation in the Cartesian room, with the location of the L480 being highlighted. Fig. 5**b** reveals that the hinge motion is mainly characterized by motions into positive y direction, corresponding of the tip of the M/E domain moving toward the G domain. The average values for the mGBP2$_{\text{apo}}$ are $\Delta x = -4.2 \pm 0.02$ Å, $\Delta y = 26.0 \pm 0.07$ Å, and $\Delta z = -6.4 \pm 0.02$ Å. In Supplementary Tab. S2, also the

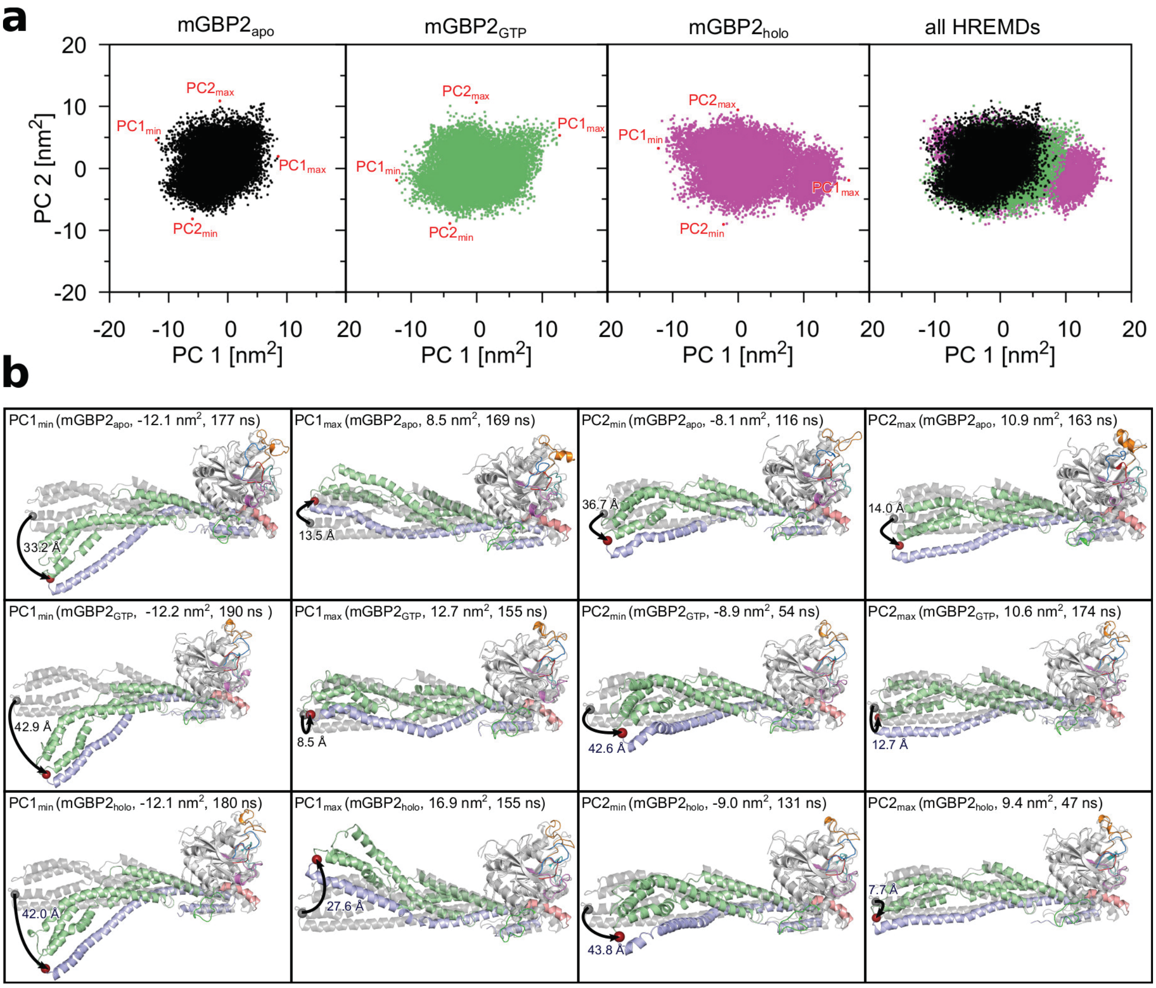


Figure 4. Main motions of the M and E domain in monomeric mGBP2 identified by PCA of the target replica of the HREMD simulations. (a) The projection of the conformations collected in the target replica onto the first two principal components (PC1 and PC2). The results for GTP-free mGBP2 (black), mGBP2 with GTP bound (green), and mGBP2 with both GTP and geranylgeranyl anchor (magenta) are shown from left to right. For a better comparability, all three projections are shown together in the rightmost panel. The minimum and maximum values of the projections are indicated by red dots and labeled. **(b)** The conformations at the minimal and maximal PC1 and PC2 values are shown for the mGBP2$_{\text{apo}}$ (top), mGBP2$_{\text{GTP}}$ (middle) and mGBP2$_{\text{holo}}$ (bottom). The corresponding PC1 and PC2 values are provided too, along with the simulation time at which these conformations were sampled. The motions of the tip of the M/E domain relative to the starting structure (gray cartoon) are indicated by arrows and the maximal amplitudes (in Å) are given.

minimum and maximum changes are listed. For the mGBP2$_{\text{GTP}}$, the values are in a similar range, however, the motion into negative Δy direction has grown. Here, values down to -19.0 Å are reached, compared to -8.1 Å in mGBP2$_{\text{apo}}$, which means that the motion can also go into the opposite direction where the tip of the M/E domain moves away from the G domain side that interacts with the α13. With the hinge motion taking place into the opposite direction, the larger PC1 space as seen for mGBP2$_{\text{GTP}}$ compared to mGBP2$_{\text{apo}}$ can be explained. Moreover, in mGBP2$_{\text{GTP}}$ more distinct motions along the z direction are observed too. In the presence of the geranylgeranyl anchor, these two effects are further increased, especially Δz values are even larger. The most extreme motions of the M/E domain are shown by representative snapshots in Supplementary Fig. S2. They show that in the presence of GTP and the geranylgeranyl anchor, the helices α9 and α12 can (reversibly) unfold, involving residues R491–E503 (Supplementary Fig. S1**a–c** and Supplementary Fig. S2**b/c**), leading to a drop in the average helix content in E domain by 3–4% compared to 88% in mGBP2$_{\text{apo}}$ (Supplementary Tab. S2). Hence, GTP binding and the addition of the geranylgeranyl anchor have an influence on the motions of the M/E domain. In the case of the geranylgeranyl anchor, these effects emerge from the direct interactions between that anchor and the M/E domain, whereas in the case of the GTP binding it is an allosteric effect.

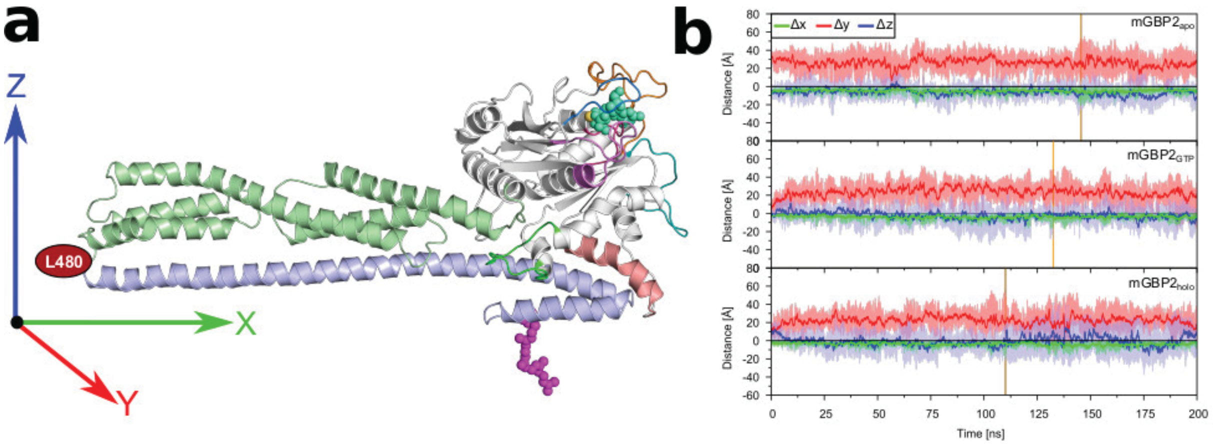


Figure 5. Motions of residue L480 in the HREMD simulations of monomeric mGBP2. (**a**) Reference conformation of the mGBP2 is shown for the calculation of the coordinate distances difference (Δx, Δy, Δz) of L480. **(b)** The evolution of these distances (Δx: green, Δy: red, Δz: blue) is shown for mGBP2$_{\text{apo}}$ (top), mGBP2$_{\text{GTP}}$ (middle), and mGBP2$_{\text{holo}}$ (bottom). The solid lines show the running average and the shades illustrate the raw data. Furthermore, the zero line is shown in black, whereas the vertical orange line indicates the frame with the maximal distance of the L480 residue between reference structure and 200 ns HREMD simulations.

2.1.5 Convergence tests

In order to test if the HREMD simulations involving 16 × 200 ns had indeed converged, we performed for mGBP2$_{\text{holo}}$ another HREMD simulation employing 40 replicas of 400 ns each, amounting to 16 μs in total for that simulation. As before, the analysis was performed for the target replica. We further performed a standard 1 μs MD simulation of the mGBP2$_{\text{holo}}$ monomer. This simulation serves as a test of how much the conformational space of the protein can be sampled without enhancing the simulation technique. This question is of relevance for our simulations of the mGBP2 dimer and membrane-bound mGBP2, where the systems were too large for running HREMD simulations.

For both simulations we performed the same kind of analysis as done for the 16 × 200 ns HREMD simulation. In Supplementary Fig. S3**a** the RMSF values are shown. In both the 40 × 400 ns HREMD and the 1 μs MD simulation the overall flexibility of the protein is very similar as in the initial HREMD simulation. The conformational space sampled by both HREMD simulations is in particular very similar, as confirmed by the RMSF values projected onto the protein (Supplementary Fig. S1**c** vs. Supplementary Fig. S1**d**) and a similar number of clusters identified for the different motifs and loops of the G domain and the geranylgeranyl group (Supplementary Tab. S3). In the 1 μs MD simulation, mGBP2$_{\text{holo}}$ shows lower flexibility (Supplementary Fig. S1**e**) and also has slightly lower numbers of clusters (Supplementary Tab. S3). The effects of the geranylgeranyl anchor on the M/E domain are similar to those observed in the 16 × 200 ns HREMD simulation. The hinge motion is stronger during the 40 × 400 ns HREMD simulation with 69.6 Å maximal displacement, whereas in the 1 μs MD simulation that motion is somewhat reduced to a 44.6 Å amplitude (Supplementary Fig. S4**a/b**). We also performed a PCA (Supplementary Fig. S4**c**) and extracted the conformations corresponding to the minimal and maximal PC1 and PC2 values (Supplementary Fig. S4**d**). Finally, we monitored the motions of M/E domain tip using the residue L480 for this (Supplementary Fig. S4**e** and Supplementary Tab. S2). In the 40 × 400 ns HREMD simulation, the M/E domain motions were of similar width as in the 16×200 ns HREMD simulation, yet shifted toward to more negative values in the x and the y direction. This is also confirmed by the PCA values with $PC1_{min}$ of -14.7 nm^2, but only $PC1_{max}$ of 9.7 nm^2, indicating that the hinge motion is more directed toward the α13 helix than in the shorter HREMD simulation of mGBP2$_{holo}$ (Supplementary Fig. S4**d**). In the 1 μs MD simulation, on the other hand, the hinge motion is less pronounced, as supported by the reduction of the sampled PC1 and PC2 space as well as the smaller Δx, Δy, and Δz values (in absolute terms) indicating the movement of the M/E domain tip.

In conclusion, the computational cost of a 40×400 ns HREMD simulation is very high, while the afforded additional sampling of is minor. Thus, we concluded that the HREMD simulation with 16 replicates of 200 ns simulation was sufficient. Furthermore, the conformational space is better sampled by HREMD simulations but the 1 μs MD simulation shows similar the results for the mGBP2$_{holo}$. Therefore, we can use this time length for the MD simulations of the bigger systems.

2.2 Dimeric mGBP2 in solution

It was demonstrated that mGBPs accumulate as hetero- and homo-multimers, in vesicle-like structures in the cytosol and at the parasitophorous vacuole membrane, during host responses[7]. We therefore wanted to probe the dynamics of the mGBP2 dimer as the smallest multimer. We generated the mGBP2 dimer based on the crystal structure of the G domain dimer that was resolved for hGBP1 (PDB ID: 2BC9)[27]. The resulting dimer structure with monomer 1 and 2 (M1 and M2) is depicted in Fig. 1**b**. We then simulated the dimer for 1 μs. As mGBP2 dimerization was shown to preferentially occur in the presence of GTP, we modeled the protein in its GTP-bound state. We further included the geranylgeranyl group in the setup, considering that the dimer is expected to bind to the parasitophorous vacuole membrane via this lipid anchor. We thus modeled the protein as mGBP2$_{holo}$ in our simulations of the dimer.

2.2.1 Motions in the G domains and interactions in the dimer interface

We first compare the overall flexibility, as quantified by the RMSF values, of the mGBP2$_{holo}$ in the dimer with that of the monomeric mGBP2$_{holo}$, using the 1 μs MD simulation of the monomer for the sake of comparability (Supplementary Fig. S3**b**). In general, the distribution of flexible and rigid protein regions is very similar in the mGBP2$_{holo}$ monomer and dimer, which is also visible from the color-coded projections of the RMSF values onto the protein structures (Supplementary Fig. S1**e/f**). Only subtle differences are observed. As in the monomer, the two G domains of the dimer are stable, except for L1 and α4' in M1 (Supplementary Fig. S1**f**). For the M/E domain, M1 only shows high flexibility at the tip (residues 415–430 and 468–504), whereas M2 exhibits a wholly flexible α12. To characterize the different motions, we executed a clustering with a cutoff value of 2.5 Å for the G domain's structural elements (Fig. 6**a/b**) and for the geranylgeranyl anchor with a cutoff value of 2.0 Å (Fig. 6**c**). The numbers for the resulting clusters and their populations are summarized in Supplementary Tab. S3. The clustering confirms that L1 (34 clusters) in M1 is more flexible than it is in M2 (15 clusters). This can be explained by the different influences of the geranylgeranyl anchor in M1 and M2. In M1, it interacts more with the L1 loop similar to the behavior in the other simulations of mGBP2$_{holo}$ in solution, while in M2 the anchor causes an unfolding in the α12 (residues 523–544), creating a cavity where it does not interact with L1.

Of special interest in the dimer is the guanine cap as it provides the interface between M1 and M2. In the dimer, both guanine caps have the same ordered structure as in the simulations of the mGBP2$_{GTP}$ and mGBP2$_{holo}$ monomer (Fig. 6**a/b**). Also the numbers of clusters are similar, with two/three cluster conformations per guanine cap found in the dimer and two clusters identified in the 1 μs simulation of the mGBP2$_{holo}$ monomer (Supplementary Tab. S3).

To identify the key residues in the dimer interface, we calculated the contact probabilities with a cutoff value of 5 Å between the residues of M1 and M2, which can be seen in Supplementary Fig. S5. The first observation is that several of the interactions between the two proteins are symmetric, such as Y47–Y199, E105–N190/H202 and D237–K243, which results from the symmetric structure of the homodimer. All special interactions like hydrogen bonds and salt bridges are summarized in the Supplementary Tab. S4, while the residues which are close to each other, yet without certain interactions are not listed. Some

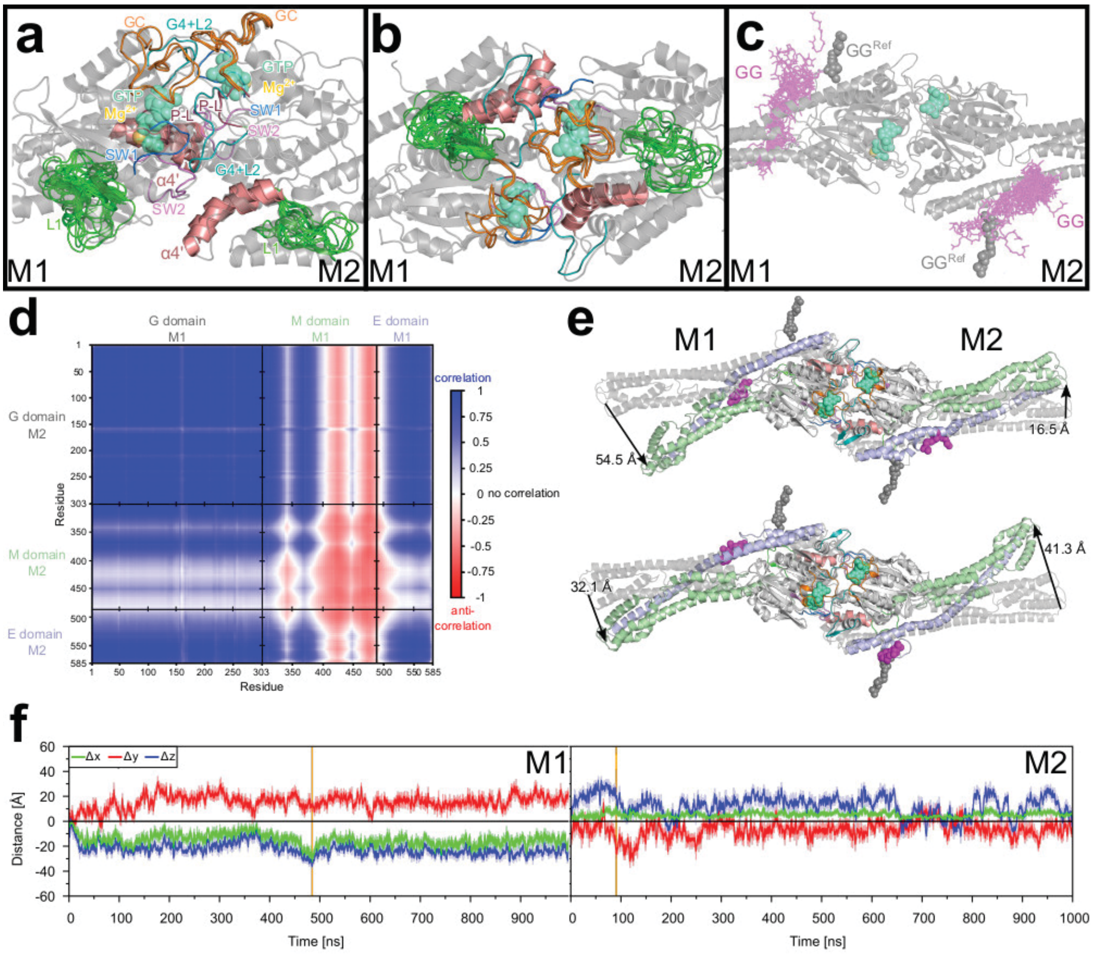


Figure 6. Molecular motions in the mGBP2$_{\mathrm{holo}}$ dimer obtained from a 1 μs MD simulation. Conformational clusters of the G-domain motifs, the L1, and the guanine cap in the side view **(a)** and front view **(b)**. In **(c)** the clusters of the geranylgeranyl lipid anchor (GG) are shown. The same coloring scheme as in Fig. 3 was used. **(d)** The dynamic cross-correlation between the motions of the C$_{\alpha}$ atoms of M1 and M2 is shown. Blue and red colors represent correlation and anti-correlation, respectively. Uncorrelated motions, corresponding to the value 0, are shown in white. The different domains are separated by black lines and the domain names given on the axes. **(e)** The largest deviations from the start structure as sampled in the MD simulation are depicted, and the amplitudes (in Å) for the tip of the M/E domain are given. **(f)** The motions of residue L480 relative to the start structure of the MD simulation are quantified in terms of Δx (green), Δy (red), and Δz (blue) in M1 (left panel) and M2 (right panel) after aligning the trajectory on the G domain of protein M1 and using the same coordinate definition as shown in Fig. 5. The solid lines display the running average and the shades illustrate the raw data. The zero line is shown in black, whereas the vertical orange line indicates the frame with the maximal distance of the L480 residue between reference structure and the simulation.

very stable hydrogen bonds are formed, which can be between backbones, side chains or between a backbone and a side chain. The two guanine caps are held together by stable salt bridges, like M1:D237–M2:K243, M1:E259–M2:R242/K243, and M1:K243–M2:E259 as well as a stable symmetric hydrogen bond between the side chain of R238 and the backbone of D182 in the G4+L2 motif. In the homologous hGBP1, the two arginine residues R240 and R242 are known to influence dimerization, as the protein remains a monomer if these two residues are mutated to alanine[33]. In mGBP2, the corresponding residues could be R238 and R242, but also K243, which form stable interprotein contacts. Furthermore, residues from the other G-domain motifs, in particular the phosphate-binding loop, SW1 and SW2, are also involved in the dimeric interface, where E105 (SW2) builds a stable hydrogen bond network with N190 and H202 (G4+L2) from the other monomer. The α4' is also affected by dimerization, as it interacts with the SW2 motif of the other monomer via a salt bridge (K213–D108), a hydrogen bond (R221–V104), and a hydrophobic contact (F217–V104).

Taken together, numerous interactions are possible within the dimer interface of the mGBP2 homodimer. However, since for hGBP1 it was demonstrated that only two arginine residues in the guanine cap have a major impact on dimerization[33], it can be speculated that the corresponding arginine or lysine residues R238, R242 and K243 play a similarly dominant role in mGBP2 dimerization.

2.2.2 Large-scale motions of the dimeric mGBP2

To analyze the dynamics of the mGBP2 dimer, and whether the motions in the two proteins (M1 and M2) composing the dimer are correlated to each other, we determined the dynamic cross-correlation (DCC) for the motions of the C_α atom. The results in Fig. 6**d** show that correlated and anti-correlated motions occur in a rather symmetric manner between the domains of the two proteins, which derives from the symmetry of the homodimer. The two G domains, the two E domains, as well as the G domain of one protein and the E domain of the other protein move in a correlated fashion. An exception is the loop L1 of both proteins (residues 151–160) whose motions are not, or only very little, correlated with the dynamics in the respective other protein. The motions of the two M domains are inversely correlated to each other. The relationship between the motions of the M domain with those of the G and E domains is mixed. The first half of the M domain up to about residue 400, which includes α7 (residues 310–340), α8 (348–369), and the first half of α9 (377–422), moves in a correlated manner with the G domain and E domain of the other protein in the dimer. The motions of the second part of the M domain, involving the second half of α9, α10 (430–447), and α11 (455–476), are inversely correlated with the motions of the G domain and E domain of the respective other protein. These anti-correlations are only weak to nonexistent for the motions between the M domain of M2 and those of the G and E domain of M1, whereas they are clearly present between the M domain of M1 and the G and E domain of M2. To understand these (anti-)correlations in the dimer dynamics, we elucidated what kind of motions take place. To this end, we monitored the motions of the tips of both M/E domains as they are the most flexible regions in the dimer according to the RMSF plot in Fig. S3b. In Fig. 6**e**, the snapshots corresponding to the largest deviations of the M/E domain relative to its starting structure, i.e., the homology model of the dimer, are shown, revealing a hinge motion similar to that taking place in the mGBP2 monomer. Protein M1 is seen to exhibit somewhat larger hinge motions than M2, with a maximal deviation of residue L480 from it starting structure of 54.5 Å in M1, as compared to 41.3 Å in M2. Fig. 6**e** further suggests that the hinge motions in M1 and M2 are anti-correlated with respect to each other, as a large (small) amplitude in M1 is recorded together with a small (large) amplitude in M2. Next, we computed the Cartesian displacements of the L480 residues to quantify the hinge motion, which can be seen in Fig. 6**f** and are summarized in Supplementary Tab. S2. As indicated by the snapshots in Fig. 6**e** and by the DCC analysis in Fig. 6**d**, the hinge motions in M1 and M2 are anti-correlated to each other. In particular, the hinge motion of M1 takes place in positive y-direction, whereas M1 moves mainly into the opposite y-direction. Interestingly, in the dimer the two M/E domains also exhibit considerable motions along the z-axis, again into opposite directions (negative z-direction in M1, positive z-direction in M2), which did not occur in the monomer (Fig. 5**b**). The same applies to motions along the x-axis, which are however less pronounced than those along y and z. In general, the amplitudes of the motions occurring in M2 are smaller than those in M1, which agrees to the findings from the snapshots in Fig. 6**e**.

To summarize our observations of the dimeric state, the two guanine caps are rather rigid and stabilizing each other. This in turn may help to hold GTP more strongly in its binding site and is thus more efficiently hydrolyzed, as observed in experiments[7]. The hinge motion of the M/E domains is also affected by dimerization, as it moves in somewhat other directions than in the monomer. Moreover, the two hinge motions occurring in M1 and M2 of the dimer are anti-correlated. We elucidated the most relevant interprotein interactions that stabilize the dimer. Via mutations studies, it could be checked if any of these interactions, in particular those involving R238, R242 and K243, are indeed necessary for the dimerization process.

2.3 mGBP2 on the POPC membrane

Finally, we conducted an 1 μs MD simulation of mGBP2 on a simple POPC (1-palmitoyl2-oleoyl-sn-glycero-3-phosphocholine) membrane starting from the structure shown in Fig. 1**c**. We considered mGBP2 without and with GTP being present. For the membrane binding, the geranylgeranyl lipid anchor was always attached to the protein and inserted into the membrane. In order to account for the presence of the geranylgeranyl anchor, but absence of GTP, we denote this system as mGBP2$^{\text{noGTP}}_{\text{mem}}$, while mGBP2$^{\text{GTP}}_{\text{mem}}$ is used for mGBP2 featuring both bound GTP and the geranylgeranyl anchor. We are aware that the mGBP2$^{\text{noGTP}}_{\text{mem}}$ system does not reflect reality as GTP is needed for mGBP2 to bind to a membrane[12]. However, by comparing mGBP2$^{\text{noGTP}}_{\text{mem}}$ with mGBP2$^{\text{GTP}}_{\text{mem}}$ we are able to investigate the influence of GTP on the interplay between mGBP2 and the membrane.

2.3.1 *Dynamics of mGBP2 on the POPC membrane*

As for the protein in solution, we find that the binding of GTP to membrane-bound mGBP2 decreases the flexibility of the G domain motifs G4+L2, SW1 and the guanine cap. This can be seen from the RMSF values in Supplementary Fig. S3**c** and from the reductions in the numbers of clusters when applying conformational clustering to the various motifs and loops (Supplementary Tab. S3). This applies in particular to the SW1, G4+L2 motif, and the guanine cap, which are very flexible in the mGBP2$^{\text{noGTP}}_{\text{mem}}$ system, with 34, 62, and 101 clusters, respectively. These numbers reduce to 4, 11, and 5 clusters upon GTP binding in the mGBP2$^{\text{GTP}}_{\text{mem}}$, which are similar, albeit somewhat higher, as in the monomeric mGBP2$_{\text{holo}}$ in solution. With regard to the movements of the M/E domain we find that it is still flexible, despite the direct neighborhood of the lipid membrane, as can be seen from the RMSF values (Supplementary Fig. S1**g/h** and Supplementary Fig. S3). The hinge motion can still take place on the membrane (Fig. 7a) and is not much affected by it, as assessed by the motions of the tip of the M/E domain. Using residue L480 as measuring point, as before, the maximal deviation from its starting structure is 57 Å for mGBP2$^{\text{noGTP}}_{\text{mem}}$ and 59 Å for mGBP2$^{\text{GTP}}_{\text{mem}}$, and is around 25 Å on average in both systems. In the 1 μs MD simulation of mGBP2$_{\text{holo}}$ in solution, very similar values were reached.

To characterize the motions of the protein relative to the membrane, we calculated the change in the z-position, which corresponds to the bilayer normal, between the center of mass of selected parts of the protein and that of the phosphor atoms of the lipids of the upper bilayer leaflet. Supplementary Fig. S6**a** shows the definition of the coordinate system as well as the parts of the protein used for the Δz calculation. These include three residue areas from the G domain, namely L1 (151–161), G4+L2 (181–196), and the guanine cap (235–256), the whole M domain (303–476) and only the M/E tip (480), as well as α12 and α13 of the E domain, and finally the geranylgeranyl lipid anchor. The Δz values of these regions are plotted in (Supplementary Fig. S6**b**) for mGBP2$^{\text{noGTP}}_{\text{mem}}$ and in (Supplementary Fig. S6**c**) for mGBP2$^{\text{GTP}}_{\text{mem}}$. A statistical analysis is summarized in Supplementary Tab. S5. The first observation is that mGBP2$^{\text{noGTP}}_{\text{mem}}$ and mGBP2$^{\text{GTP}}_{\text{mem}}$ exhibit no big differences in their motions. Secondly, no dissociation of the E domain from the membrane is observed. Thirdly, with regard to the G domain one finds that in both systems, the loop L1 is generally closer to the membrane than G4+L2 and the guanine cap are. However, even L1 does not interact with the membrane, as its distance from the membrane surface is almost always 20 Å or higher. The center of mass of the M domain has also no membrane interactions, while the tip of the M/E domain touches recurrently the membrane. The repeated contact formation with the membrane surface and loss thereof are a result of the hinge motion of the M/E domain. However, the contacts formed between the membrane and both helices of the E domain are less affected by the hinge motion, as they are close to the membrane most of the time (fluctuating around 10–11 Å, Supplementary Tab. S5). Especially α13 interacts intimately with the membrane. The geranylgeranyl lipid anchor lies under the zero line, meaning that it remains fully membrane-inserted throughput the whole simulation. The similarity of the motions of mGBP2$^{\text{noGTP}}_{\text{mem}}$ and mGBP2$^{\text{GTP}}_{\text{mem}}$ relative to the membrane suggests that GTP has no direct effect on the mGBP2-membrane interactions.

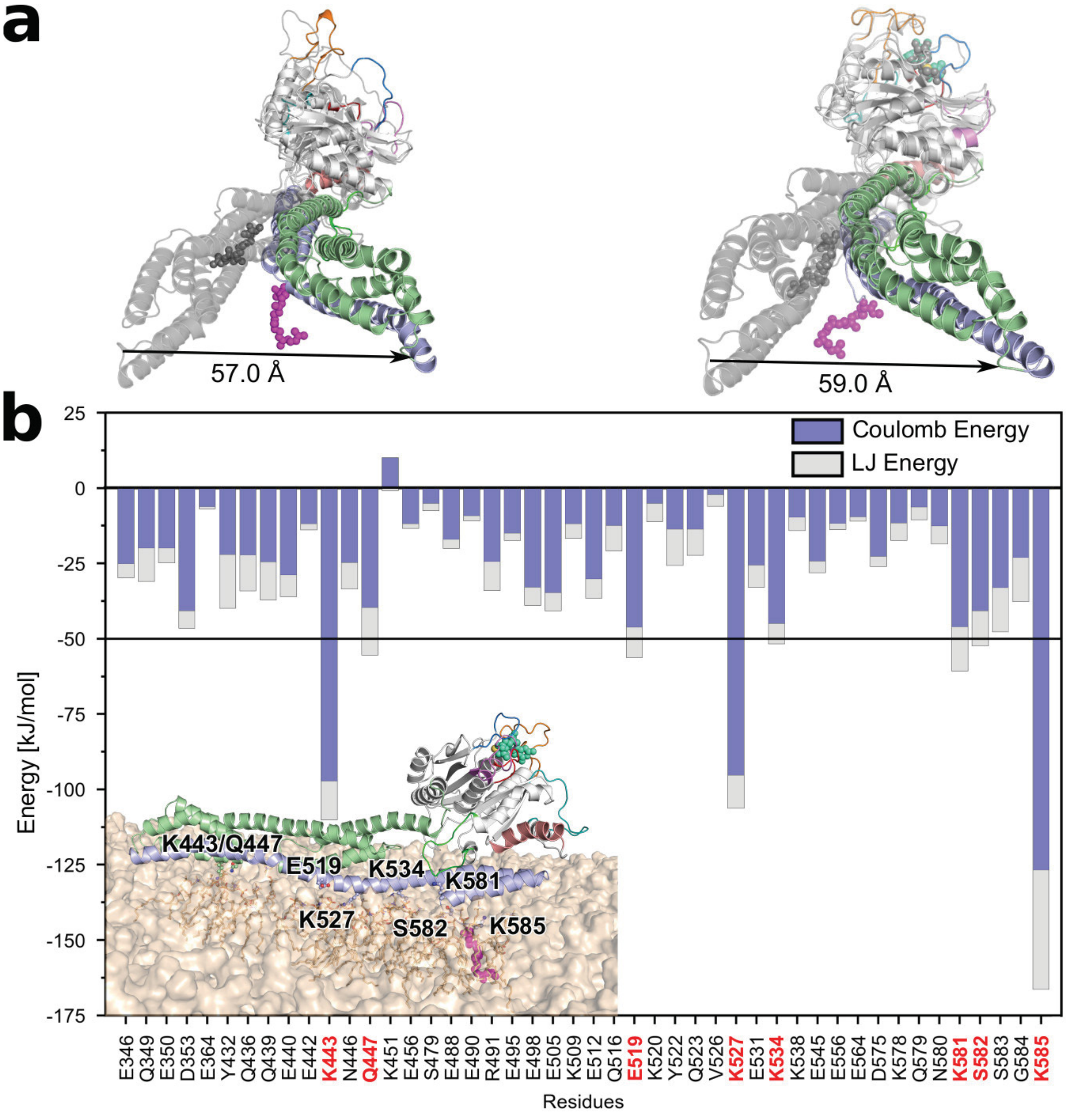


Figure 7. Hinge motion and energies of the strongest interactions between mGBP2 residues and the POPC membrane. **(a)** The conformation with the highest distance to the start structure (L480–L480 for E domain distance, in gray) are shown for mGBP2$_{\text{mem}}^{\text{noGTP}}$ (left) and mGBP2$_{\text{mem}}^{\text{GTP}}$ (right). **(b)** The energies are decomposed into Coulomb (blue) and Lennard-Jones (gray) contributions. The residues with a total interaction below −50 kJ/mol (horizontal black line) are highlighted with red labels and considered as key residues for interacting with the membrane. They are all part of the M/E domain, as can be seen in the structure figure in the lower left corner. The membrane is displayed as a brown surface in that structure figure.

2.3.2 Membrane interactions via energy calculation

In order to elucidate the residues that contribute the most to the interaction with the POPC membrane, we calculated for the mGBP2$_{\text{mem}}^{\text{GTP}}$ system the protein-membrane interaction energy and decomposed it into Lennard-Jones and Coulomb energies, E_{LJ} and E_{Coul}. In Fig. 7**b**, these energies are plotted for all residues that feature a considerable interaction with the lipids. These residues belong all to the M/E domain, while the whole G domain shows no interaction with the membrane, although some of the areas come close. In particular, the lysines K443, K527 and K585 show the highest energies (with values below −90 kJ/mol; negative energies indicating attraction), together with Q447, E519, K581 and S582 (below −50 kJ/mol). These residues, especially the positively charged ones, interact with the head group of the POPC lipids, in particular with their negatively charged PO4$^-$ groups, giving rise to substantial electrostatic energies. The LJ interactions, on the other hand, are generally weaker and only appear together with electrostatic interactions, and not on their own. This indicates that, apart from the geranylgeranyl lipid anchor, none of the other parts of mGBP2 inserted into the membrane but remained on its surface. It was already demonstrated for mGBP2 and hGBP1 that the lipid anchor is necessary for the accumulation of the proteins at the parasitophorous vacuole membrane or giant unilamellar vesicles[7,8]. From the current analysis we conclude that in addition to the lipid anchor, also some of the lysine residues of the M/E domain are strongly involved the mGBP2-membrane interactions.

In summary, mGBP2 is stable on the POPC membrane, with little to no effects caused by the presence of GTP. The protein can still perform its characteristic hinge motion. However, this motion has no noteworthy effect on the membrane properties.

3 Discussion and conclusion

In this study, we provide insights into the conformational changes of mouse GBP2, the homologue of human GBP1. We modeled mGBP2 as a monomer in different states and as a dimer in solution as well as on a simple POPC membrane, by using atomistic molecular dynamics simulations and an enhanced sampling method, Hamiltonian replica exchange molecular dynamics. This is the first report of a GBP being simulated in the GTP-bound state, in addition to considering its isoprenylation.

We find that GTP binding considerably stabilizes the G domain of mGBP2. Even the guanine cap, which is highly flexible without GTP, becomes more ordered and less flexible upon GTP binding. In the biochemical study of Kravets et al.[12], the GTP binding site was characterized in detail with the K51A and R48A mutants affecting phosphate binding, and E99A in the SW2 and D182A in the G4 motif influencing GTP activity and oligomerization. Compared to hGBP1 with its previously reported dissociation constant K_D of 1.1 μM[36], mGBP2 has a 2.4-fold higher affinity to GTP with a K_D of 0.45 μM[12]. This difference might be explained by their slightly different amino acid sequences. Upon GTP binding, the key residues of the mGBP2's G domain interact with the nucleotide, as was previously demonstrated for hGBP1[28,36]. The decreased flexibility of the guanine cap following GTP binding seen here for the mGBP2 can explain why this shifts the monomer–dimer equilibrium to the dimeric state, like in hGBP1[24,33,37], as this should reduce the penalty for dimer formation arising from the loss of conformational entropy of the guanine cap. In the dimeric interaction interface, we identified two residues in the guanine cap, R238/R244, which could have the same function as R240/R244 in the human orthologue[33].

The characteristic hinge motion that we previously identified in hGBP1[31] was seen in all states of mGBP2 that we simulated. Interestingly, this hinge motion may lead to a closed state. Having both a closed and open state is a property that was identified for other dynamin-like proteins, including the human myxovirus resistance protein and also the bacterial dynamin-like protein[38–40]. In solution, the hinge motion of mGBP2 can cover up to 70 Å whereas this decreases to only 30 Å in the dimeric state or on the membrane. As we know, that different mGBPs localize at the parasitophorous vacuole membrane of *T. gondii* as either dimer or higher multimer states, the hypothesis arises that this hinge motion can cause membrane damage, whereupon the membrane might be breaking with sufficient amount of GBPs. This has been shown to be the case for bin-amphiphysin-rvs domain proteins[41,42]. However, with only one mGBP2 molecule being bound to the membrane, we observed no significant changes in the membrane properties, like no high curvature, independent of whether or not GTP was bound to the mGBP2. Therefore, our future studies will focus on simulating mGBP2 assemblies on a membrane. Moreover, as biological membrane has more than one lipid type and considering that the E domain of mGBP2 is highly polar and charged, we can expect that charged lipids may also change the interplay between mGBP2 and membrane. This hypothesis is supported by our observation that several charged residues of the M/E domain strongly interact with the membrane surface. Thus, in the next steps more complex membrane composition will be considered in the simulation setup.

In recent studies, it was demonstrated that hGBP1 can polymerize by detaching the E domain from the other two domains to swing it out, giving rise to disk-like structures, where the G domains are at the outside of the disk and the farnesyl groups assembled together inside the disk[8,43]. A conformation with the E domain in the folded-out position and dimerization via the G domains, as described here, could also tether two giant unilamellar vesicles together via the farnesyl lipid anchor[8,20,35]. Until now, a similar folding out of the E domain has not be shown yet for mGBPs. Nonetheless, such a motion should be considered in future simulations studies of mGBP2 too. It should be noted that the dimer state studied here should have a hard time to insert both geranylgeranyl lipid anchors into the membrane at the same time as they are oriented into different directions. Also this aspect needs attention in the future.

4 Methods

4.1 Molecular dynamic simulations

For the MD and HREMD simulations, the Amber99SB*-ILDNP force field[44–46], TIP3P water[47], and GROMACS 2016[48,49] were applied. All simulations were run at a temperature of 310 K (37°C) and a pressure of 1 bar. In Table 2, all simulations performed in this study are summarized, amounting to a total simulation time of 29.6 μs. Detailed model preparation and simulation descriptions can be found in the **Supplementary Information**.

Table 2. Summary of simulations performed in this work.

Simulation	System	Size in atoms	Replicas/Runs ×Length	Cumulated time
HREMD	$\text{mGBP2}_{\text{apo}}$	131,493	16×200 ns	3.2 μs
HREMD	$\text{mGBP2}_{\text{GTP}}$	129,116	16×200 ns	3.2 μs
HREMD	$\text{mGBP2}_{\text{holo}}$	130,665	16×200 ns	3.2 μs
HREMD	$\text{mGBP2}_{\text{holo}}$	130,665	40×400 ns	16 μs
MD	$\text{mGBP2}_{\text{holo}}$	130,665	1×1 μs	1 μs
MD	$\text{mGBP2}_{\text{holo}}$ dimer	573,673	1×1 μs	1 μs
MD	$\text{mGBP2}_{\text{mem}}^{\text{noGTP}}$	659,946	1×1 μs	1 μs
MD	$\text{mGBP2}_{\text{mem}}^{\text{GTP}}$	662,603	1×1 μs	1 μs
Total simulation time				**29.6 μs**

4.2 Analysis

If not stated otherwise, all analyses were performed using GROMACS 2016. For the analysis of the HREMD simulations, only the data collected by the target replica was used.

Flexibility. To describe the stability and flexibility of mGBP2, the RMSF of the C_α atoms around their average positions was determined for each residue. The RMSF describes the positional change of the selected atoms as time-average. A residue with an RMSF value over 2 Å was considered as flexible. We also projected the RMSF values onto the start structure using a color code, where red colors indicate highly flexible regions (≥ 2 Å) and blue colors are for rigid residues.

Secondary structure analysis. To calculate the average helix content of the E domain (α_{E}) averaged over the whole trajectory, the DSSP (Define Secondary Structure of Proteins) algorithm[50] was employed. It determines the secondary structure for all residues, from which the α-helical content of the E domain was determined.

Clustering analysis. Conformational clustering was performed to identify the most populated conformations of the G domain motifs and loops (see Supplementary Tab. S1) and the geranylgeranyl group. To this end, the algorithm of Daura et al.[51] was applied to the C_α atoms of the structural element in question. This algorithm is based on the root mean square deviation (RMSD) between all structures collected in a trajectory, and an RMSD cutoff value of 2.5 Å was used to define cluster membership for the G domain motifs and loops. Prior the calculation, the trajectories were fitted onto the G domain of the starting structure. A similar clustering was performed for the geranylgeranyl lipid anchor, with the differences that **(i)** the trajectory was fitted using the whole protein and **(ii)** a cutoff value of 2 Å was used.

Hinge motion analysis. The main structural changes and movements of the M/E domain were identified based on a principal component analysis (PCA). For being able to compare the motions of the different mGBP2 systems within the same principle component (PC) space, we applied the PCA to the concatenated target replicas of the HREMD simulations (16 × 200 ns) of mGBP2$_{\mathrm{apo}}$, mGBP2$_{\mathrm{GTP}}$, and mGBP2$_{\mathrm{holo}}$. This step involved the removal of not only the solvent and ions, but also of GTP and Mg^{2+} from both mGBP2$_{\mathrm{GTP}}$ and mGBP2$_{\mathrm{holo}}$ and in the case of the latter, also the geranylgeranyl lipid anchor was removed, as the same number of atoms is needed for a joint PCA. Moreover, the trajectories were aligned to the G domain of the MD starting structure. Of the resulting principal components, the two main motions described by PC1 and PC2 were analyzed in detail. This included to separately project the target replica trajectories of the three HREMD simulations onto the space defined by PC1 and PC2. A PCA was also done for the additional simulations performed for mGBP2$_{\mathrm{apo}}$, i.e., the target replica trajectory of the 40×400 ns HREMD simulation and the 1 μs MD simulation.

Dynamic cross correlation. To characterize the influence of the two proteins composing the dimer on each other's motions, we applied a dynamic cross-correlation (DCC) as available in MD-TASK[52] (`https://md-task.readthedocs.io/en/latest/corr.html`). This allows to determine the dynamic correlation between the atoms of a system, i.e., to what degree the atoms move together and whether their motion is in the same of the opposite direction. This analysis generated an $N \times N$ matrix, where N is the number of the C$_\alpha$ atoms, and each element (i, j) corresponds to the dynamical cross correlation between atoms i and j. This matrix was plotted using a color code, with blue colors representing correlations with values between 0.1 and 1 (complete correlation) and red colors indicating the anti-correlation with valued from -0.1 to -1 (complete anti-correlation). No or only very weak correlation is given, if the value is 0 ± 0.1, as represented by white in the DCC plot.

Protein-protein interactions in the mGBP2 dimer. For the characterization of the mGBP2 dimer interface, all interprotein residue pairs where the two residues were within 12 Å of each other at any time during the MD simulation were identified, resulting in a 149 × 149 matrix. To further separate the strong residue-residue contacts from the weaker ones, the time-averaged contact probability was calculated. To this end, for each of the 149 × 149 residue pairs it was for every 1 ns determined if the pair in question was indeed in contact, using a 5 Å distance cutoff for the minimum distance between the residues to define the presence of a contact.

Analysis of the mGBP2 monomer on the POPC membrane. We monitored the motions of the three mGBP2 domains with respect to the POPC membrane by calculating the distance of the centers of mass of the three domains with respect to the membrane surface along the z coordinate (representing the membrane normal). To define the membrane surface position, the average position of the phosphor atoms of the POPC lipid headgroups were used. To further specifiy which parts of the domains are closest to the membrane, we also calculated the distances of the G domain motifs L1 (151–161), G4+L2 (181–196), and guanine cap (235–256), of the tip of the M/E domain (L480), of α12, α13, and the geranylgeranyl lipid anchor from the membrane surface. This analysis was conducted for every 1 ns. To identify the residues with the strongest interactions with the membrane, we calculated the time-averaged residue-membrane interaction energies E_{LJ} for the Lennard-Jones interactions and E_{Coul} for the electrostatic interactions.

Figures. For the generation of 3D protein structure figures we used PyMol[53].

Acknowledgements (not compulsory)

We thank Dr. Bogdan Barz and Wibke Schumann for fruitful discussions.

Author contributions statement

Conceptualization: B.S.; data curation: J.L., X.W.; formal analysis: J.L.; funding acquisition: J.L., B.S.; investigation: J.L., X.W, B.S; methodology: J.L., B.S; project administration: B.S.; software: J.L.; supervision: B.S.; validation: J.L., B.S.; visualization: J.L.; writing – original draft: J.L.; writing – review & editing: X.W., B.S.

Funding

J.L. and B.S. received funding for this project from the Deutsche Forschungsgemeinschaft (German Research Foundation, https://www.dfg.de/) through Grant 267205415 (CRC 1208, Project A07). We gratefully acknowledge the computing time granted through JARA-HPC (project JICS6A) on the supercomputer JURECA at Forschungszentrum Jülich, through the Leibniz Supercomputing Centre (LRZ) of the Bavarian Academy of Sciences on the supercomputer SuperMUC-NG (project pn98zo), and through the Centre for Information and Media Technology at Heinrich Heine University Düsseldorf.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information .

Correspondence and requests for materials should be addressed to B.S..

References

1. Praefcke, G. & McMahon, H. The dynamin superfamily: universal membrane tubulation and fission molecules? *Nat. Rev. Mol. Cell Biol.* **5**, 133–147 (2004).
2. Vestal, D. J. & Jeyaratnam, J. A. The Guanylate-Binding Proteins: Emerging Insights into the Biochemical Properties and Functions of This Family of Large Interferon-Induced Guanosine Triphosphatase. *J. Interf. Cytokine Res.* **31**, 89–97, DOI: 10.1089/jir.2010.0102 (2011). PMID: 21142871, https://doi.org/10.1089/jir.2010.0102.
3. Faelber, K. *et al.* Chapter Fifteen – Oligomerization of Dynamin Superfamily Proteins in Health and Disease. In Giraldo, J. & Ciruela, F. (eds.) *Oligomerization in Health and Disease*, vol. 117 of *Prog Mol Biol Transl Sci*, 411–443, DOI: https://doi.org/10.1016/B978-0-12-386931-9.00015-5 (Academic Press, 2013).
4. Rafeld, H. L., Kolanus, W., van Driel, I. R. & Hartland, E. L. Interferon-induced GTPases orchestrate host cell-autonomous defence against bacterial pathogens. *Biochem. Soc. Trans.* **49**, 1287–1297, DOI: 10.1042/BST20200900 (2021). https://portlandpress.com/biochemsoctrans/article-pdf/49/3/1287/915998/bst-2020-0900c.pdf.
5. Haller, O., Staeheli, P., Schwemmle, M. & Kochs, G. Mx GTPases: dynamin-like antiviral machines of innate immunity. *Trends Microbiol.* **23**, 154–163, DOI: https://doi.org/10.1016/j.tim.2014.12.003 (2015).
6. Carter, C., Gorbacheva, V. & Vestal, D. Inhibition of VSV and EMCV replication by the interferon-induced GTPase, mGBP-2: differential requirement for wild-type GTP binding domain. *Arch. Virol.* **150**, 1213–1220 (2005).
7. Kravets, E. *et al.* Guanylate binding proteins directly attack *Toxoplasma gondii* via supramolecular complexes. *eLife* **5**, e11479, DOI: 10.7554/eLife.11479 (2016).
8. Shydlovskyi, S. *et al.* Nucleotide-dependent farnesyl switch orchestrates polymerization and membrane binding of human guanylate-binding protein 1. *Proc. Natl. Acad. Sci. U. S. A* **114**, E5559–E5568, DOI: 10.1073/pnas.1620959114 (2017). https://www.pnas.org/content/114/28/E5559.full.pdf.

9. Santos, J. C. *et al.* LPS targets host guanylate-binding proteins to the bacterial outer membrane for non-canonical inflammasome activation. *The EMBO J.* **37**, e98089, DOI: https://doi.org/10.15252/embj.201798089 (2018). https://www.embopress.org/doi/pdf/10.15252/embj.201798089.

10. Praefcke, G. J. Regulation of innate immune functions by guanylate-binding proteins. *Int. J. Med. Microbiol* **308**, 237–245, DOI: https://doi.org/10.1016/j.ijmm.2017.10.013 (2018). Intracellular Compartments as Places of Pathogen-Host Interaction.

11. Haque, M., Siegel, R. J., Fox, D. A. & Ahmed, S. Interferon-stimulated GTPases in autoimmune and inflammatory diseases: promising role for the guanylate-binding protein (GBP) family. *Rheumatology* **60**, 494–506, DOI: 10.1093/rheumatology/keaa609 (2020). https://academic.oup.com/rheumatology/article-pdf/60/2/494/36167979/keaa609.pdf.

12. Kravets, E. *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*. *Int. J. Biol. Chem.* **287**, 27452–27466 (2012).

13. Degrandi, D. *et al.* Murine Guanylate Binding Protein 2 (mGBP2) controls Toxoplasma gondii replication. *Proc. Natl. Acad. Sci. U. S. A* **110**, 294–299, DOI: 10.1073/pnas.1205635110 (2013). http://www.pnas.org/content/110/1/294.full.pdf.

14. Lindenberg, V. *et al.* Broad recruitment of mGBP family members to Chlamydia trachomatis inclusions. *PLOS ONE* **12**, 1–14, DOI: 10.1371/journal.pone.0185273 (2017).

15. Kim, B.-H. *et al.* A Family of IFN-γ–Inducible 65-kD GTPases Protects Against Bacterial Infection. *Science* **332**, 717–721, DOI: 10.1126/science.1201711 (2011). https://science.sciencemag.org/content/332/6030/717.full.pdf.

16. Feeley, E. M. *et al.* Galectin-3 directs antimicrobial guanylate binding proteins to vacuoles furnished with bacterial secretion systems. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E1698–E1706, DOI: 10.1073/pnas.1615771114 (2017). https://www.pnas.org/content/114/9/E1698.full.pdf.

17. Shenoy, A. R. *et al.* GBP5 Promotes NLRP3 Inflammasome Assembly and Immunity in Mammals. *Science* **336**, 481–485, DOI: 10.1126/science.1217141 (2012). https://science.sciencemag.org/content/336/6080/481.full.pdf.

18. Ince, S. *et al.* Catalytic activity of human guanylate-binding protein 1 coupled to the release of structural restraints imposed by the C-terminal domain. *The FEBS J.* **288**, 582–599, DOI: https://doi.org/10.1111/febs.15348 (2021). https://febs.onlinelibrary.wiley.com/doi/pdf/10.1111/febs.15348.

19. Cui, W. *et al.* Structural basis for GTP-induced dimerization and antiviral function of guanylate-binding proteins. *Proc. Natl. Acad. Sci. U. S. A* **118**, DOI: 10.1073/pnas.2022269118 (2021). https://www.pnas.org/content/118/15/e2022269118.full.pdf.

20. Sistemich, L. *et al.* The Molecular Mechanism of Polymer Formation of Farnesylated Human Guanylate-binding Protein 1. *J. Mol. Biol.* **432**, 2164–2185, DOI: https://doi.org/10.1016/j.jmb.2020.02.009 (2020).

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. Interf. Cytokine Res.* **26**, 328–352, DOI: 10.1089/jir.2006.26.328 (2006). PMID: 16689661, https://doi.org/10.1089/jir.2006.26.328.

22. Degrandi, D. *et al.* Extensive Characterization of IFN-Induced GTPases mGBP1 to mGBP10 Involved in Host Defense. *J. Immunol. Res.* **179**, 7729–7740, DOI: 10.4049/jimmunol.179.11.7729 (2007). https://www.jimmunol.org/content/179/11/7729.full.pdf.

23. Kresse, A. *et al.* Analyses of murine GBP homology clusters based on in silico, in vitro and in vivo studies. *BMC Genomics* **9**, 1–12, DOI: https://doi.org/10.1186/1471-2164-9-158 (2008).

24. 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* **403**, 567, DOI: 10.1038/35000617 (2000).

25. Prakash, B., Renault, L., Praefcke, G., Herrmann, C. & Wittinghofer, A. Triphosphate structure of guanylate-binding protein 1 and implications for nucleotide binding and GTPase mechanism. *The EMBO J.* **19**, 4555–4564, DOI: 10.1093/emboj/19.17.4555 (2000). http://emboj.embopress.org/content/19/17/4555.full.pdf.

26. Legewie, L. *et al.* Biochemical and structural characterization of murine GBP7, a guanylate binding protein with an elongated C-terminal tail. *Biochem. J.* **476**, 3161–3182 (2019).

27. 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* **440**, 101–104, DOI: 10.1038/nature04510 (2006).

28. Tripathi, R., Glaves, R. & Marx, D. The GTPase hGBP1 converts GTP to GMP in two steps via proton shuttle mechanisms. *Chem. Sci.* **8**, 371–380, DOI: 10.1039/C6SC02045C (2017).

29. Tripathi, R., Noetzel, J. & Marx, D. Exposing catalytic versatility of GTPases: taking reaction detours in mutants of hGBP1 enzyme without additional energetic cost. *Phys. Chem. Chem. Phys.* **21**, 859–867, DOI: 10.1039/C8CP06343E (2019).

30. Vöpel, T. *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* **53**, 4590–4600, DOI: 10.1021/bi500524u (2014). PMID: 24991938, https://doi.org/10.1021/bi500524u.

31. 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.* **15**, 1–29, DOI: 10.1371/journal.pcbi.1007193 (2019).

32. Vestal, D., Gorbacheva, V. & Sen, G. Different Subcellular Localizations for the Related Interferon-Induced GTPases, MuGBP-1 and MuGBP-2: Implications for Different Functions? *J. Interf. Cytokine Res.* **20**, 991–1000, DOI: 10.1089/10799900050198435 (2000). PMID: 11096456, https://doi.org/10.1089/10799900050198435.

33. 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. *The FEBS J.* **279**, 203–210, DOI: https://doi.org/10.1111/j.1742-4658.2011.08415.x (2012). https://febs.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1742-4658.2011.08415.x.

34. Nantais, D. E., Schwemmle, M., Stickney, J. T., Vestal, D. J. & Buss, J. E. Prenylation of an interferon-γ-induced GTP-binding protein: the human guanylate binding protein, huGBP1. *J. Leukoc. Biol.* **60**, 423–431, DOI: https://doi.org/10.1002/jlb.60.3.423 (1996). https://jlb.onlinelibrary.wiley.com/doi/pdf/10.1002/jlb.60.3.423.

35. Sistemich, L. *et al.* Structural requirements for membrane binding of human guanylate-binding protein 1. *The FEBS J.* **288**, 4098–4114, DOI: https://doi.org/10.1111/febs.15703 (2021). https://febs.onlinelibrary.wiley.com/doi/pdf/10.1111/febs.15703.

36. Praefcke, G. *et al.* Identification of Residues in the Human Guanylate-binding Protein 1 Critical for Nucleotide Binding and Cooperative GTP Hydrolysis. *J. Mol. Biol.* **344**, 257–269, DOI: https://doi.org/10.1016/j.jmb.2004.09.026 (2004).

37. Lorenz, C. *et al.* Farnesylation of human guanylate-binding protein 1 as safety mechanism preventing structural rearrangements and uninduced dimerization. *The FEBS J.* **287**, 496–514, DOI: https://doi.org/10.1111/febs.15015 (2020). https://febs.onlinelibrary.wiley.com/doi/pdf/10.1111/febs.15015.

38. 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* **139**, 1342–1352, DOI: 10.1016/j.cell.2009.11.003 (2009).

39. Low, H. H. & Löwe, J. Dynamin architecture — from monomer to polymer. *Curr. Opin. Struct. Biol.* **20**, 791–798, DOI: 10.1016/j.sbi.2010.09.011 (2010).

40. Chen, Y. *et al.* Conformational dynamics of dynamin-like MxA revealed by single-molecule FRET. *Nat. Commun.* **8**, 15744, DOI: doi:10.1038/ncomms15744 (2017).

41. Daumke, O., Roux, A. & Haucke, V. BAR Domain Scaffolds in Dynamin-Mediated Membrane Fission. *Cell* **156**, 882–892, DOI: https://doi.org/10.1016/j.cell.2014.02.017 (2014).

42. Carman, P. & Dominguez, R. Bar domain proteins—a linkage between cellular membranes, signaling pathways, and the actin cytoskeleton. *Biophys. Rev* **10**, 1587–1604 (2018).

43. Budde, J.-H. *et al.* FRET nanoscopy enables seamless imaging of molecular assemblies with sub-nanometer resolution. *arxiv* (2021). 2108.00024.

44. Lindorff-Larsen, K. *et al.* Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **78**, 1950–1958, DOI: 10.1002/prot.22711 (2010).

45. Best, R. B. & Hummer, G. Optimized Molecular Dynamics Force Fields Applied to the Helix-Coil Transition of Polypeptides. *J. Phys. Chem. B* **113**, 9004–9015, DOI: 10.1021/jp901540t (2009).

46. Aliev, A. E. *et al.* Motional timescale predictions by molecular dynamics simulations: Case study using proline and hydroxyproline sidechain dynamics. *Proteins* **82**, 195–215, DOI: 10.1002/prot.24350 (2014). https://onlinelibrary.wiley.com/doi/pdf/10.1002/prot.24350.

47. 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.* **79**, 926–935 (1983).

48. Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* (2015).

49. Abraham, M. J., van der Spoel, D., Lindahl, E. & Hess, B. the GROMACS development team GROMACS User Manual Version 2016.4.

50. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637 (1983).

51. Daura, X. *et al.* Peptide Folding: When Simulation Meets Experiment. *Angew. Chem. Int. Ed.* **38**, 236–240 (1999).

52. Brown, D. *et al.* MD-TASK: a software suite for analyzing molecular dynamics trajectories. *Bioinformatics* **33**, 2768–2771 (2017).

53. PyMol. The pymol molecular graphics system, version 1.8 (2015). Schrödinger, LLC.

4.3 Manuscript III

Biochemical and structural characterization of murine GBP7, a guanylate binding protein with an elongated C-terminal tail

Larissa Legewie, Jennifer Loschwitz, Nora Steffens, Martin Prescher, Xue Wang, Sander H. J. Smits, Lutz Schmitt, Birgit Strodel, Daniel Degrandi and Klaus Pfeffer

Biochemical Journal (2019). https://doi.org/10.1042/BCJ20190364

X.W. contributed to this paper by building the structure model of mGBP7, running MD simulations of the mGBP7 monomer and the membrane-bound CT tail and analyzing these simulations.

Biochemical Journal (2019) **476** 3161–3182
https://doi.org/10.1042/BCJ20190364

PORTLAND PRESS

Research Article

Biochemical and structural characterization of murine GBP7, a guanylate binding protein with an elongated C-terminal tail

Larissa Legewie[1], **Jennifer Loschwitz**[2,3], **Nora Steffens**[1], **Martin Prescher**[4], **Xue Wang**[2,3], **Sander H. J. Smits**[4,5], **Lutz Schmitt**[4], **Birgit Strodel**[2,3], **Daniel Degrandi**[1] and **Klaus Pfeffer**[1]

[1]Institute of Medical Microbiology and Hospital Hygiene, Heinrich Heine University, Duesseldorf, Germany; [2]Institute of Complex Systems: Structural Biochemistry, Forschungszentrum Juelich, Juelich, Germany; [3]Institute of Theoretical and Computational Chemistry, Heinrich Heine University, Duesseldorf, Germany; [4]Institute of Biochemistry, Heinrich Heine University, Duesseldorf, Germany; [5]Center for Structural Studies, Heinrich Heine University, Duesseldorf, Germany

Correspondence: Daniel Degrandi (daniel.degrandi@hhu.de) or Klaus Pfeffer (klaus.pfeffer@hhu.de)

Guanylate-binding proteins (GBPs) constitute a family of interferon-inducible guanosine triphosphatases (GTPases) that are key players in host defense against intracellular pathogens ranging from protozoa to bacteria and viruses. So far, human GBP1 and GBP5 as well as murine GBP2 (mGBP2) have been biochemically characterized in detail. Here, with murine GBP7 (mGBP7), a GBP family member with an unconventional and elongated C-terminus is analyzed. The present study demonstrates that mGBP7 exhibits a concentration-dependent GTPase activity and an apparent GTP turnover number of 20 min^{-1}. In addition, fluorescence spectroscopy analyses reveal that mGBP7 binds GTP with high affinity ($K_D = 0.22\ \mu M$) and GTPase activity assays indicate that mGBP7 hydrolyzes GTP to GDP and GMP. The mGBP7 GTPase activity is inhibited by incubation with γ-phosphate analogs and a K51A mutation interfering with GTP binding. SEC-MALS analyses give evidence that mGBP7 forms transient dimers and that this oligomerization pattern is not influenced by the presence of nucleotides. Moreover, a structural model for mGBP7 is provided by homology modeling, which shows that the GTPase possesses an elongated C-terminal (CT) tail compared with the Caa*X* motif-containing mGBP2 and human GBP1. Molecular dynamics simulations indicate that this tail has transmembrane characteristics and, interestingly, confocal microscopy analyses reveal that the CT tail is required for recruitment of mGBP7 to the parasitophorous vacuole of *Toxoplasma gondii*.

Introduction

The 65–73 kDa guanylate-binding proteins (p65 GBPs) belong to the interferon (IFN)-induced GTPases of the dynamin superfamily [1]. Members of this superfamily utilize the energy of GTP hydrolysis to execute various structural changes, which are necessary to remodel cellular membranes or to initiate fusion events [2,3]. Several studies on GBPs demonstrated that their accumulation at membranes of intracellular pathogens is essential for host immunity [4–6]. For example, GBPs have been shown to target the parasitophorous vacuole (PV) of the protozoan parasite *Toxoplasma gondii* (*T. gondii*) and the bacteria-containing vacuoles of *Chlamydia trachomatis* (*C. trachomatis*), *Legionella pneumophila*, and *Francisella* spp. [7–12]. However, only scarce data are available about the requirements for GTP binding, GTP hydrolysis, and oligomerization, which are essential to understand the exact mechanisms by which GBPs exert their role in host defense [7,8,13].

In general, the GBP family is highly conserved in vertebrates and exhibits a high degree of homology among each other. The GBP domain architecture consists of two parts: The N-terminal globular GTPase (G) domain and the elongated helical domain that can be subdivided into a middle (M) region, and a C-terminal (CT) GTPase effector (E) region. The G domain harbors the GTP-binding

Received: 15 May 2019
Revised: 8 October 2019
Accepted: 9 October 2019

Accepted Manuscript online:
10 October 2019
Version of Record published:
5 November 2019

site containing four conserved sequence elements: the canonical G1 motif or phosphate-binding loop (P-loop) G*XXXX*GK(S/T), the switch1/G2 T motif, the phosphate- and Mg^{2+}-binding switch2/G3 D*XX*G sequence, and the nucleotide-specificity providing T(V/L)RD G4 motif [14]. The T(V/L)RD motif deviates from the classical N/TK*X*D motif in other GTP-binding proteins [15,16]. The helical M region consists of two helix bundles (α7–11) and extends away from the G domain which gives the molecule an elongated shape [17]. The CT E region consists of two helices (α12/13) which folds back along the M region of the helical domain to the G domain and allows the long α12 helix to interact via eight contacts with the M region and via four contacts with the G domain [17].

To date, seven human GBPs (hGBP1 to hGBP7) and 11 murine GBPs (mGBP1 to mGBP11) have been identified [14,18,19]. Three of these 11 mGBPs, mGBP1, −2, and −5, contain a CT Caa*X* motif, which is posttranslationally modified by a farnesyl or geranylgeranyl isoprene moiety and is required for membrane interaction [7,19–21]. Until now, mainly GBPs bearing such a membrane anchoring motif have been biochemically characterized in detail [7,22,23]. Other mGBPs lacking this particular motif, such as mGBP7, the largest mGBP family member, or its closest human relatives (hGBP4 and hGBP6) have not yet been investigated with regard to their biochemical characteristics and abilities to interact with membranes. mGBP7 was shown to be highly induced in murine cells by IFNγ and, to a lower extent, by IFNβ [18]. Also, it was found to be induced in lung, liver, and spleen of *Listeria monocytogenes* and *T. gondii* infected mice, implying a significant role of mGBP7 in infection [18].

It has been reported that upon binding of GTP GBPs self-assemble into dimers which stimulates their enzymatic activity [24,25]. The dimerization of GBPs has been reported to be established through contact between two G domains [20]. This dimerization induces the positioning of a conserved arginine residue of the P-loop toward the γ-phosphate and stimulates its cleavage by stabilizing the transition state of GTP hydrolysis [26]. A recent quantum mechanics/molecular mechanics study revealed that S73 is also highly important for the hydrolysis reaction, which involves a transfer of a proton from the water nucleophile to a non-bridging phosphoryl oxygen after activation by a composite base consisting of S73, E99, a bridging water molecule, and GTP itself [27]. S73 is introduced into the active site of hGBP1 only upon dimerization [27]. For the unique feature of the majority of GBPs to be able to hydrolyze GTP consecutively to GDP and on to GMP, the β-phosphate in the GDP-bound form has to move to the position where the γ-phosphate is located in the GTP-bound form [7,26,28–30]. The physiological function of this unique feature of GBPs to hydrolyze GDP is not known so far. However, the product ratio varies from 0% GMP (hGBP5 and mGBP5) up to 85–90% GMP (hGBP1) for different GBPs and experimental conditions [1,13,24,31,32]. Besides the nucleotide-dependent self-assembly, further common characteristic properties of GBPs are their high intrinsic GTPase activity with turnover numbers between 2 to 102 min^{-1} and their ability to bind agarose-immobilized GMP, GDP, and GTP with μM affinities [33,34].

Taken together, the biochemistry of GBPs as well as their structure requires further investigation to link their biochemical properties and mode of action to their biological function in host defense. In this respect, data for mGBP7 have been missing up to now. In this study, the GTPase activity of mGBP7 and a hydrolysis deficient mGBP7 mutant (K51A) is characterized. In addition, the oligomerization pattern of mGBP7 and the K51A mutant in the presence of different nucleotides is analyzed and the products of the GTP hydrolysis reaction are determined. Finally, computational modeling is used to predict the structure of mGBP7 and confocal microscopy experiments are performed to elucidate the biological function of the elongated CT tail of mGBP7.

Experimental procedures

Chemicals

All chemicals for protein purification and GTPase activity measurements were ordered from Sigma if not otherwise stated. Components and chemicals used for SDS–PAGE and BN-PAGE were obtained from Thermo Scientific.

Expression constructs

The wild-type (WT) open reading frame (ORF) of mGBP7 (NCBI accession number NM_001083312.2) was cloned into the pQE-80L plasmid (Qiagen). Site-directed mutagenesis (QuikChange II XL Site-Directed Mutagenesis Kit, Agilent Technologies) was carried out for the generation of the GTPase deficient mutant K51A. The K51A mutant was introduced into mGBP7 using forward primer 5′-AGGACTATACCGTACGGGAGCATCCTACTTGATGAACC GC-3′ (Metabion) and reverse primer 5′-GCGGTTCATCAAGTAGGATGCTCCCGTACGGTATAGTCCT-3′ (Metabion). The sequences of both constructs were verified by DNA sequencing (Beckman Coulter Genomics).

The mGBP7 ORF (NCBI accession number NM_001083312.2) without the last 147 base pairs (mGBP7$^{\Delta 49}$) was amplified by PCR using forward primer 5′-ATTCCCGGGAGCATCTGGTCC-3’ (Metabion) and reverse primer 5′-ATTCTTCTCCTTATTTAGTTGAATC-3′ (Metabion). The corresponding clones were subcloned into the pWPXL plasmid (Trono Laboratory [35]) as N-terminal GFP fusion constructs. WT GFP-mGBP7 and GFP-mGBP7$^{\Delta 49}$ sequences were verified by DNA sequencing (BMFZ, Duesseldorf, Germany) and used for lentiviral transduction.

Transformation of *E. coli*

mGBP7 expression constructs were transformed into Rosetta$^{\mathrm{TM}}$ 2(DE3)pLysS competent cells (Novagen) using standard procedures (Novagen). The Rosetta$^{\mathrm{TM}}$ 2 host strain was choosen to enhance the expression of eukaryotic proteins that contain codons rarely used in *E. coli*.

Protein expression and purification

For large scale expression, a volume of six liters of selective 2YT medium (1.6% (w/v) tryptone, 1% (w/v) yeast extract, 0.5% (w/v) NaCl) was inoculated with 150 ml of an *E. coli* Rosetta$^{\mathrm{TM}}$ 2(DE3)pLysS overnight culture. The expression of WT and the K51A mGBP7 mutant with N-terminal His$_6$ tag were induced by adding 150 μM isopropyl-β-D-thiogalactopyranoside (IPTG) at an A_{600} of 0.5. During both steps, the temperature was set at 37°C and the mixer was set at 1800 rpm. After 4 h of incubation, cells were harvested by centrifugation (8000 rpm, 20 min, 4°C) and resuspended in high salt buffer (50 mM Tris–HCl, 5 mM $MgCl_2$, 300 mM NaCl, 10 mM imidazole, 1 mM β-mercaptoethanol, 250 μM Pefabloc (Roche), pH 8). Bacteria were disrupted by multiple passes through a cell disrupter (Constant Systems) at 2.7 kbar. After centrifugation for 1 h at 48 000 rpm, 40 ml of the supernatant was added to 10 ml of equilibrated Ni-NTA resin suspension (Qiagen) and incubated overnight on a rotary shaker (200 rpm) at 4°C. After three washing steps with 45 ml low salt buffer (50 mM Tris–HCl, 5 mM $MgCl_2$, 250 mM NaCl, 10 mM imidazole, 1 mM β-mercaptoethanol, pH 8) mGBP7 and the mGBP7 K51A mutant were eluted with 5 ml of high imidazole buffer (50 mM Tris–HCl, 5 mM $MgCl_2$, 150 mM NaCl, 300 mM imidazole, 1 mM β-mercaptoethanol, pH 8). The 5 min centrifugation steps to sediment the Ni-NTA suspension were performed at 4°C and 1500 rpm. Protein purity was verified by Coomassie blue stained SDS–PAGE and immunoblotting. The 5 ml elution fraction was applied to a Superdex200 26/60 column (GE Healthcare) equilibrated in gel filtration buffer (50 mM Tris–HCl, 5 mM $MgCl_2$, 2 mM dithioerythritol). Fractions containing monomeric mGBP7 were pooled and concentrated using centrifugal ultrafiltration devices with a 50 kDa MWCO (Merck). An amount of 5 ml of concentrated protein was again used for a Size Exclusion Chromatography (SEC) run and eluted protein fractions were analyzed by Coomassie blue stained SDS–PAGE and immunoblotting. mGBP7 fractions were concentrated up to 10–20 mg/ml using 50 kDa MWCO centrifugal ultrafiltration devices, aliquoted, and stored at −80°C.

Immunoblotting and SDS–PAGE analysis

All SDS–PAGE and Western blotting experiments were performed with standard laboratory techniques. For immunoblotting, a polyclonal anti-mGBP7 antibody (Eurogentec), a monoclonal anti-Tetra His antibody (Qiagen) and a monoclonal anti-β-actin antibody (AC-74, Sigma) were used. Non-reducing SDS–PAGE was performed after preincubation with different concentrations (0, 5, 10, 50 mM) of β-mercaptoethanol (Merck).

GTPase activity measurements

The GTPase activity of mGBP7 was examined with the malachite green assay by determining the release of free inorganic phosphate. The assay was performed as described previously with the given modifications [36,37]. Reactions were carried out in a total volume of 100 μl in 50 mM Tris–HCl, pH 7.5 (at 37°C) containing 10 mM $MgCl_2$ and 0–5 mM GMP, GDP or GTP. The reaction was initiated by adding 0.68 μM (total amount of protein: 5 μg) or 0.83 μM (total amount of protein: 6.2 μg) of purified mGBP7 protein at 37°C and stopped after 0 and 20 min by the addition of 25 μl of the reaction mixture into 175 μl of 20 mM ice-cold H_2SO_4. Next, 50 μl of dye solution (0.096% (w/v) malachite green, 1.48% (w/v) ammonium molybdate, and 0.173% (w/v) Tween 20 in 2.36 M H_2SO_4) was added. After 15 min of incubation, the amount of free phosphate was quantified by measuring the absorbance at 595 nm. Obtained data points were fitted using a standard

Biochemical Journal (2019) **476** 3161–3182
https://doi.org/10.1042/BCJ20190364

Michaelis–Menten equation:

$$v = \frac{v_{\max}\ [\mathrm{S}]}{K_{\mathrm{M}} +\ [\mathrm{S}]} \tag{1}$$

Here, v describes the reaction velocity as a function of the substrate concentration [S], $v_{\max}$ is the maximum reaction velocity and the Michaelis constant K_{M} denotes the substrate concentration at which the reaction rate is half of $v_{\max}$. It is important to stress that the reaction mechanism of GBPs includes a dimerization step and therefore also dimerization-dependent catalysis. However, the classic Michaelis–Menten equation, which was used in this study, does not account for dimerization-dependent catalysis. In light of the above-mentioned limitations, all kinetic parameters were reported as apparent values to emphasize the point that these equations do not take dimerization processes into account. In addition, the GTPase activity was determined keeping the GTP concentration constant at 2 mM GTP and varying the amount of added protein from 0 to 1 μM (total amount of protein: 0–7.4 μg). Here, the data were fitted using an allosteric sigmoidal (enzyme activity) or a quadratic (specific activity) equation. In this case, the reaction was initiated by adding 2 mM GTP and all other steps were performed as described previously. Inhibition of the GTPase activity by γ-phosphate analogs was assayed with 1, 10, and 100 mM stock solutions of $\mathrm{BeF_x}$ (Santa Cruz Biotechnology, Inc.) and 1, 10, and 100 mM of $\mathrm{AlCl_3}$ complemented with 5, 50, and 500 mM NaF (1, 10, 100 mM $\mathrm{AlF_x}$), respectively. 1, 10, and 100 mM Orthovanadate stock solutions were prepared from $\mathrm{Na_3VO_4}$ (NEB) at pH 10 and boiled for 2 min prior to use [38,39]. To determine the half-maximal inhibitory concentration ($\mathrm{IC_{50}}$) values, the GTPase activity was plotted against the log of inhibitor concentration. The data were analyzed according to the following four-parameter logistic equation:

$$y = \frac{y_{\max} - y_{\min}}{1 +\ 10^{((\log \mathrm{IC}_{50} - x)\cdot h)}} \tag{2}$$

Here, y represents the GTPase activity value, $y_{\max}$ depicts the GTPase activity of the starting plateau and $y_{\min}$ depicts the GTPase activity of the final inhibited plateau. x represents the logarithmic concentration of the inhibitor. The Hill slope h describes the steepness of the curve. The half-maximal inhibitory concentration ($\mathrm{IC_{50}}$) is calculated as the value of the inhibitor concentration used at an enzyme activity inhibition of 50%. This value matches the inflection point of the resulting curves.

For data evaluation, an additional reaction with EDTA (final concentration of 20 mM) was performed and the autohydrolysis of GTP was subtracted. A $\mathrm{Na_2HPO_4}$ standard curve was used for the determination of free phosphate concentrations. All experiments were conducted three times. Data were fitted using GraphPad Prism version 5.01 for Windows (GraphPad Software, www.graphpad.com).

Fluorimetry

The fluorescent GTP analog mant-GTPγS (2′/3′-O-(*N*-Methyl-anthraniloyl)-guanosine-5′-(γ-thio)-triphosphate (Jena Bioscience) was used to assess the binding affinity of GTP to mGBP7 and to the K51A mGBP7 mutant. The binding of mant-nucleotide to the protein was monitored at 448 nm using a Fluorolog-3 spectrometer (Instruments S.A., HORIBA). The excitation wavelength was set at 355 nm, the slit width at 2 nm, and the temperature was maintained at 25(±1)°C by a circulating water bath. All measurements were done in gel filtration buffer supplemented with 0.5 μM mant-GTPγS. The equilibrium dissociation constant K_{D} was obtained by fitting the data with a quadratic equation as described previously [7,40].

MALDI-TOF mass spectrometry

The GTPase activity measurements were carried out in a total volume of 100 μl in 50 mM Tris–HCl, pH 7.5 (at 37°C) containing 10 mM $\mathrm{MgCl_2}$ and 4 mM GTP. The reaction was started by adding 22.5 μM (total amount of protein: 166.2 μg) of purified mGBP7 protein at 37°C for 30 min. The samples were then spotted with 1 μl of a MALDI matrix solution of α-cyano-4-hydroxy-cinnamic acid (HCCA) in 0.1% trifluoroacetic acid (TFA)/acetonitrile 1 : 10. MALDI-TOF mass spectrometry was performed on a MALDI-TOF/TOF ultrafleXtreme (Bruker Daltonics). The spectra were recorded in a linear mode within a mass range from m/z 200 to m/z 3500. The mass spectra were analyzed using flexAnalysis 3.0 (Bruker Daltonics).

Oligomerization and cooperativity analysis

To determine the oligomerization status of mGBP7, a SEC column (Superdex 200 Increase 10/300 GL, GE Healthcare) was equilibrated with gel filtration buffer using a HPLC system (Agilent 1260 HPLC System) connected to a multi-angle light scattering detector (miniDAWN Treos 2, Wyatt technology) and a differential refractive-index detector (Optilab T-rEX differential refractometer, Wyatt technology). An amount of 100 μl of purified mGBP7 with a concentration of 3.5 mg/ml (47 μM) or 12 mg/ml (162 μM) were centrifuged for 10 min at 4°C, preincubated for 5 min with ±5 mM GTPγS (Jena Bioscience) and loaded onto the equilibrated SEC column at a flow rate of 0.6 ml/min in buffer without added nucleotide. Data collection and processing was performed using the ASTRA7 software (Wyatt Technologies).

BN-PAGE was generally performed as described in the Native PAGE Bis–Tris Gel Manual (Thermo Scientific). Prior to BN-PAGE analysis, 1.4 μM (total amount of protein: 2 μg) of purified WT or K51A mGBP7 mutant protein were incubated in GTPase buffer for 20 min at 37°C (or 4°C, not shown) and supplemented with 2 mM GTPγS or 2 mM GTP in combination with γ-phosphate analogs. Inhibitor concentrations of five times the IC_{50} were used (135 μM Orthovanadate, 1.85 mM AlF_x, and 1.85 mM BeF_x) and the reactions were performed in a total volume of 20 μl. After the incubation, the samples were chilled on ice, supplemented with 5 μl of Native PAGE Sample Buffer and loaded onto a 3–12% Bis–Tris BN-PAGE gel. Native PAGE Running Anode Buffer and Native PAGE Dark Blue Cathode Buffer (containing 0.02% Coomassie G-250) were used. Gels were run at 4°C for 45 min at 150 V and another 45 min at 250 V. Gels were fixed (40% methanol, 10% acetic acid) and de-stained (30% acetic acid) prior to visualization. To estimate molecular masses, denaturated mGBP7 protein was separated electrophoretically. Densitometry analysis was performed using Fiji Software [41] and data were visualized using GraphPad Prism Software version 5.01 for Windows (GraphPad Software, www.graphpad.com). To evaluate the band densities, each lane was measured three times and an appropriate background measurement was substracted. The sum of the protein monomer and dimer densities of each row was set as 100%; the calculated mean values are illustrated in Supplementary Figure S2A,B.

Homology modeling

Homology modeling was performed using the Iterative Threading ASSEmbly Refinement (I-TASSER) web server [42–44] to predict the tertiary structure of mGBP7. Only the sequence of mGBP7 (the target) was required as input, while for all other options the default choices were used. First, the top ten sequence alignments to targets of similar folds from the Protein Data Bank (PDB) using a meta-threading approach were identified. From this, structure models were produced via Monte Carlo simulations to get thousands of structural conformation variants of the target, which were clustered into five representative models ranked by the cluster-size and their quality based on a confidence score (C-score). The C-score ranges between −5 and +2, with a higher C-score indicating a high structural confidence. Here, the models are also ranked by the C-score. The best I-TASSER model was finally aligned to all structures in the PDB library to identify the ten proteins with the closest structural similarity to the mGBP7 model.

MD simulations of mGBP7 in solution

To test the structural stability of the homology models, a 100 ns MD simulation was performed for each of them on the supercomputer JURECA [45]. For the preparation, conduct, and analysis of the MD simulations, the MD package Gromacs 2016 [46,47] was applied, using Amber99SB*-ILDNP as protein force field [48,49] and the TIP3P water model [50]. Each system consisted of mGBP7 in the GTP-bound state (GTP and Mg^{2+} ion), water molecules as solvent and 12 Na^+ ions for neutralization, resulting in a system size of ∼200 000 atoms. After simulation-box setup, each system was first relaxed by minimizing the energy to a maximal force of 1000 $kJ\ mol^{-1}\ nm^{-1}$ using a steepest descent algorithm and then equilibrated to adapt it to the desired temperature of 37°C and pressure of 1 atm for mimicking the physiological environment. First, a 0.1 ns *NVT* equilibration was performed in which the number of atoms (*N*), the box volume (*V*) and temperature (*T*) were kept constant, followed by a 1 ns *NPT* equilibration to adjust the pressure (*P*). During equilibration, mGBP7 and GTP were restrained with a force constant of 1000 $kJ\ mol^{-1}\ nm^{-2}$, allowing the water molecules to relax around the solute. Afterwards, the 100 ns MD production runs in the *NPT* ensemble were performed. In order to decrease the computational cost, a triclinic box measuring 10 × 10 × 19 nm was used and positional restraints were applied on residues of the β-sheets in the G domain so that mGBP7 was not able to rotate within the box. No other restraints were applied during the production runs. The velocity rescaling thermostat [47] was used to regulate the temperature in the *NVT* simulations, while the Nosé–Hoover thermostat [51,52] and the semi-isotropic Parrinello–Rahman barostat [53] were employed for the *NPT* simulations. The electrostatic interactions were calculated with the particle mesh-Ewald method [54,55] in connection with periodic boundary conditions. A cutoff of 12 Å was used for the short-range electrostatic and van der Waals interactions. The LINCS algorithm [56] was applied to constrain all bond lengths, allowing for a time step of 2 fs for the integration of the equations of motion. The coordinates and energies of the systems were recorded every 20 ps.

Bioinformatic analyses

To obtain predictions for the secondary structure and transmembrane (TM) propensity based on the mGBP7 sequence, various bioinformatic tools were applied. To identify possible TM regions, three amino acid scales were calculated using the ProtScale web server [57]: (i) the Kyte & Doolittle hydrophobicity scale using a window size (ws) of 19 [58], (ii) the TM tendency scale with ws = 9 [59], (iii) the percentage of buried residues scale with ws = 9 [60]. For the calculation of these scales, a window size, i.e. the number of residues in the neighborhood of the residue in question that are to be considered needs to be provided. This implies that for the first and last ws/2–1 residues of a given protein sequence these scales are not calculated. The optimal ws depends on the scale to be calculated; the default values as given above were selected in this work. As an additional method to determine TM regions the TMpred server was employed [58]. Here, one has to pre-set the possible TM helix length, which was chosed 14–35 residues. TMpred calculates scores for the probability of inside-to-outside and outside-to-inside TM helices. As for mGBP7, the resulting scores were rather similar, the two scores were averaged for each residue. A score value >0 indicates the presence of a TM helix, and scores >500 are significant. All bioinformatic tools used in this work required the mGBP7 sequence, for which the sequence with Genbank code DAA05846.1 was used.

MD simulations of the membrane-embedded CT tail

The stability of the membrane-inserted CT tail was tested in an MD simulation. To this end, a system was prepared with residues 590–620 being pre-inserted into a lipid bilayer consisting of 68 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) molecules. As starting structure the conformation from homology model 3 (Supplementary Fig. S4) was used, as this is a mainly helical structure in a linear conformation and therefore considered to be a good choice for the initial TM helix model. It was inserted in such a way that the amino acids from residue 598 onwards, i.e. where the predicted TM region starts, where in the hydrophobic core of the membrane, whereas residues 590–597 were placed in the head-group region of the upper bilayer leaflet. Charmm36 was used as force field for both the protein [61] and lipids [62], combined with the TIP3P model for water [50] and standard Charmm36 potentials for the 8 K^+ and 11 Cl^- ions that were added to the system. The resulting system size included 19 770 atoms. For the set-up of the system, its energy minimization and equilibration at a temperature of 310 K and pressure of 1 atm the Charmm-GUI web server [63] was used. The resulting structure, which can be seen in Supplementary Figure S6, was subsequently simulated for 250 ns using Gromacs 2016 [47,48] with the same MD settings used in the simulations of mGBP7 in solution.

Analysis of the MD simulations

To create pictures of the 3D protein structure, PyMol [64] was used. For the calculations, we used Gromacs 2016 [46,47]. To quantify the stability and flexibility of the mGBP7 models during the MD simulations, the root mean square deviations (RMSD) and fluctuations (RMSF) of the C_α atoms of mGBP7 were calculated. The RMSD is the average distance between the atoms of the superimposed homology model and conformations sampled during the MD simulation. For the structural alignment, the last 49 residues were ignored. The RMSF measures the time-averaged fluctuations of the C_α atoms around their average positions. The time-resolved secondary structure of mGBP7 was also determined using the DSSP algorithm (*Define Secondary Structure of Proteins*) [65] to analyze the structural changes of mGBP7 during the MD simulations.

Cell culture and transduction

Primary murine embryonic fibroblasts (MEFs) and human foreskin fibroblasts (HS27, ATCC) were cultured in Dulbecco's modified Eagle's medium (DMEM, Thermo Scientific) supplemented with 10% (v/v) heat-inactivated low endotoxin fetal bovine serum (FBS, PAN-Biotech), and 0.05 mM β-mercaptoethanol (Thermo Scientific). 293FT cells (Thermo Scientific) were cultivated in DMEM supplemented with 10% FBS, 0.05 mM β-mercaptoethanol and 100 μg/ml penicillin/streptomycin (Merck). All recombinant lentiviruses were produced by transient transfection of 293FT cells as described previously [7]. mGBP7$^{-/-}$ MEFs (knockout cell line generated in our laboratory, not shown) were transduced with an appropriate amount of lentivirus and 10 µg/ml polybrene (Merck). After 48 h of incubation, the medium was exchanged by fresh culture medium. Transduced cells were enriched by fluorescence-activated cell sorting (BD FACSAria IIu). Tachyzoites from *T. gondii* strain ME49 (ATCC) were maintained by serial passage using confluent monolayers of HS27 cells as described previously [18].

Infection of MEFs with *T. gondii*

MEFs were stimulated with 200 U/ml IFNγ (R&D Systems) 16 h prior to infection. For immunofluorescence analysis, MEFs were seeded in 24-well plates (Merck) on 13 mm diameter coverslips (VWR International) and inoculated with freshly harvested *T. gondii* at a ratio of 50 to 1. After 2 h of incubation, extracellular parasites were removed by washing with PBS (Thermo Scientific).

Immunofluorescence analysis

Cells were fixed with 4% paraformaldehyde (Santa Cruz Biotechnology, Inc.), permeabilized with 0.02% saponin (Merck) in PBS, blocked with 2% goat serum (Agilent Dako) in 0.002% saponin/PBS, and stained as described previously [18]. The outer membrane protein SAG1 of *T. gondii* tachyzoites was visualized by anti-SAG1 (Abcam) at a concentration of 1 : 700 and subsequent incubation with Alexa Fluor 688-conjugated goat anti-rabbit IgG (Sigma) at 1 : 200. Host cell and pathogen DNA were stained with 1 : 2500 4,6-diamidino-2-phenylindole (DAPI, Thermo Scientific). The coverslips were fixed on microscope slides using Fluoromount-G (Thermo Scientific). Fluorescence was visualized using an LSM780 confocal microscope with Airyscan detection (Zeiss). Image analysis and processing was performed using ZEN 2012 (blue edition, Zeiss), AutoQuant X3 (MediaCybernetics) and Imaris (Bitplane) software. Quantification data were analyzed by the Student t-test using GraphPad Prism 5.01 (GraphPad Software, www.graphpad.com).

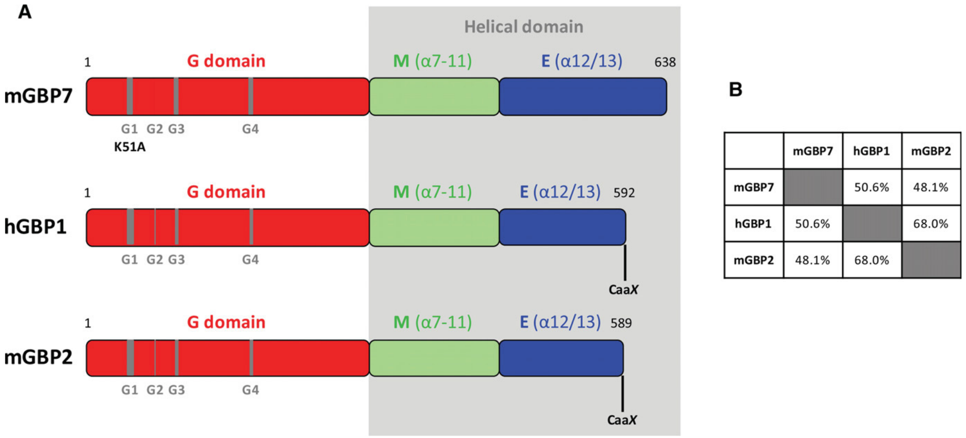
Results

Expression and purification of mGBP7

Previously, we described the biochemical properties of mGBP2 and the GTP hydrolysis deficient mGBP2 mutant K51A [7]. In this study, mGBP7 wild-type (WT) protein and a mutated protein (K51A) with an exchange of the lysine residue in the P-loop at position 51 to alanine were analyzed (Figure 1). Both mGBP7 proteins were expressed in *E. coli* Rosetta^{TM} 2(DE3)pLysS. The highest protein purity was achieved using Ni-NTA batch purification in combination with two consecutive SEC steps (Figure 2A). Purified WT mGBP7 and the K51A mutant were analyzed using Coomassie Brilliant Blue stained SDS gels showing a high purity. WT and K51A mGBP7 were identified by immunoblotting using a polyclonal mGBP7- or a Tetra His-antibody (Figure 2B,C). Detection of WT mGBP7 with the mGBP7-antibody displayed a slight amount of protein degradation.

GTPase activity of wild-type mGBP7 and the K51A mGBP7 mutant

The GTPase activities of WT mGBP7 and the K51A mutant were analyzed using the malachite green phosphate assay by determining the release of free inorganic phosphate. Here, not only GTP but also GDP and GMP were tested as substrates considering that GBPs have the unique feature to hydrolyze GTP to both GDP and GMP [7,28,29]. Offering GMP to WT mGBP7 and the K51A mutant as a substrate did not result in any measurable phosphate release (Figure 3A and Table 1). Also in the presence of GDP, only a background hydrolysis rate for the purified WT mGBP7 and no phosphate hydrolysis for the K51A mutant were detected (Figure 3B and Table 1). In contrast, upon offering GTP a comparable strong GTPase activity with an apparent maximal reaction velocity ($v_{\text{max, app}}$) of 265.2 ± 4.7 nmol min^{-1} per mg of protein and an apparent turnover number ($k_{\text{cat, app}}$) of 19 min^{-1} for WT mGBP7 were determined. In comparison, the K51A mutant only had a neglectable $v_{\text{max, app}}$ of 5.2 ± 4.5 nmol min^{-1} per mg (Figure 3C and Table 1) with a $k_{\text{cat, app}}$ of 1 min^{-1} (Figure 3B and Table 1). The obtained hyperbolic curves for the specific activity of WT mGBP7 indicate that there is no cooperativity in relation to GTP binding (Figure 3C). To corroborate the calculated values for v_{max} and k_{cat}, the GTPase assay was repeated this time varying the amount of added mGBP7 protein and keeping the nucleotide concentration constant at 2 mM GTP (Figure 3D) [66]. Here, plotting the enzyme activity against the mGBP7 concentration a sigmoidal binding curve was obtained (Figure 3D). This is in agreement with a dimerization-dependent mechanism, in which GTP hydrolysis is promoted by dimerization of the



	mGBP7	hGBP1	mGBP2
mGBP7		50.6%	48.1%
hGBP1	50.6%		68.0%
mGBP2	48.1%	68.0%	

Figure 1. Domain structure comparison of mGBP7, hGBP1, and mGBP2.

(**A**) All three GBP members contain an N-terminal globular GTPase (G) domain and an extended CT helical domain that can be subdivided in a middle (M) region (α7–11), and a GTPase effector (E) region (α12/13). The G domains encompass the nucleotide-binding motifs G1–G4 (gray, for explanation see introduction). The investigated K51A mutation in the G1 motif of mGBP7 is depicted. Domain structure comparisons of the three GBPs reveal that the mGBP7 E region is ~50 amino acids longer than the E region of hGBP1 and mGBP2. Moreover, mGBP7 does not contain a CT Caa*X* motif for isoprenoid modification. (**B**) Percentage amino acid identity of mGBP7, hGBP1, and mGBP2 (ClustalW alignment).

enzyme. In line with this finding, the protein concentration was plotted against the apparent rates yielding an apparent dimerization constant of 0.8 μM for mGBP7. Moreover, the result in Figure 3D indicated that at the used protein concentration of 0.68 μM only ~80% of the protein is active. Therefore, the kinetic parameters of WT mGBP7 were again computed for a concentration of 0.83 μM at which a higher dimer fraction is expected. Here, a higher apparent $v_{\mathrm{max,\ app}}$ of 278.9 ± 9.8 nmol min^{-1} mg^{-1} and a marginal higher apparent $k_{\mathrm{cat,\ app}}$ of 20 min^{-1} were calculated. This finding demonstrates that the hydrolysis of GTP is concentration-dependent.

In summary, GTPase measurements demonstrate that mGBP7 possesses a high intrinsic GTPase activity and that the K51 is essential for this activity. In addition, the results illustrate that predominantly GTP and not GDP or GMP serve as a substrate for the mGBP7 induced hydrolysis reaction.

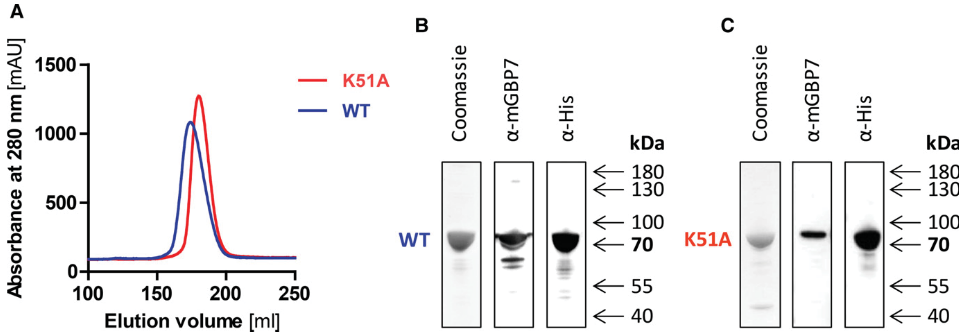


Figure 2. Purification of wild-type mGBP7 and the mGBP7 K51A mutant.

(**A**) SEC profiles of purified WT mGBP7 and the mGBP7 K51A mutant. The mGBP7 proteins containing an N-terminal His_6 tag were expressed in Rosetta^{TM} 2(DE3)pLysS and purified as described under 'Experimental Procedures'. Five micrograms of purified WT mGBP7 (**B**) and the mGBP7 K51A mutant (**C**) were resolved on a 4–12% SDS–PAGE and either stained with Coomassie Brilliant Blue (left panel) or detected by Immunoblotting using the polyclonal anti-mGBP7 antibody (middle panel) and the monoclonal anti-Tetra His antibody (right panel). Due to the His_6 tag, mGBP7 has a theoretical mass of 73.9 kDa. mAU, milli-absorbance units.

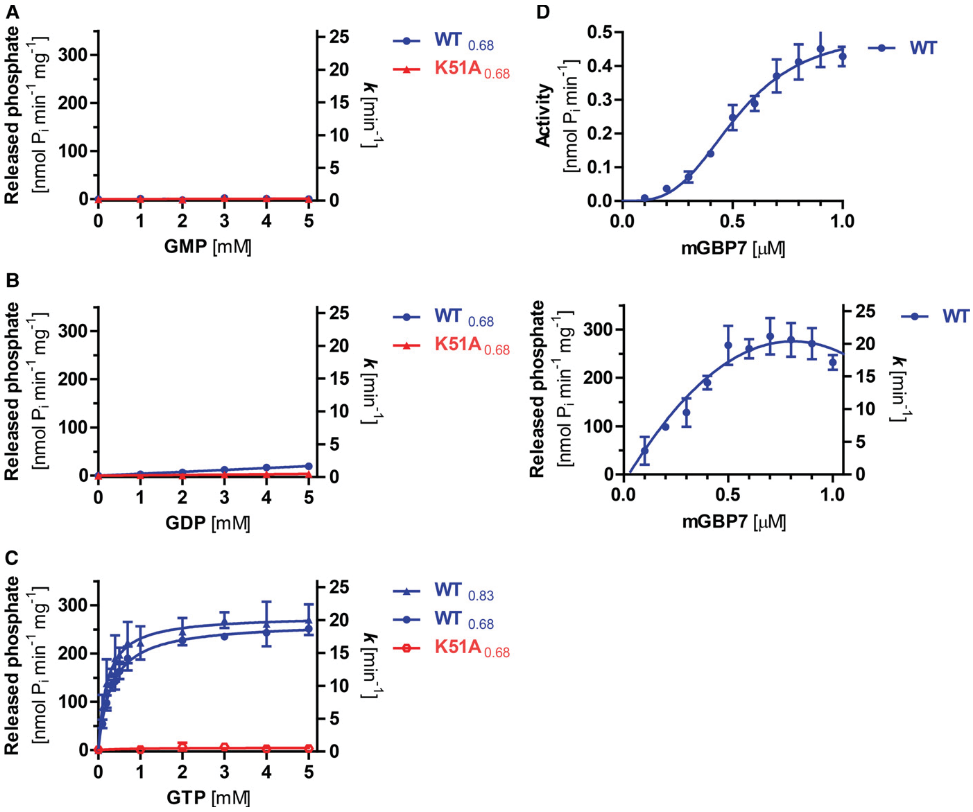


Figure 3. GTPase activity of purified mGBP7 and the K51A mutant.

GTPase activity of WT mGBP7 (blue) and the mGBP7 K51A mutant (red) depending on the GMP (**A**), GDP (**B**), or GTP (**C**) concentration. The GTPase activity was measured in the presence of 0.68 µM (**A**–**C**, WT$_{0.68}$ and K51A$_{0.68}$) or 0.83 µM (**C**, WT$_{0.83}$) protein. In addition, the GTPase activity was measured keeping the GTP concentration constant at 2 mM and varying the mGBP7 concentration (**D**). Datasets **A**, **B**, and **C** were fitted according to the Michaelis–Menten equation (Eq. 1). Dataset D was fitted using an allosteric sigmodal (enzyme activity) or a quadratic (*k*, specific activity) equation, respectively. Results are means ± S.D. from three independent experiments.

Table 1. Kinetic parameters of mGBP7 GTPase activity in the presence of different nucleotides

	GTP			GDP			GMP		
	$v_{\max,\ app}$ nmol min^{-1} mg^{-1}	$K_{m,\ app}$ µM GTP	$k_{cat,\ app}$ min^{-1}	$v_{\max,\ app}$ nmol min^{-1} mg^{-1}	$K_{m,\ app}$ µM GDP	$k_{cat,\ app}$ min^{-1}	$v_{\max}$ nmol min^{-1} mg^{-1}	$K_{m,\ app}$ µM GMP	$k_{cat,\ app}$ min^{-1}
WT$_{0.83}$	278.9 ± 9.8[a]	207.0 ± 32.3[a]	20[a]	-	-	-	-	-	-
WT$_{0.68}$	265.2 ± 4.7[b]	323.4 ± 21.7[b]	19[b]	ND	ND	ND	ND	ND	ND
WT	-	-	20[c]	-	-	-	-	-	-
K51A	5.2 ± 4.5[b]	ND	1[b]	ND	ND	ND	ND	ND	ND

Results are means ± S.D. from three independent experiments. $v_{\max,\ app}$, apparent maximal reaction velocity; $K_{m,\ app}$, apparent Michaelis–Menten constant; $k_{cat,\ app}$, apparent turnover number; ND, not detected.

[a] The GTPase activity was measured in the presence of 0.83 µM (total amount of protein: 6.2 µg) mGBP7 protein (WT$_{0.83}$, Figure 3C);

[b] The GTPase activity was measured in the presence of 0.68 µM (total amount of protein: 5 µg) mGBP7 protein (WT$_{0.68}$, Figure 3A–C);

[c] The GTPase activity was measured in the presence of 2 mM GTP (WT, Figure 3D).

GTP-binding affinity of wild-type mGBP7 and the K51A mGBP7 mutant

To compare the previously studied mGBP2 to mGBP7 analyzed in this study, the GTP-binding affinities of WT mGBP7 and the K51A mutant were examined using fluorescence spectroscopy. The non-hydrolyzable GTP analog GTPγS labeled with a fluorescent mant group was used as substrate. Interestingly, for WT mGBP7 a K_D of 0.22 μM was determined (Figure 4). Using the same approach, it was previously demonstrated that WT mGBP2 possesses a high affinity to mant-GTPγS with a K_D of 0.45 μM [7]. Thus, it can be concluded that WT mGBP7 has a twofold higher GTP-binding affinity than mGBP2 and that under physiological intracellular conditions mGBP7 is rather expected to exist in the GTP-bound form than nucleotide-free [7]. Moreover, whereas for mGBP2 the K51A mutation was described to decrease GTPγS binding affinity drastically (K_D of 44.1 μM), for the K51A mGBP7 mutant virtually no binding in the presence of mant-GTPγS was detected (Figure 4) [7].

mGBP7 hydrolyzes GTP to GDP and GMP

Previously, we reported for mGBP2 that similar to hGBP1, GMP is the main end product of GTP hydrolysis. Here, the products of mGBP7 specific GTP hydrolysis were determined by Matrix-Assisted Laser Desorption Ionization Time-Of-Flight Mass Spectrometry (MALDI-TOF-MS) which revealed that also for mGBP7, not only GDP but also GMP represents a main product of the GTP-hydrolysis reaction (Figure 5C). In light of the observation that mGBP7 possesses a high affinity for mant-GTPγS, a control measurement was included to ensure that the purified protein is indeed nucleotide free (Figure 5A).

mGBP7 forms a transient dimer

The GBP proteins studied to date dimerize upon binding of GTP [7,67]. In previous gel filtration studies, we had demonstrated that mGBP2 elutes as a dimer in the presence of GTPγS [7]. Moreover, we had observed that mGBP2 assembles into tetramers in the presence of GDP and AlF_x [7]. To address whether mGBP7 is also able to form dimers or even tetramers, the oligomeric state of WT mGBP7 was analyzed via SEC coupled to Multi-Angle Light Scattering (MALS) which allows the determination of the molecular mass of the eluting species independently of their hydrodynamic radius [68]. Using this approach, at 3.5 mg/ml protein, an average molecular mass of 74 kDa in the absence of nucleotides and an average molecular mass of 83 kDa after preincubation with 5 mM GTPγS were determined (Figure 6A,B). These molar masses are consistent with the monomeric form of the 73.9 kDa His_6-mGBP7. At 12 mg/ml protein two molecular species of 93.6 and 142.2 kDa in the absence of nucleotides and two different mGBP7 species of 89.8 and 135.1 kDa in the presence of GTPγS were determined (Figure 6B,C). However, it has to be noted that the two peaks are eluting too close for a clear separation. Thus, the higher oligomer contains contributions of the lower oligomer and vice versa. This shifts the calculated mass of the dimer toward a lower apparent molecular mass and the monomer toward a higher molecular mass, whereas the shift in the calculated molecular mass of smaller molecules is higher since bigger particles scatter more light. Consequently, when only considering the upper threshold of 148 and 145 kDa for the dimer and the lower threshold of 88 and 83 kDa for the monomer, the calculated values fit quite well to

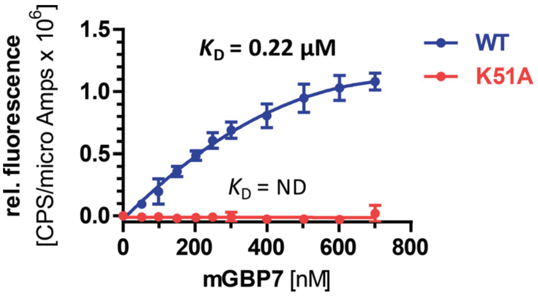


Figure 4. Equilibrium binding of mant-GTPγS to mGBP7.
WT or K51A mGBP7 was added stepwise to gel filtration buffer containing 0.5 μM mant-GTPγS. The fluorescence was excited at 355 nm and measured at 448 nm. The fluorescence data were analyzed as described under 'Experimental Procedures'. Results are means ± S.D. from at least three independent experiments. K_D, dissociation constant; CPS, counts per second.

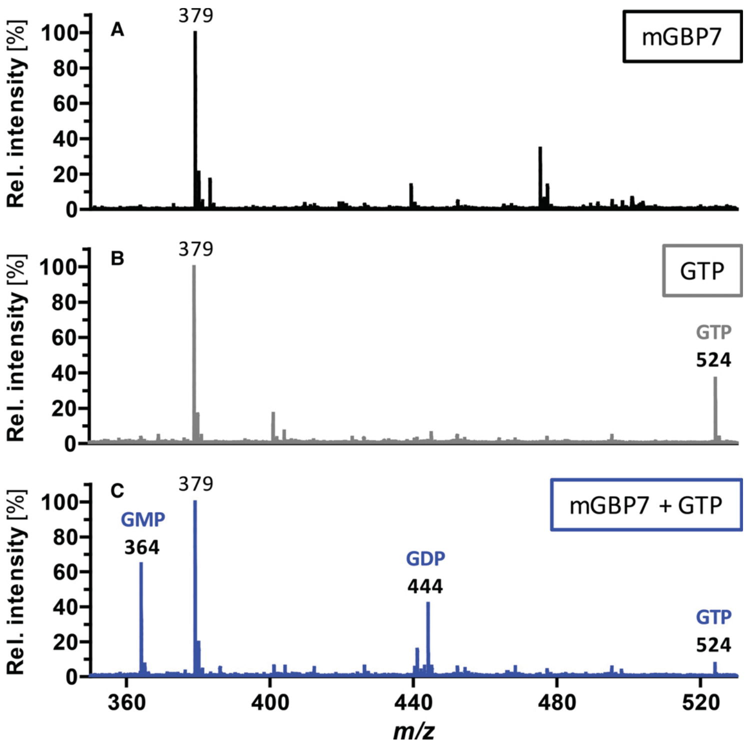


Figure 5. Identification of mGBP7 hydrolysis reaction products.

GTPase activity measurements were analyzed by MALDI-TOF-MS. The mass spectrum of a control sample containing mGBP7 but no additional GTP (**A**), a control sample containing GTP but no protein (**B**) and the mass spectrum of a sample containing mGBP7 after the addition of GTP (**C**) are shown ($n = 3$). Peaks corresponding to mono protonated ions $[M + H]^+$ of GMP, GDP, and GTP are indicated. *m/z*, mass-to-charge ratio.

the expected molecular masses of an mGBP7 dimer and monomer, respectively. These data indicate that the mGBP7 dimer formation is rather dependent on protein concentration than affected by the presence of nucleotides. Nevertheless, it has to be considered that the GTPγS nucleotide and the mGBP7 protein possibly get separated in the course of the SEC run and that this is the explanation for the missing shift to more dimer formation. To exclude the possibility that mGBP7 forms a non-physiological disulfide bridge at high protein concentration, SDS–PAGE analyses of the analyzed protein fractions with and without reducing agents were performed (Supplementary Fig. S3). With and without the reducing agent β-mercaptoethanol, only monomeric mGBP7 was detected indicating that the mGBP7 dimer is not formed due to unspecific disulfide bridge formation.

Thus, it can be postulated that mGBP7 exists in a monomeric and dimeric form and that the addition of GTPγS has no detectable effect on the mGBP7 monomer-to-dimer ratio.

Inhibition of mGBP7 GTPase activity by GTP-transition and ground state analogs

Next, the inhibition of mGBP7-specific GTP hydrolysis by phosphate GTP transition and ground state analogs was analyzed (Figure 7). To mimic the terminal phosphate group (γ-phosphate) in the transition state, orthovanadate or aluminum fluoride (AlF_x) were added to the reaction mixture. To mimic the γ-phosphate in its ground state, beryllium fluoride (BeF_x) was added to the sample. The GTPase activity is inhibited by binding of orthovanadate, AlF_x, or BeF_x in the position of the γ-phosphate after the first hydrolysis step of the added GTP. To assure that GDP is not the limiting compound for the formation of these GDP-conjugated inhibitor complexes, the measurements were also performed at 4 mM GDP and 2 mM GTP (Supplementary Fig. S1). In

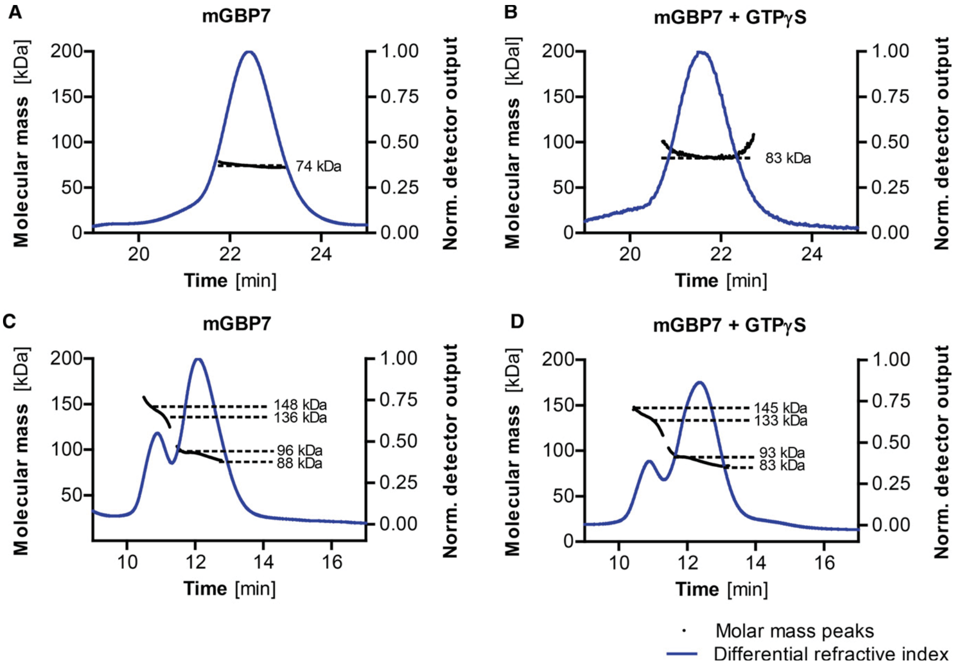


Figure 6. Concentration-dependent dimerization of mGBP7 assessed by SEC-MALS.

Chromatograms of WT mGBP7 showing the differential refractive index (blue) and the calculated molar mass peaks (black). The MALS analyses were performed in the absence of nucleotides (**A** and **C**) or after preincubation with 5 mM GTPγS (**B** and **D**). At protein concentrations of 3.5 mg/ml (**A** and **B**), determined average molar masses were 74 ± 2 kDa (**A**) and 83 ± 1 kDa (**B**), respectively. At 12 mg/ml, SEC-MALS elution profiles showed two distinct molecular mass species. Molar mass species in the absence of nucleotides: 142.2 ± 6.9 kDa, 93.6 ± 3.0 kDa. Molar mass species in the presence of GTPγS: 89.8 ± 3.0 kDa, 135.1 ± 6.0 kDa. Including the His_6 tag, the mGBP7 protein has a theoretical mass of 73.9 kDa. SEC, size-exclusion chromatography; MALS, multi-angle light scattering.

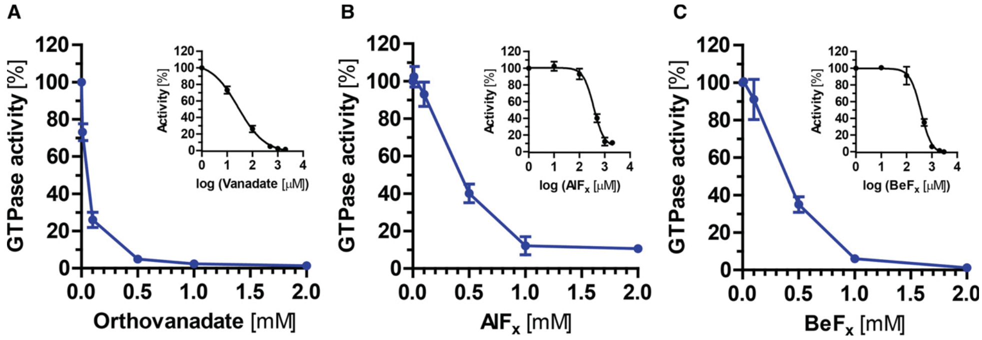


Figure 7. Inhibition of mGBP7 by γ-phosphate analogs.

GTPase activity of WT mGBP7 in the presence of 0.01, 0.1, 0.5, 1.0, and 2.0 mM orthovanadate (**A**), AlF_x (**B**), or BeF_x (**C**). The GTPase activity was analyzed in the presence of 2 mM GTP and 0.68 μM (5 μg) mGBP7 protein. The data were fitted according to equation 2 (Eq. 2) and the corresponding IC_{50} curves are shown in the insets. Results are means ± S.D. from three independent experiments. IC_{50}, half-maximal inhibitory concentration.

Table 2. *IC*$_{50}$ values of mGBP7 GTPase activity inhibited by γ-phosphate analogs

γ-phosphate analog	
Orthovanadate	31 ± 2[a] 31 ± 6[b]
AlF$_x$	405 ± 61[a] 390 ± 48[b]
BeF$_x$	355 ± 53[a] 279 ± 71[b]

Results are means ± S.D. from three independent experiments. *IC*$_{50}$, half-maximal inhibitory concentration.

[a] The GTPase activity was measured in the presence of 2 mM GTP and 0.68 μM (total amount of protein: 5 μg) mGBP7 protein;

[b] The GTPase activity was measured in the presence of 2 mM GTP, 4 mM GDP, and 2.5 μM (total amount of protein: 18.5 μg) mGBP7 protein.

either case, similar concentrations with up to 2 mM of γ-phosphate analog were required to get a strong (~90%) or nearly complete inhibition of mGBP7 initiated GTP hydrolysis.

Furthermore, the half-maximal inhibitory concentration (IC$_{50}$) of all three inhibitors was determined (Table 2). For orthovanadate, an IC$_{50}$ of 31 μM was calculated, representing the lowest IC$_{50}$ of the examined inhibitors. The IC$_{50}$ for AlF$_x$ was computed to be 405/390 μM, and thus ~13-fold higher than the IC$_{50}$ for orthovanadate. For BeF$_x$, the IC$_{50}$ was more about in the same range of AlF$_x$, and with 355/279 μM ~10-fold higher than the IC$_{50}$ of orthovanadate.

Subsequently, it was investigated whether the phosphate analogs could stabilize mGBP7 in its dimeric state. Therefore, the GTPase assay was performed in the presence of GTPγS, orthovanadate, AlF$_x$, or BeF$_x$ and the samples were analyzed by Blue Native (BN) PAGE (Supplementary Fig. S2). In all cases, WT proteins were found as both monomers and dimers (Supplementary Fig. S2A). Also, in the absence of any nucleotides, dimers constituted up to ~8% of the WT mGBP7 proteins. In the presence of the non-hydrolyzable GTP analogs only a very slight increase to ~12% dimers was observed. Subsequently, the K51A mutant was analyzed via BN-PAGE (Supplementary Fig. S2B), and similar monomer-to-dimer ratios were observed.

In summary, these results support the MALS data which demonstrate that the dimerization pattern of mGBP7 is not influenced by the presence of nucleotides.

Homology modeling of mGBP7

To obtain a structural model of mGBP7, homology modeling using I-TASSER was employed [42–44]. The top ten sequence alignments were found for hGBP1 (PDB) codes 1F5N [69] and 1DG3 [17]) as templates with ~55% sequence identity between mGBP7 and hGBP1. However, in contrast with hGBP1, mGBP7 exhibits no lipid anchor but has 49 additional residues at the CT end (residues 590–638), which will be denoted as CT tail. Five structure models with C-scores between −1.76 and −0.18 were generated (Supplementary Fig. S4A), where a larger C-score indicates higher model reliability and a C-score of +2 is the maximum that can be obtained with I-TASSER. The G domain of mGBP7 is very similar to that of hGBP1 as they share a sequence identity of 64%, which was calculated with BLAST [17,70]. As a result, the G domains of the five mGBP7 models differ by <0.4 Å in terms of their mutual root mean square deviation (RMSD). Moreover, the GTP-binding sites of mGBP7 and hGBP1 involving loops G1–G4 have similar amino acid compositions (Figure 8A). It should be noted that N-terminally of the G4 motif in mGBP7 there is a gap compared with hGBP1, shifting D182 of mGBP7 to the position D184 of hGBP1 in the sequence alignment. The structural superposition of the first I-TASSER model of mGBP7 and the hGBP1 crystal structure (PDB code 1F5N [69]) shows that the orientation of the side chains of the key residues R48 and K51 in the P-loop (G1), E99 in switch2, and D182/184 in the G4 motif are slightly different in the two proteins, but overall the structures of the two G domains are very

Biochemical Journal (2019) **476** 3161–3182
https://doi.org/10.1042/BCJ20190364

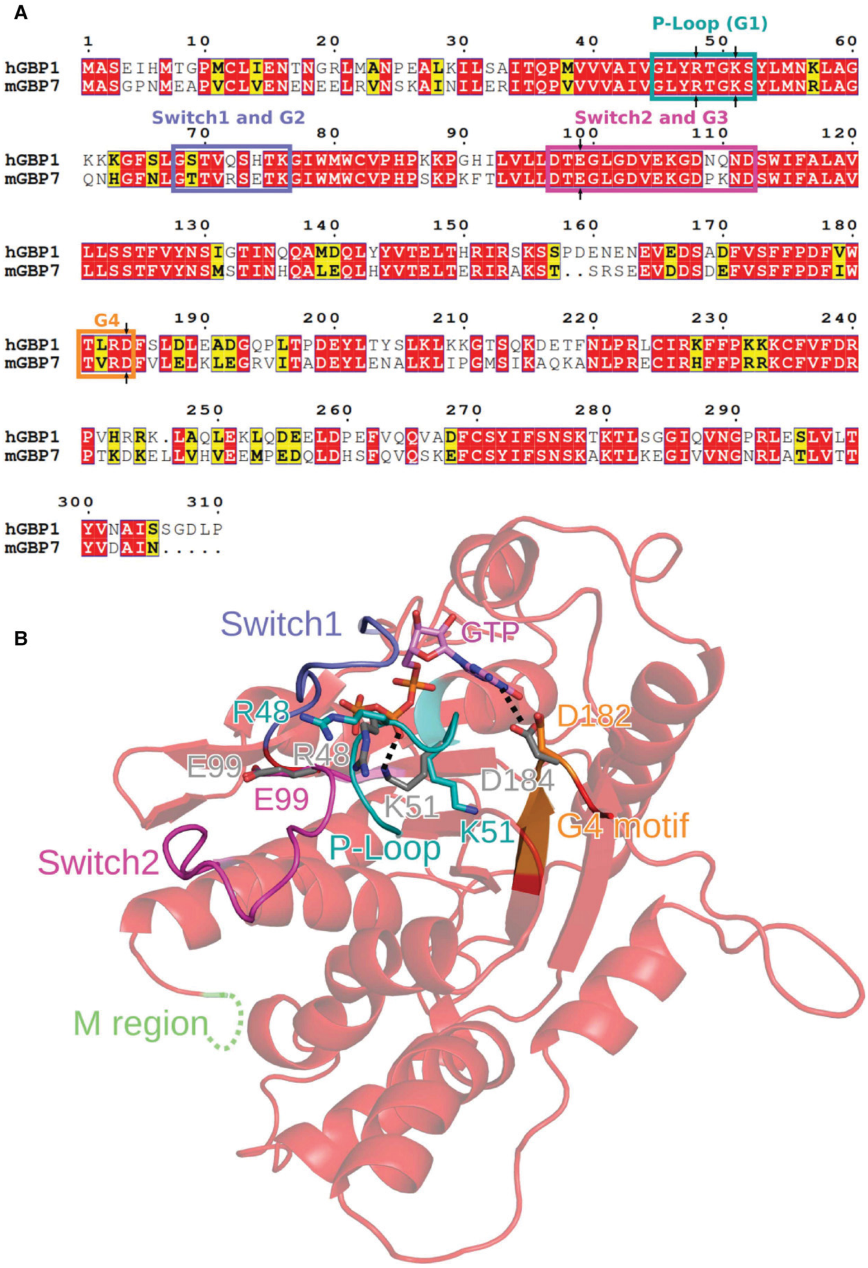


Figure 8. Amino acid sequence and structural alignments of mGBP7 and hGBP1.

(**A**) The sequence alignment between both proteins, calculated by Blast [17,70] and created with ESPript 3.0 [80], is shown where residues highlighted in red are conserved and residues marked in yellow exhibit similar chemical properties. Furthermore, the four conserved GTP-binding site motifs are displayed: P-Loop (G1) in turquoise, switch1 (G2: T75) in blue, switch2 (G3: DTEG) in magenta, and G4 motif in light orange. The black arrows indicate the key residues for GTP binding and hydrolysis. (**B**) The G domain of the mGBP7 model 1 (Supplementary Figure S4B) is depicted as cartoon. The four GTP-binding loops (cartoon representation) and the key residues (sticks) are highlighted in the corresponding colors as in **A**. For the comparison of hGBP1 with mGBP7, hGBP1 key residues are shown in gray. The black dashed lines indicate the interactions with GTP (purple sticks).

similar to each other (Figure 8B). For the homology modeling of mGBP7, no GTP was used, therefore, the orientation of the key residues in the GTP-binding site are different between mGBP7 and GNP-bound hGBP1 but they reoriented to the GTP during the MD simulations. The uncertainty of the five I-TASSER models of mGBP7, as evident by the C-scores below zero, stems from the structure prediction for the CT tail, which adopts different helical and/or random coil structures in the five models (Supplementary Fig. S4). This was confirmed by homology modeling of mGBP7 without the additional 49 CT residues, which resulted in only one model with a high C-score of 1.91, indicating a high reliability of the mGBP7 models for residues 1–589. The best of the five full-length mGBP7 models is shown in Figure 9A. To further elucidate the structural preferences and flexibility of the CT tail, a 100 ns all-atom MD simulation, considering GTP and Mg^{2+} in the active site was performed for each of the five models. These simulations revealed that the CT tail is one of the most flexible parts of mGBP7, which can be deduced from the root mean square fluctuations (RMSF) of the C_α atoms (Figure 9B and Supplementary Figure S4C). Another mobile part of mGBP7 is the region where the M and the E regions are connected, resulting from the overall mobility of these two regions with respect to the G domain. Despite this flexibility, the helices in the M and E regions are stable during the MD simulations, which is not the case for the CT tail. The structures at the end point of the MD simulations (Supplementary Fig. S4B) as well as the analysis of the evolution of the secondary structure (Supplementary Fig. S5) show that the helices predicted by I-TASSER for some of the last 49 residues are not stable in solution as they started to unfold during the MD simulation. This indicates that the CT tail has a certain propensity for a helical structure, which is, however, not stable in solution but may be stable under different environments, such as a lipid membrane.

To test whether the CT tail may indeed form a TM helix and might thus act as lipid anchor, which would help in the recruitment of mGBP7 to membranes, various bioinformatic tools, that make such predictions based on the physicochemical properties of the amino acids and their sequence, were employed. First, three amino acid scales using the ProtScale web server [57] were calculated. When interpreted together, these scales allow making predictions about possible TM regions. In particular, the hydrophobicity using the Kyle & Doolittle scale [58], the TM tendency [59], and the buried-residue probability [60] were calculated for each residue (results not shown). The combination of the three scales led to the identification of three possible TM regions for mGBP7: ^{39}VVVAIVGLY^{47}, ^{114}WIFALAVLL^{121}, both in the G domain, and ^{598}FGDVLISVVPGSGKYFGLGLKIL^{620} in the CT tail. It should be noted that the two possible TM regions identified this way for the G domain are unlikely to be TM helices. Both regions involve only nine residues, while for a stable TM helix at least 14 residues are required. In addition, residues 39–47 are part of the β-sheet present in the G domain. To support our finding for the CT tail, the TMpred server [58] was invoked. It

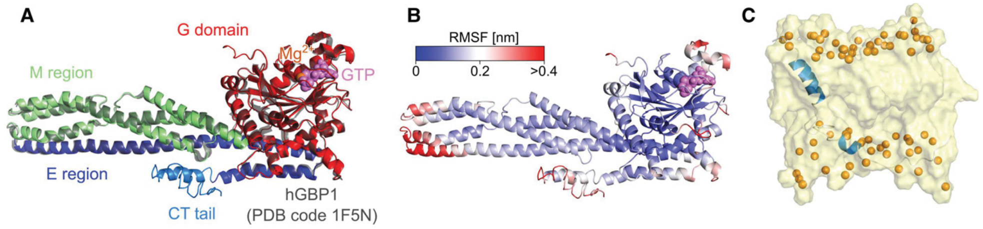


Figure 9. Homology model for full-length mGBP7 and a model for the membrane-inserted CT tail.
(**A**) The mGBP7 homology model produced by I-TASSER (C-score −0.18) is shown as cartoon, including GTP and Mg^{2+} shown as spheres. mGBP7 consists of three different parts: the G domain (red) with GTP (purple) and Mg^{2+} (orange), the M region (green), the E region (blue) including the CT, 49-residues spanning tail (light blue). The crystal structure of hGBP1 (PDB code 1F5N), which was one of the templates, is overlayed in gray. The RMSD between these two structures is 0.5 Å. (**B**) The protein flexibility, quantified as the RMSF of the C_α atoms during the MD simulation, is projected onto the mGBP7 model. Rigid residues are displayed in blue and flexible residues in red according to the color scale shown above the protein model. (**C**) A model for residues 589–620 of the CT tail inserted into a POPC membrane, which was obtained at the end of a 250 ns MD simulation, is shown. Amino acid residues in a helical conformation are colored in blue, coil and turn conformations in gray. The membrane is shown as a yellow transparent surface with the head-group region being indicated by yellow spheres representing the phosphorus atoms.

Biochemical Journal (2019) **476** 3161–3182
https://doi.org/10.1042/BCJ20190364

assigns a score value >0 to a residue if it is likely to be part of a TM helix. Based on the TMpred results this is the case for the CT tail residues 598–620, i.e. the same residues identified by the combined application of the above-mentioned amino acid scales. For some of these residues score values of up to 690 are found, which is above the score threshold of 500 for a significant probability of TM helix formation. Motivated by these findings, in a next step the stability of such a TM CT-tail helix was tested in an MD simulation. To this end, residues 590–620 were pre-inserted into a POPC membrane and their dynamics followed for 250 ns. As initial conformation for residues 590–620 the helical conformation from homology model 3 (Supplementary Fig. S4) was used, which was placed in the membrane such that residues 590–597 were in head-group region of the upper membrane leaflet, allowing the predicted TM region 598–620 to be located in the hydrophobic membrane core (Supplementary Fig. S6). During the MD simulation the TM region adopted a tilted conformation (Figure 9C) as it is longer than the hydrophobic region of a POPC membrane. Another change is that the TM helix is not fully stable as the sequence 598–620 includes two lysine residues, and especially K611 disturbs the TM helix, as it instead prefers to interact with the head-group region of the lower membrane leaflet. Nonetheless, the remaining TM helix is stable, which was not the case during the MD simulations of the corresponding mGBP7 homology model in water, where the CT helix started to unfold within a few nanoseconds (Supplementary Fig. S5). Thus, based on the current simulations and bioinformatic analysis it is suggested that the CT tail may act as lipid anchor for mGBP7.

The elongated CT tail of mGBP7 is required for its localization in VLS and efficient accumulation at the *T. gondii* PV

Previously, mGBP2 and several other mGBPs (mGBP1, 3, 6) have been shown to reside in larger complexes in the cytosol of IFNγ-treated cells that we termed vesicle-like structures (VLS) [6,7,18,71]. Furthermore, using sophisticated fluorescence microscopy techniques it was shown that mGBPs get depleted of these VLS and relocate quickly to the *T. gondii* PV when the parasite is detected [6].

In an effort, to investigate whether the last 49 CT residues (CT tail) of mGBP7 are required for the localization of mGBP7 in VLS and/or for the recruitment of mGBP7 to the *T. gondii* PV, we stably transduced mGBP7$^{-/-}$ MEFs with the respective N-terminal GFP fusion constructs. To verify protein expression levels and the integrity of GFP-mGBP7 and GFP-mGBP7$^{\Delta 49}$, Western Blot analyses were performed (Supplementary Fig. S7A). Confocal microscopy analyses of fixed cells revealed that the truncated mGBP7 (mGBP7$^{\Delta 49}$) does no longer localize in VLS and is evenly distributed within the cytoplasm (Figure 10A). Moreover, and even more intriguing, *in vitro* analysis of *T. gondii* infected cells demonstrated that in contrast with the WT protein, mGBP7$^{\Delta 49}$ showed significant decreased accumulation effects at the PV of the parasite (Figure 10B and Supplementary Fig. S7B).

From these results, it can be concluded that the 49 CT residues are essential for the typical mGBP7 localization in VLS and translocation of the protein to the PV of *T. gondii*.

Discussion

In this study, the purification and biochemical characterization of mGBP7 is reported for the first time. mGBP7 is a member of the IFN-inducible GTPase superfamily which has gained attention in the last decades due to its outstanding ability to specifically target intracellular vacuolar pathogens such as *T. gondii* or *C. trachomatis* and to inhibit their replication by leading to the destruction of their vacuolar compartment [6,72]. Previously, we systematically analyzed mGBP2, the closest murine orthologue of hGBP1. For mGBP2, it was demonstrated that different mutations cause individual defects in nucleotide binding, GTPase activity, and oligomerization capability [7]. We were particularly interested in examining mGBP7 since this with 638 amino acids largest mGBP family member possesses an elongated CT domain and does not have an isoprenylation motif, in contrast with the previously studied GBPs (GBP1, −2, −5) [7,13,15,23,25,73]. Furthermore, this study aimed at determining whether mGBP7 shows distinct differences to mGBP2 with respect to GTP affinity, GTP hydrolysis, and domain structure that could help shed light on their individual biological functions in host defense.

The GTP-binding affinity measurement revealed that mGBP7 has a twofold higher affinity for GTP (K_D of 0.22 μM) than mGBP2 and an approximately fivefold higher affinity for GTP than hGBP1 [7,16,24,74]. As a result, we propose that under physiological intracellular conditions with ∼470 ± 220 μM GTP mGBP7 can efficiently bind and hydrolyze GTP [7,75]. An explanation for this slightly enhanced GTP affinity of mGBP7

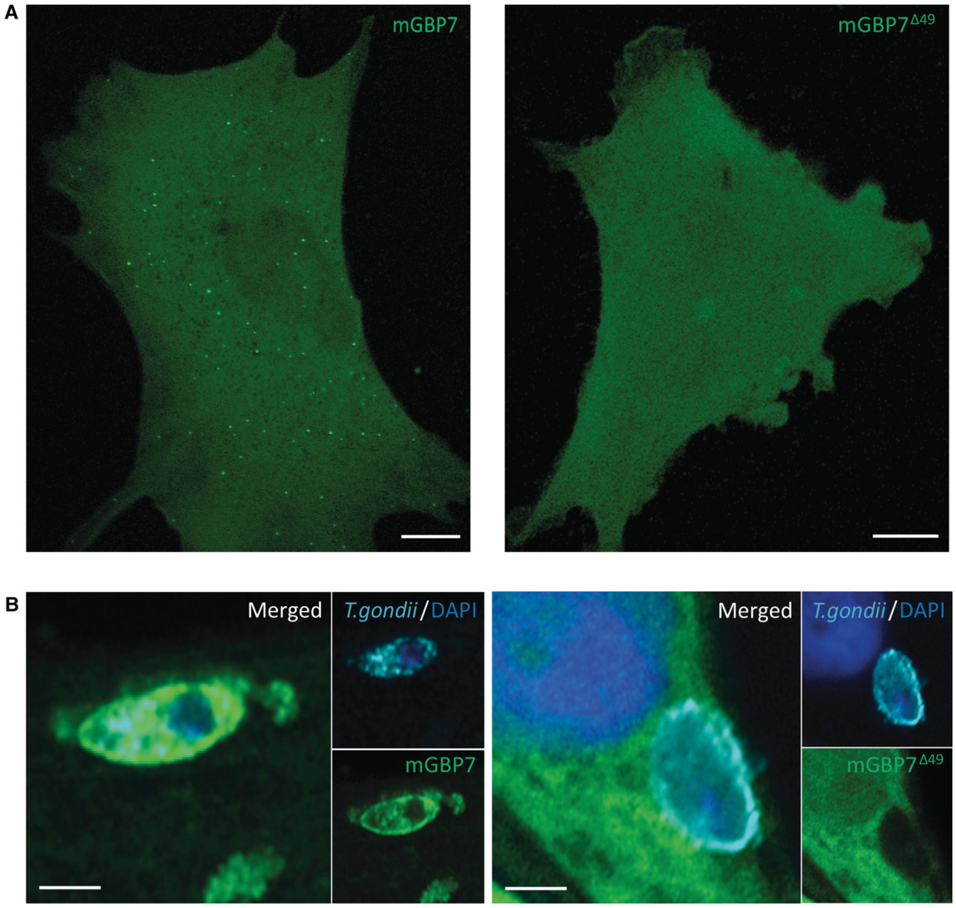


Figure 10. Subcellular localization of WT mGBP7 and mGBP7$^{\Delta 49}$.

Confocal images of N-terminally GFP-tagged mGBP7 proteins in mGBP7$^{-/-}$ MEF cells. Cells were stimulated with IFNγ for 16 h prior to fixation. (**A**) Intracellular localization of WT mGBP7 (left) and mGBP7$^{\Delta 49}$(right). WT mGBP7 forms distinct subcellular VLS whereas mGBP7$^{\Delta 49}$ shows a mostly homogeneous distribution in the cytoplasm ($n = 3$). Scale bars, 5 μm. (**B**) Accumulation of WT mGBP7 and mGBP7$^{\Delta 49}$ at the PV of ME49 *T. gondii* 2 h post-infection. WT mGBP7 accumulates at the PV of *T. gondii*. For the truncation mutant mGBP7$^{\Delta 49}$ virtually no accumulation at the PV is detectable. The cell nuclei were stained with DAPI and *T. gondii* was visualized by staining the parasite with the surface marker SAG1 using a mAB to [TP3] (secondary AB: Alexa Fluor 633, cyan). Scale bars, 2 μm.

compared with hGBP1 and mGBP2 is possibly originating from differences in their GTPase domain and/or minute changes in their conformation. Analysis of the GTP-binding motifs in the G domain reveals that mGBP7 has a 'TVRD' motif in contrast with hGBP1 and mGBP2 that possess a 'TLRD' motif (Figure 8A) [14,16]. This amino acid variation in the G4 motif results in a conservative exchange of aliphatic (L/V) amino acids and could be interesting to analyze by mutational analysis for its effects on GTP-binding affinity. In general, the apparent substrate affinity of mGBP7 for GTP is in the expected range for members of the dynamin superfamily, but is certainly much lower than the GTP-binding affinities of members of the small GTPase families such as Ras or Gα proteins [16]. It was previously reported that the K51A mutation drastically impairs the GTP-binding affinity for hGBP1 (~50-fold) and mGBP2 (~100-fold) [7,24]. For mGBP7, this effect is even more pronounced, since no binding of the K51A mutant protein to GTPγS was measured, emphasizing the critical role of the K51 residue for nucleotide binding.

First, the GTPase measurements were performed at constant protein concentration to determine the optimal nucleotide concentration, and then the amount of protein was varied to define the protein concentration that is required for 100% protein activity. The GTPase measurements that were finally performed at the optimal mGBP7 concentration and fitted using the Michaelis–Menten equation yielded an apparent $v_{\text{max, app}}$ of 278.9 ± 9.8 and an apparent $K_{\text{M, app}}$ of 207.0 ± 32.3 μM GTP (Figure 3C, $\text{WT}_{0.83}$). Furthermore, the obtained sigmoidal binding curve when plotting the enzyme activity in relation to the protein concentration indicated a positive cooperativity of GTP hydrolysis (Figure 3D). Such positive cooperativity that GTP hydrolysis is promoted enzyme dimerization has previously also been reported for mGBP2 and hGBP1 [7,17]. Another characteristic for GBPs is that they exhibit a high intrinsic GTPase activity. The GBPs usually possess a turnover rate constant k_{cat} of 2–100 min^{-1} and for mGBP2, we could already determine a comparatively high apparent turnover rate of 102 min^{-1} [7]. The apparent $k_{\text{cat, app}}$ for mGBP7 is 20 min^{-1} (Table 1) and therefore fivefold lower than for mGBP2 or hGBP1 ($k_{\text{cat}} = 95\ \text{min}^{-1}$) but similar to that of hGBP2 with a k_{cat} of 23 min^{-1} [7,24,76]. A direct comparison of mGBP7, having a twofold higher affinity for GTP, and mGBP2, showing a fivefold greater turnover rate, suggests that the overall catalytic efficiency of mGBP7 and mGBP2 is in about the same range.

According to the GTPase activity measurements, GMP and GDP can only serve as substrates for hydrolysis in negligible amounts. However, the MALDI-TOF-MS results revealed that mGBP7 hydrolyzes GTP to both GDP and GMP proposing a two-step, consecutive hydrolysis mechanism of GTP by mGBP7. For hGBP1, it was shown that GTP hydrolysis to GMP indeed occurs via two consecutive cleavages of single phosphate groups and not by a single cleavage of pyrophosphate [28]. We therefore assume that not only hGBP1 (85–90% GMP production) and mGBP2 (74% GMP production) but also mGBP7 releases GMP as reaction product and that the GTP hydrolysis reaction likely involves two consecutive cleavage steps [1,7,24]. This capability distinguishes mGBP2 and mGBP7 from mGBP5 and hGBP5, which mediate hydrolysis of GTP to GDP only [13,32]. This unique feature of GBPs to hydrolyze GTP to a mixture of GDP and GMP with unequal ratios is quite remarkable and opens up the question of its physiological functions, such as whether GBPs are able to target intracellular pathogens more effectively and/or faster by gaining additional energy from the second hydrolysis step, or whether the production of GMP is beneficial or even necessary for the host response of the target cells. These are questions that will have to be addressed in the future to fully understand the complexity of GBPs in host defense.

The MALS results show that mGBP7 forms a transient dimer and that this dimerization is not influenced by the presence of the non-hydrolyzable GTP analog GTPγS. Thus, mGBP7 shows a clear difference to the multimerization properties of mGBP2 [7]. For WT mGBP2, it was demonstrated in gel filtration studies that it elutes as dimer in the presence of GTPγS but as monomer in the absence of nucleotides [7]. These results illustrate that despite the high homology of GBPs, there exist considerable differences in GBP oligomerization behavior that requires further investigation. However, a separated dimer peak was only observed when using very high concentrations of mGBP7 (12 mg/ml). Otherwise, only one broad peak, presumably containing a mixture of monomeric and dimeric species, was observed. This suggests that the dimer formation is transient and that the dimeric state during GTP hydrolysis is only short-lived.

To study the oligomerization pattern of mGBP7 in more detail, BN-PAGE analyses were performed. First, it was investigated whether mGBP7 GTP hydrolysis could be inhibited by γ-phosphate analogs such as orthovanadate, AlF_x, and BeF_x that trap the GTPase in its ground or transition state of GTP hydrolysis in complex with GDP after the first hydrolysis step. So far, only GDP*AlF_x was analyzed and reported to inhibit hGBP1 and mGBP2 GTP hydrolysis [7,17]. Here, it was shown that besides GDP*AlF_x the GTPase activity of GDP-bound mGBP7 can be inhibited by the two other additionally tested γ-phosphate analogs orthovanadate and BeF_x (Table 2). The finding that mGBP7 can be inhibited by AlF_x brings the GBPs closer to the G_α family of proteins that form a stable complex with GDP*AlF_x and the small G proteins that interact with AlF_x in the presence of their respective GTPase-activating proteins [77,78]. Comparable to the MALS result, the BN-PAGE analyses that were performed with GTP and excess of γ-phosphate analogs showed almost no effect on the oligomerization pattern of mGBP7 and the mGBP7 K51A mutant. Thus, in contrast with mGBP2, the GTP hydrolysis reaction does not seem to be a prerequisite for the dimerization of mGBP7.

I-TASSER homology modeling revealed that mGBP7 in comparison with hGBP1 possesses an elongated CT tail of 49 residues, which adopts different helical and/or random coil structures and represents one of the most flexible parts of mGBP7. Furthermore, modeling a structure for mGBP7 without the CT tail indicated that the reliability of the proposed five structure models of the truncated mGBP7 is very high. In addition, MD simulations provided evidence that the 49 CT residues are not stable in solution as they started to unfold during the analyzed 100 ns. We consequently propose that the CT tail of mGBP7 has a certain propensity to form a helical structure, which is flexible but unstable in solution (Supplementary Fig. S4). Moreover, the CT tail alone is predicted to be stable when inserted into a membrane. Thus, this suggested that the CT tail of mGBP7 replaces the role of a Caa*X* motif present in other GBPs (GBP1, −2, and −5) for membrane anchoring.

To test this hypothesis, the subcellular localization of WT mGBP7 and an mGBP7 truncation mutant lacking these last 49 residues (mGBP7$^{\Delta 49}$) was investigated in mGBP7$^{-/-}$ cells. WT mGBP7, as previously observed for mGBP2, formed VLS of heterogeneous size in the cytosol [7]. In contrast, mGBP7$^{\Delta 49}$ showed a more homogeneously distribution and failed to localize to VLS. For mGBP2, this localization to VLS was shown to be dependent on isoprenylation of the Caa*X* box motif [71]. In fact, prenylation was reported to be required for the membrane association for all Caa*X* box-containing hGBPs [79]. Moreover, the Caa*X* sequences of GBP1, −2 and −5 are conserved which suggests an important biological function and allows speculating if other GBPs might have evolved different structures or modes of membrane interactions to compensate for this binding motif.

In addition, the recruitment potential of mGBP7$^{\Delta 49}$ to the outer membrane of the parasite *T. gondii* was assayed since WT mGBP7 and further mGBPs (mGBP1, −2, −3, −6, and −9) efficiently recruit to Toxoplasma parasites [18]. Interestingly, there was almost no translocation of mGBP7$^{\Delta 49}$ to the PV of *T. gondii* detectable. These results are in line with the MD simulations and propose a nonredundant function of the divergent mGBP7 CT tail for correct protein localization, membrane anchoring and/or interaction.

Taken together, the GTPase activity, the hydrolyzation products, the structure and the oligomerization pattern of mGBP7 were characterized. Furthermore, the present study demonstrates a so far not described GTP independent mechanism for mGBP7 dimerization and proposes a new mechanism for the interaction of GBPs with intracellular membranes.

Abbreviations

CT, C-terminal; DAPI, 4,6-diamidino-2-phenylindole; DMEM, Dulbecco's modified Eagle's medium; FBS, fetal bovine serum; GBPs, Guanylate-binding proteins; GTPases, guanosine triphosphatases; I-TASSER, Iterative Threading ASSEmbly Refinement; IC_{50}, half-maximal inhibitory concentration; IFN, interferon; MALDI-TOF-MS, Matrix-Assisted Laser Desorption Ionization Time-Of-Flight Mass Spectrometry; MEFs, murine embryonic fibroblasts; ORF, open reading frame; PDB, Protein Data Bank; POPC, phosphatidylcholine; PV, parasitophorous vacuole; RMSD, root mean square deviations; RMSF, root mean square fluctuations; SEC, Size Exclusion Chromatography; TM, transmembrane; VLS, vesicle-like structures.

Author Contribution

L.L. performed all experiments and wrote the paper draft; J.L., X.W., and B.S. performed molecular modeling and simulation; N.S. performed confocal microscopy experiments; M.P. assisted in MALS experiments; S.S. and L.S. assisted in the analysis and interpretation of biochemical and MALS experiments; D.D. and K.P. designed and supervised the experiments, assisted in drafting and critical reading. All the authors discussed the results and contributed to the writing of the manuscript.

Funding

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—project number 267205415—CRC 1208 to L.S., B.S., D.D., and K.P. J.L., X.W., and B.S. gratefully acknowledge the computing time granted through JARA-HPC (project JICS6A) on the supercomputer JURECA at the Forschungszentrum Juelich. The Center for Structural Studies is funded by the Deutsche Forschungsgemeinschaft (DFG Grant number 417919780).

Acknowledgements

We would like to thank Julia Mock, Karin Buchholz, Nicole Kuepper, Anne Tersteegen, Jens Lichte, Elisabeth Kravets, Olivia Spitz, and Tobias Beer for experimental assistance, and Ursula R. Sorg for critical reading of the manuscript. Furthermore, we thank Dr. Bogdan Barz for fruitful discussions.

Competing Interests

The Authors declare that there are no competing interests associated with the manuscript.

References

1. Vestal, D.J. and Jeyaratnam, J.A. (2011) The guanylate-binding proteins: emerging insights into the biochemical properties and functions of this family of large interferon-induced guanosine triphosphatase. *J. Interferon Cytokine Res.* **31**, 89–97 https://doi.org/10.1089/jir.2010.0102
2. Praefcke, G.J. and McMahon, H.T. (2004) The dynamin superfamily: universal membrane tubulation and fission molecules? *Nat. Rev. Mol. Cell Biol.* **5**, 133–147 https://doi.org/10.1038/nrm1313
3. Faelber, K., Gao, S., Held, M., Posor, Y., Haucke, V., Noe, F. et al. (2013) Oligomerization of dynamin superfamily proteins in health and disease. *Prog. Mol. Biol. Transl. Sci.* **117**, 411–443 https://doi.org/10.1016/B978-0-12-386931-9.00015-5
4. Haller, O., Staeheli, P., Schwemmle, M. and Kochs, G. (2015) Mx GTPases: dynamin-like antiviral machines of innate immunity. *Trends Microbiol.* **23**, 154–163 https://doi.org/10.1016/j.tim.2014.12.003
5. Carter, C.C., Gorbacheva, V.Y. and Vestal, D.J. (2005) Inhibition of VSV and EMCV replication by the interferon-induced GTPase, mGBP-2: differential requirement for wild-type GTP binding domain. *Arch. Virol.* **150**, 1213–1220 https://doi.org/10.1007/s00705-004-0489-2
6. Kravets, E., Degrandi, D., Ma, Q., Peulen, T.O., Klumpers, V., Felekyan, S. et al. (2016) Guanylate binding proteins directly attack *Toxoplasma gondii* via supramolecular complexes. *eLife* **5**, e11479 https://doi.org/10.7554/eLife.11479
7. Kravets, E., Degrandi, D., Weidtkamp-Peters, S., Ries, B., Konermann, C., Felekyan, S. et al. (2012) 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.* **287**, 27452–27466 https://doi.org/10.1074/jbc.M112.379636
8. Degrandi, D., Kravets, E., Konermann, C., Beuter-Gunia, C., Klumpers, V., Lahme, S. et al. (2013) Murine guanylate binding protein 2 (mGBP2) controls *Toxoplasma gondii* replication. *Proc. Natl Acad. Sci. U.S.A.* **110**, 294–299 https://doi.org/10.1073/pnas.1205635110
9. Lindenberg, V., Molleken, K., Kravets, E., Stallmann, S., Hegemann, J.H., Degrandi, D. et al. (2017) Broad recruitment of mGBP family members to *Chlamydia trachomatis* inclusions. *PLoS ONE* **12**, e0185273 https://doi.org/10.1371/journal.pone.0185273
10. Tietzel, I., El-Haibi, C. and Carabeo, R.A. (2009) Human guanylate binding proteins potentiate the anti-chlamydia effects of interferon-gamma. *PLoS ONE* **4**, e6499 https://doi.org/10.1371/journal.pone.0006499
11. Kim, B.H., Shenoy, A.R., Kumar, P., Das, R., Tiwari, S. and MacMicking, J.D. (2011) A family of IFN-gamma-inducible 65-kD GTPases protects against bacterial infection. *Science* **332**, 717–721 https://doi.org/10.1126/science.1201711
12. Feeley, E.M., Pilla-Moffett, D.M., Zwack, E.E., Piro, A.S., Finethy, R., Kolb, J.P. et al. (2017) Galectin-3 directs antimicrobial guanylate binding proteins to vacuoles furnished with bacterial secretion systems. *Proc. Natl Acad. Sci. U.S.A.* **114**, E1698–EE706 https://doi.org/10.1073/pnas.1615771114
13. Shenoy, A.R., Wellington, D.A., Kumar, P., Kassa, H., Booth, C.J., Cresswell, P. et al. (2012) GBP5 promotes NLRP3 inflammasome assembly and immunity in mammals. *Science* **336**, 481–485 https://doi.org/10.1126/science.1217141
14. Kresse, A., Konermann, C., Degrandi, D., Beuter-Gunia, C., Wuerthner, J., Pfeffer, K. et al. (2008) Analyses of murine GBP homology clusters based on in silico, in vitro and in vivo studies. *BMC Genomics* **9**, 158 https://doi.org/10.1186/1471-2164-9-158
15. Cheng, Y.S., Patterson, C.E. and Staeheli, P. (1991) Interferon-induced guanylate-binding proteins lack an N(T)KXD consensus motif and bind GMP in addition to GDP and GTP. *Mol. Cell Biol.* **11**, 4717–4725 https://doi.org/10.1128/MCB.11.9.4717
16. Praefcke, G.J., Geyer, M., Schwemmle, M., Robert Kalbitzer, H. and Herrmann, C. (1999) Nucleotide-binding characteristics of human guanylate-binding protein 1 (hGBP1) and identification of the third GTP-binding motif. *J. Mol. Biol.* **292**, 321–332 https://doi.org/10.1006/jmbi.1999.3062
17. Prakash, B., Praefcke, G.J., Renault, L., Wittinghofer, A. and Herrmann, C. (2000) Structure of human guanylate-binding protein 1 representing a unique class of GTP-binding proteins. *Nature.* **403**, 567–571 https://doi.org/10.1038/35000617
18. Degrandi, D., Konermann, C., Beuter-Gunia, C., Kresse, A., Wurthner, J., Kurig, S. et al. (2007) Extensive characterization of IFN-induced GTPases mGBP1 to mGBP10 involved in host defense. *J. Immunol.* **179**, 7729–7740 https://doi.org/10.4049/jimmunol.179.11.7729
19. Olszewski, M.A., Gray, J. and Vestal, D.J. (2006) In silico genomic analysis of the human and murine guanylate-binding protein (GBP) gene clusters. *J. Interferon Cytokine Res.* **26**, 328–352 https://doi.org/10.1089/jir.2006.26.328
20. Vopel, T., Hengstenberg, C.S., Peulen, T.O., Ajaj, Y., Seidel, C.A., Herrmann, C. et al. (2014) 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* **53**, 4590–4600 https://doi.org/10.1021/bi500524u
21. Nantais, D.E., Schwemmle, M., Stickney, J.T., Vestal, D.J. and Buss, J.E. (1996) Prenylation of an interferon-gamma-induced GTP-binding protein: the human guanylate binding protein, huGBP1. *J. Leukoc. Biol.* **60**, 423–431 https://doi.org/10.1002/jlb.60.3.423
22. Vopel, T., Syguda, A., Britzen-Laurent, N., Kunzelmann, S., Ludemann, M.B., Dovengerds, C. et al. (2010) Mechanism of GTPase-activity-induced self-assembly of human guanylate binding protein 1. *J. Mol. Biol.* **400**, 63–70 https://doi.org/10.1016/j.jmb.2010.04.053
23. Ince, S., Kutsch, M., Shydlovskyi, S. and Herrmann, C. (2017) The human guanylate-binding proteins hGBP-1 and hGBP-5 cycle between monomers and dimers only. *FEBS J.* **284**, 2284–2301 https://doi.org/10.1111/febs.14126
24. Praefcke, G.J., Kloep, S., Benscheid, U., Lilie, H., Prakash, B. and Herrmann, C. (2004) Identification of residues in the human guanylate-binding protein 1 critical for nucleotide binding and cooperative GTP hydrolysis. *J. Mol. Biol.* **344**, 257–269 https://doi.org/10.1016/j.jmb.2004.09.026
25. Wehner, M., Kunzelmann, S. and Herrmann, C. (2012) The guanine cap of human guanylate-binding protein 1 is responsible for dimerization and self-activation of GTP hydrolysis. *FEBS J.* **279**, 203–210 https://doi.org/10.1111/j.1742-4658.2011.08415.x
26. Kochs, G. and Haller, O. (2010) Chapter 226—Mx Proteins: High Molecular Weight GTPases with Antiviral Activity. In *Handbook of Cell Signaling*, 2nd edn, (Bradshaw, R.A. and Dennis, E.A., eds), pp. 1855–1864, Academic Press, San Diego
27. Tripathi, R., Glaves, R. and Marx, D. (2017) The GTPase hGBP1 converts GTP to GMP in two steps via proton shuttle mechanisms. *Chem. Sci.* **8**, 371–380 https://doi.org/10.1039/C6SC02045C
28. Schwemmle, M. and Staeheli, P. (1994) The interferon-induced 67-kDa guanylate-binding protein (hGBP1) is a GTPase that converts GTP to GMP. *J. Biol. Chem.* **269**, 11299–11305 PMID:7512561
29. Ghosh, A., Praefcke, G.J., Renault, L., Wittinghofer, A. and Herrmann, C. (2006) How guanylate-binding proteins achieve assembly-stimulated processive cleavage of GTP to GMP. *Nature* **440**, 101–104 https://doi.org/10.1038/nature04510
30. Abdullah, N., Balakumari, M. and Sau, A.K. (2010) Dimerization and its role in GMP formation by human guanylate binding proteins. *Biophys. J.* **99**, 2235–2244 https://doi.org/10.1016/j.bpj.2010.07.025

31 Fres, J.M., Muller, S. and Praefcke, G.J. (2010) Purification of the CaaX-modified, dynamin-related large GTPase hGBP1 by coexpression with farnesyltransferase. *J. Lipid Res.* **51**, 2454–2459 https://doi.org/10.1194/jlr.D005397
32 Wehner, M. and Herrmann, C. (2010) Biochemical properties of the human guanylate binding protein 5 and a tumor-specific truncated splice variant. *FEBS J.* **277**, 1597–1605 https://doi.org/10.1111/j.1742-4658.2010.07586.x
33 Cheng, Y.S., Colonno, R.J. and Yin, F.H. (1983) Interferon induction of fibroblast proteins with guanylate binding activity. *J. Biol. Chem.* **258**, 7746–7750 PMID:6305951
34 Kim, B.H., Shenoy, A.R., Kumar, P., Bradfield, C.J. and MacMicking, J.D. (2012) IFN-inducible GTPases in host cell defense. *Cell Host Microbe* **12**, 432–444 https://doi.org/10.1016/j.chom.2012.09.007
35 pWPXL was a gift from Didier Trono (Addgene plasmid # 12257; http://n2t.net/addgene:12257; RRID:Addgene_12257)
36 Baykov, A.A., Evtushenko, O.A. and Avaeva, S.M. (1988) A malachite green procedure for orthophosphate determination and its use in alkaline phosphatase-based enzyme immunoassay. *Anal. Biochem.* **171**, 266–270 https://doi.org/10.1016/0003-2697(88)90484-8
37 Reimann, S., Poschmann, G., Kanonenberg, K., Stuhler, K., Smits, S.H. and Schmitt, L. (2016) Interdomain regulation of the ATPase activity of the ABC transporter haemolysin B from *Escherichia coli*. *Biochem. J.* **473**, 2471–2483 https://doi.org/10.1042/BCJ20160154
38 Gordon, J.A. (1991) Use of vanadate as protein-phosphotyrosine phosphatase inhibitor. *Methods Enzymol.* **201**, 477–482 https://doi.org/10.1016/0076-6879(91)01043-2
39 Kluth, M., Stindt, J., Dröge, C., Linnemann, D., Kubitz, R. and Schmitt, L. (2015) A mutation within the extended X loop abolished substrate-induced ATPase activity of the human liver ATP-binding cassette (ABC) transporter MDR3. *J. Biol. Chem.* **290**, 4896–4907 https://doi.org/10.1074/jbc.M114.588566
40 Herrmann, C. and Nassar, N. (1996) *Ras* and its effectors. *Prog. Biophys. Mol. Biol.* **66**, 1–41 https://doi.org/10.1016/S0079-6107(96)00015-6
41 Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T. et al. (2012) Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 https://doi.org/10.1038/nmeth.2019
42 Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J. and Zhang, Y. (2015) The I-TASSER suite: protein structure and function prediction. *Nat. Methods* **12**, 7–8 https://doi.org/10.1038/nmeth.3213
43 Roy, A., Kucukural, A. and Zhang, Y. (2010) I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.* **5**, 725–738 https://doi.org/10.1038/nprot.2010.5
44 Zhang, Y. (2008) I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* **9**, 40 https://doi.org/10.1186/1471-2105-9-40
45 Mitoma, H., Hanabuchi, S., Kim, T., Bao, M., Zhang, Z., Sugimoto, N. et al. (2013) The DHX33 RNA helicase senses cytosolic RNA and activates the NLRP3 inflammasome. *Immunity* **39**, 123–135 https://doi.org/10.1016/j.immuni.2013.07.001
46 Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J.C., Hess, B. et al. (2015) GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1-2**, 19–25 https://doi.org/10.1016/j.softx.2015.06.001
47 Abraham, M.J., van der Spoel, D., Lindahl, E. and Hess, B. and the GROMACS development team. (2018) GROMACS User Manual version 2016.4 [www.gromacs.org]
48 Best, R.B. and Hummer, G. (2009) Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides. *J. Phys. Chem. B* **113**, 9004–9015 https://doi.org/10.1021/jp901540t
49 Aliev, A.E., Kulke, M., Khaneja, H.S., Chudasama, V., Sheppard, T.D. and Lanigan, R.M. (2014) Motional timescale predictions by molecular dynamics simulations: case study using proline and hydroxyproline sidechain dynamics. *Proteins* **82**, 195–215 https://doi.org/10.1002/prot.24350
50 Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W. and Klein, M.L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 https://doi.org/10.1063/1.445869
51 Hoover, W.G. (1985) Canonical dynamics: equilibrium phase-space distributions. *Phys. Rev. A Gen. Phys.* **31**, 1695–1697 https://doi.org/10.1103/PhysRevA.31.1695
52 Nose, S. (1984) A molecular-dynamics method for simulations in the canonical ensemble. *Mol. Phys.* **52**, 255–268 https://doi.org/10.1080/00268978400101201
53 Parrinello, M. and Rahman, A. (1981) Polymorphic transitions in single-crystals—a new molecular-dynamics method. *J. Appl. Phys.* **52**, 7182–7190 https://doi.org/10.1063/1.328693
54 Darden, T., York, D. and Pedersen, L. (1993) Particle mesh Ewald—an N.Log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092 https://doi.org/10.1063/1.464397
55 Essmann, U., Perera, L., Berkowitz, M.L., Darden, T., Lee, H. and Pedersen, L.G. (1995) A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577–8593 https://doi.org/10.1063/1.470117
56 Hess, B., Bekker, H., Berendsen, H.J.C. and Fraaije, J.G.E.M. (1997) LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* **18**, 1463–1472 https://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H
57 Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D. et al. (2005) Protein Identification and Analysis Tools on the ExPASy Server. In *The Proteomics Protocols Handbook* (Walker, J.M., ed.), pp. 571–607, Humana Press, Totowa, NJ
58 Kyte, J. and Doolittle, R.F. (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132 https://doi.org/10.1016/0022-2836(82)90515-0
59 Zhao, G. and London, E. (2006) An amino acid “transmembrane tendency” scale that approaches the theoretical limit to accuracy for prediction of transmembrane helices: relationship to biological hydrophobicity. *Protein Sci.* **15**, 1987–2001 https://doi.org/10.1110/ps.062286306
60 Janin, J. (1979) Surface and inside volumes in globular proteins. *Nature* **277**, 491–492 https://doi.org/10.1038/277491a0
61 Huang, J. and MacKerell, Jr, A.D. (2013) CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *J. Comput. Chem.* **34**, 2135–2145 https://doi.org/10.1002/jcc.23354
62 Pastor, R.W. and Mackerell, Jr, A.D. (2011) Development of the CHARMM force field for lipids. *J. Phys. Chem. Lett.* **2**, 1526–1532 https://doi.org/10.1021/jz200167q
63 Jo, S., Kim, T., Iyer, V.G. and Im, W. (2008) CHARMM-GUI: a web-based graphical user interface for CHARMM. *J. Comput. Chem.* **29**, 1859–1865 https://doi.org/10.1002/jcc.20945
64 DeLano, W.L. (2002) *The PyMOL Molecular Graphics System*, DeLano Scientific, Palo Alto, CA, U.S.A.

65 Kabsch, W. and Sander, C. (1983) Dictionary of protein secondary structure—pattern-recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637 https://doi.org/10.1002/bip.360221211

66 Zaitseva, J., Jenewein, S., Wiedenmann, A., Benabdelhak, H., Holland, I.B. and Schmitt, L. (2005) Functional characterization and ATP-induced dimerization of the isolated ABC-domain of the haemolysin B transporter. *Biochemistry* **44**, 9680–9690 https://doi.org/10.1021/bi0506122

67 Praefcke, G.J.K. (2017) Regulation of innate immune functions by guanylate-binding proteins. *Int. J. Med. Microbiol.* **308**, 237–245 https://doi.org/10.1016/j.ijmm.2017.10.013

68 Wyatt, P.J. (1998) Submicrometer particle sizing by multiangle light scattering following fractionation. *J. Colloid Interface Sci.* **197**, 9–20 https://doi.org/10.1006/jcis.1997.5215

69 Prakash, B., Renault, L., Praefcke, G.J., Herrmann, C. and Wittinghofer, A. (2000) Triphosphate structure of guanylate-binding protein 1 and implications for nucleotide binding and GTPase mechanism. *EMBO J.* **19**, 4555–4564 https://doi.org/10.1093/emboj/19.17.4555

70 Altschul, S.F., Wootton, J.C., Gertz, E.M., Agarwala, R., Morgulis, A., Schaffer, A.A. et al. (2005) Protein database searches using compositionally adjusted substitution matrices. *FEBS J.* **272**, 5101–5109 https://doi.org/10.1111/j.1742-4658.2005.04945.x

71 Vestal, D.J., Gorbacheva, V.Y. and Sen, G.C. (2000) Different subcellular localizations for the related interferon-induced GTPases, MuGBP-1 and MuGBP-2: implications for different functions? *J. Interferon Cytokine Res.* **20**, 991–1000 https://doi.org/10.1089/10799900050198435

72 Meunier, E. and Broz, P. (2016) Interferon-inducible GTPases in cell autonomous and innate immunity. *Cell. Microbiol.* **18**, 168–180 https://doi.org/10.1111/cmi.12546

73 Modiano, N., Lu, Y.E. and Cresswell, P. (2005) Golgi targeting of human guanylate-binding protein-1 requires nucleotide binding, isoprenylation, and an IFN-gamma-inducible cofactor. *Proc. Natl Acad. Sci. U.S.A.* **102**, 8680–8685 https://doi.org/10.1073/pnas.0503227102

74 Kunzelmann, S., Praefcke, G.J. and Herrmann, C. (2005) Nucleotide binding and self-stimulated GTPase activity of human guanylate-binding protein 1 (hGBP1). *Methods Enzymol.* **404**, 512–527 https://doi.org/10.1016/S0076-6879(05)04045-0

75 Traut, T.W. (1994) Physiological concentrations of purines and pyrimidines. *Mol. Cell. Biochem.* **140**, 1–22 https://doi.org/10.1007/BF00928361

76 Neun, R., Richter, M.F., Staeheli, P. and Schwemmle, M. (1996) GTPase properties of the interferon-induced human guanylate-binding protein 2. *FEBS Lett.* **390**, 69–72 https://doi.org/10.1016/0014-5793(96)00628-X

77 Bigay, J., Deterre, P., Pfister, C. and Chabre, M. (1985) Fluoroaluminates activate transducin-GDP by mimicking the gamma-phosphate of GTP in its binding site. *FEBS Lett.* **191**, 181–185 https://doi.org/10.1016/0014-5793(85)80004-1

78 Ahmadian, M.R., Mittal, R., Hall, A. and Wittinghofer, A. (1997) Aluminum fluoride associates with the small guanine nucleotide binding proteins. *FEBS Lett.* **408**, 315–318 https://doi.org/10.1016/S0014-5793(97)00422-5

79 Britzen-Laurent, N., Bauer, M., Berton, V., Fischer, N., Syguda, A., Reipschlager, S. et al. (2010) Intracellular trafficking of guanylate-binding proteins is regulated by heterodimerization in a hierarchical manner. *PLoS ONE* **5**, e14246 https://doi.org/10.1371/journal.pone.0014246

80 Robert, X. and Gouet, P. (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320–W324 https://doi.org/10.1093/nar/gku316

5 Conclusions and outlook

In this thesis work, two kinds of membrane-binding proteins were investigated: GABARAP and GBPs (mGBP2 and mGBP7).

One of the aims of the simulations of GABARAP was to assess the degree of conformational flexibility of this protein. To this end, we performed multiple MD simulations with different force fields and water models to rule out force-field artefacts, employed the HREMD method to enhance the simulations, and conducted mutation studies with the goal to initiate protein flexibility. Special attention was put on the α1-2 helices of GABARAP as predictions based on Förster resonance energy transfer spectroscopy suggested these two helices to be very flexible. However, none of our simulation approaches identified large motions of α1-2. The MD simulations of monomeric GABARAP anchored to a membrane allowed us to identify several residues which contribute the most to the membrane-binding process, in addition to the lipid anchor itself. These simulation studies helped to better understand the S^2 order parameters measured in parallel in NMR experiments. Previously predicted dimeric models of GABARAP were also investigated in solution and bound to a membrane. However, these simulations revealed these dimer models to be unstable, questioning the validity of these models. This suggests that predicting the dimeric pattern of a protein remains a difficult task, and in the particular case of GABARAP further efforts are needed. In general, combining experimental data with computational methods is a promising approach to predict the dimeric structure of a protein.

With regard to the GBPs, mGBP2 and mGBP7 were investigated in this thesis. We found similar large-scale hinge motions of mGBP2 with and without GTP being bound; the same kind of motion was also found in hGBP1 in a previous simulation study. The binding of the mGBP2 monomer to a lipid membrane did not affect much the hinge motion. On the other hand, this motion did also not cause significant changes in membrane properties. While we suspect that the hinge motion plays a role in the destruction of the parasitophorous vacuole membrane, it seems that membrane-bound mGBP2 or mGBP7 polymers are needed for this, in order to create a force large enough to cause a membrane-damaging effect. Therefore, future studies should simulate larger mGBP2/7 systems to verify this hypothesis. The power of combining experimental and simulation methods was demonstrated in the third part of this thesis work, as predictions made from simulation studies led to mutation studies in the wet lab, which confirmed the necessity of the CT tail of mGBP7 for the binding of this protein to the membrane of the parasitophorous vacuole of *Toxoplasma gondii*.

The findings from this thesis work confirm the general conclusion that *in silico* methods are able to help figuring out the mechanisms underlying a large number of biological problems. For instance, biomolecular simulations help to predict and optimize the structure of biomolecules, such as proteins (as demonstrated here for mGBP7), RNAs or DNAs. They further aid in discovering the functional mechanisms of proteins and other biomolecules at atomic level, as was established here for the membrane binding of mGBP7. The ability of uncovering the structural basis of diseases even provides novel insights for the development of new drugs or treatments, which however is beyond the scope of the current thesis.

Nonetheless, there are also still limitations and deficiencies in the use of *in silico* methods, especially when not combined with experimental data. This even applies when it comes to the prediction of protein structures using AlphaFold, even though its invention can be considered as the most notable milestone in protein structure determination in recent years: "*Alphafold can accurately predict 3D models of prote*in *structures and has the potential to accelerate research in every field of biology* [147]." Nevertheless, AlphaFold still has its shortcomings. It uses training data to detect a protein structure based on its amino acid sequence, which means the intelligence of AlphaFold is confined to the training data. It still needs to face the challenge of a nearly infinite number of possible amino acid sequences of proteins and its limitation of finite training data. Moreover, as proteins are usually flexible molecules, as best seen in this work for mGBP2, some of AlphaFold's predictions are less accurate than those from more traditional methods, further necessitating the use of MD simulations to predict protein structures in their different conformational states [148].

In this work, MD simulations were employed to study the dynamics of different proteins. The usage of the REMD technique can substantially enhance the computational efficiency, as measured by the decrease in the simulation time required to achieve a particular statistical accuracy [149]. However, in the case of GABARAP even the application of the

REMD technique did not lead to protein structures substantially different from the numerous crystal and NMR structures that were deposited for GABARAP in the Protein Data Bank. This either suggests that GABARAP is a very stable protein resistant to large-scale protein motions, or that even longer or more enhanced simulations are required to observe the slow protein motions that need to overcome large energy barriers. Another way of reducing the computational cost needed for an MD simulation is the application of a coarse-grained force field, instead of an all-atom force field used in the current work. Apart from this, the reduction of the degrees of freedom in the system also makes it easier to interpret the simulation data. However, the disadvantages of coarse graining are obvious too: the limited accuracy of a CG model and the fewer information generated in a CGMD simulation compared to the corresponding AAMD simulation [150]. Therefore, the integration of AAMD, REMD and CGMD simulations in one project is a popular trend, as a wealth of information obtained from simulations of various scales will enrich the results. When combined with biology, biochemical or biophysical experiments, the accuracy of the results will improve even further, as was shown in this work for mGBP7. In order to better simulate the biological systems with *in silico* methods, there is still a high demand on increasing the sampling in terms of system size and simulation time. With the development of computing power and the continuous efforts of the scientists from different fields, it is believed that this goal will be achieved in the near future.

Bibliography

[1] A. P. A. Eric Alm, “Biological networks,” *Curr. Opin. Struct. Biol.*, vol. 13, no. 2, pp. 193–202, 2003.

[2] I. Xenarios, Ł. Salwínski, X. J. Duan, P. Higney, S. M. Kim, and D. Eisenberg, “DIP, the Database of Interacting Proteins: A research tool for studying cellular networks of protein interactions,” *Nucleic Acids Res.*, vol. 30, no. 1, pp. 303–305, 2002.

[3] M. A. Williams, J. M. Goodfellow, and J. M. Thornton, “Buried waters and internal cavities in monomeric proteins,” *Protein Sci.*, vol. 3, no. 8, pp. 1224–1235, 1994.

[4] S. J. Hubbard, K. H. Gross, and P. Argos, “Intramolecular cavities in globular proteins,” *Protein Eng. Des. Sel.*, vol. 7, no. 5, pp. 613–626, 1994.

[5] J. Monod, J. Wyman, and J.-P. Changeux, “On the nature of allosteric transitions: A plausible model,” *J. Mol. Biol.*, vol. 12, no. 1, pp. 88–118, 1965.

[6] R. A. Cook and D. E. J. Koshland, “Specificity in the assembly of multisubunit proteins.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 64, no. 1, pp. 247–254, Sep. 1969.

[7] G. Weber, “Phenomenological description of the association of protein subunits subjected to conformational drift. Effects of dilution and of hydrostatic pressure,” *Biochemistry*, vol. 25, no. 12, pp. 3626–3631, Jun. 1986.

[8] A. Maeno and K. Akasaka, “High-Pressure Fluorescence Spectroscopy.,” *Subcell. Biochem.*, vol. 72, pp. 687–705, 2015.

[9] G. Mei, A. Di Venere, N. Rosato, and A. Finazzi-Agrò, “The importance of being dimeric,” *FEBS J.*, vol. 272, no. 1, pp. 16–27, 2005.

[10] H. Berman, K. Henrick, and H. Nakamura, “Announcing the worldwide Protein Data Bank,” *Nat. Struct. Mol. Biol.*, vol. 10, no. 12, p. 980, 2003.

[11] M. A. J. Ferguson, “Lipid anchors on membrane proteins,” *Curr. Opin. Struct. Biol.*, vol. 1, no. 4, pp. 522–529, 1991.

[12] P. J. Casey and M. C. Seabra, “Protein Prenyltransferases (∗),” *J. Biol. Chem.*, vol. 271, no. 10, pp. 5289–5292, Mar. 1996.

[13] G. Novelli and M. R. D’Apice, “Protein farnesylation and disease,” *J. Inherit. Metab. Dis.*, vol. 35, no. 5, pp. 917–926, 2012.

[14] M. D. Resh, “Trafficking and signaling by fatty-acylated and prenylated proteins.,” *Nat. Chem. Biol.*, vol. 2, no. 11, pp. 584–590, Nov. 2006.

[15] J. P. Wilson, A. S. Raghavan, Y.-Y. Yang, G. Charron, and H. C. Hang, “Proteomic analysis of fatty-acylated proteins in mammalian cells with chemical reporters reveals S-acylation of histone H3 variants.,” *Mol. Cell. Proteomics*, vol. 10, no. 3, p. M110.001198, Mar. 2011.

[16] T. Kinoshita and M. Fujita, “Biosynthesis of GPI-anchored proteins: special emphasis on GPI lipid remodeling.,” *J. Lipid Res.*, vol. 57, no. 1, pp. 6–24, Jan. 2016.

[17] M. D. S. Resh, “Covalent Lipid Modifications of Proteins,” *Curr. Biol.*, vol. 23, no. 10, pp. 431–435, 2013.

[18] C. de Duve, “The Lysosome Concept,” *Ciba Foundation Symposium - Anterior Pituitary Secretion (Book I of Colloquia on Endocrinology)*. pp. 1–35, 01-Jan-1963.

[19] N. Mizushima and M. Komatsu, “Autophagy: renovation of cells and tissues.,” *Cell*, vol. 147, no. 4, pp. 728–741, Nov. 2011.

[20] D. R. Green and B. Levine, “To be or not to be? How selective autophagy and cell death govern cell fate.,” *Cell*, vol. 157, no. 1, pp. 65–75, Mar. 2014.

[21] T. N. Nguyen, B. S. Padman, J. Usher, V. Oorschot, G. Ramm, and M. Lazarou, “Atg8 family LC3 / GAB ARAP proteins are crucial for autophagosome – lysosome fusion but not autophagosome formation during PINK1 / Parkin mitophagy and starvation,” *J. Cell Biol.*, vol. 215, no. 6, pp. 1–18, 2016.

[22] B. S. Padman, T. N. Nguyen, and M. Lazarou, “Autophagosome formation and cargo sequestration in the absence of LC3/GABARAPs.,” *Autophagy*, vol. 13, no. 4, pp. 772–774, Apr. 2017.

[23] T. Decker, S. Stockinger, M. Karaghiosoff, M. Müller, and P. Kovarik, “IFNs and STATs in innate immunity to microorganisms,” *J. Clin. Invest.*, vol. 109, no. 10, pp. 1271–1277, May 2002.

[24] J. D. MacMicking, “Interferon-inducible effector mechanisms in cell-autonomous immunity,” *Nat. Rev. Immunol.*, vol. 12, no. 5, pp. 367–382, 2012.

[25] D. Degrandi *et al.*, “Murine Guanylate Binding Protein 2 (mGBP2) controls Toxoplasma gondii replication,” *Proc. Natl. Acad. Sci.*, vol. 110, no. 1, pp. 294 LP – 299, Jan. 2013.

[26] E. Meunier *et al.*, “Guanylate-binding proteins promote activation of the AIM2 inflammasome during infection with Francisella novicida,” *Nat. Immunol.*, vol. 16, no. 5, pp. 476–484, 2015.

[27] 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.

[28] B. Prakash, G. J. K. Praefcke, L. Renault, A. Wittinghofer, and C. Herrmann, “Structure of human guanylate-binding protein 1 representing a unique class of GTP-binding proteins,” *Nature*, vol. 403, no. 6769, pp. 567–571, 2000.

[29] A. Kresse *et al.*, “Analyses of murine GBP homology clusters based on in silico, in vitro and in vivo studies.,” *BMC Genomics*, vol. 9, p. 158, Apr. 2008.

[30] S. Virreira Winter *et al.*, “Determinants of GBP recruitment to Toxoplasma gondii vacuoles and the parasitic factors that control it.,” *PLoS One*, vol. 6, no. 9, p. e24434, 2011.

[31] L. Sistemich and C. Herrmann, “Purification of Farnesylated hGBP1 and Characterization of Its Polymerization and Membrane Binding.,” *Methods Mol. Biol.*, vol. 2159, pp. 67–81, 2020.

[32] N. Steffens *et al.*, “Essential Role of mGBP7 for Survival of Toxoplasma gondii Infection.,” *MBio*, vol. 11, no. 1, Jan. 2020.

[33] T. U. Consortium, “UniProt: a worldwide hub of protein knowledge,” *Nucleic Acids Res.*, vol. 47, no. D1, pp. D506–D515, Jan. 2019.

[34] X. Robert and P. Gouet, “Deciphering key features in protein structures with the new ENDscript server,” *Nucleic Acids Res.*, vol. 42, no. W1, pp. W320–W324, Jul. 2014.

[35] J. Jumper *et al.*, “Highly accurate protein structure prediction with AlphaFold.,” *Nature*, vol. 596, no. 7873, pp. 583–589, Aug. 2021.

[36] M. J. Bower, F. E. Cohen, and R. L. Dunbrack, “Prediction of protein side-chain rotamers from a backbone-dependent rotamer library: a new homology modeling tool11Edited by B. Honig,” *J. Mol. Biol.*, vol. 267, no. 5, pp. 1268–1282, 1997.

[37] T. A. Jones and S. Thirup, “Using known substructures in protein model building and crystallography,” *EMBO J.*, vol. 5, no. 4, pp. 819–822, Apr. 1986.

[38] T. Schwede, J. Kopp, N. Guex, and M. C. Peitsch, “SWISS-MODEL: An automated protein homology-modeling server,” *Nucleic Acids Res.*, vol. 31, no. 13, pp. 3381–3385, Jul. 2003.

[39] K. Arnold, L. Bordoli, J. Kopp, and T. Schwede, “The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling,” *Bioinformatics*, vol. 22, no. 2, pp. 195–201, Jan. 2006.

[40] S. Bienert *et al.*, “The SWISS-MODEL Repository-new features and functionality,” *Nucleic Acids Res.*, vol. 45, no. D1, pp. D313–D319, Jan. 2017.

[41] N. Eswar *et al.*, “Tools for comparative protein structure modeling and analysis,” *Nucleic Acids Res.*, vol. 31, no. 13, pp. 3375–3380, Jul. 2003.

[42] H. B. Uchôa, G. E. Jorge, N. J. Freitas Da Silveira, J. C. Camera, F. Canduri, and W. F. De Azevedo, “Parmodel: a web server for automated comparative modeling of proteins,” *Biochem. Biophys. Res. Commun.*, vol. 325, no. 4, pp. 1481–1486, 2004.

[43] J. A. R. Dalton and R. M. Jackson, “An evaluation of automated homology modelling methods at low target–template sequence similarity,” *Bioinformatics*, vol. 23, no. 15, pp. 1901–1908, Aug. 2007.

[44] U. Pieper *et al.*, “MODBASE: a database of annotated comparative protein structure models and associated resources,” *Nucleic Acids Res.*, vol. 34, no. suppl_1, pp. D291–D295, Jan. 2006.

[45] U. Pieper *et al.*, “ModBase, a database of annotated comparative protein structure models and associated resources,” *Nucleic Acids Res.*, vol. 42, no. D1, pp. D336–D346, Jan. 2014.

[46] S. Montgomerie, J. A. Cruz, S. Shrivastava, D. Arndt, M. Berjanskii, and D. S. Wishart, “PROTEUS2: a web server for comprehensive protein structure prediction and structure-based annotation,” *Nucleic Acids Res.*, vol. 36, no. suppl_2, pp. W202–W209, Jul. 2008.

[47] A. Fiser, R. K. G. Do, and A. Šali, “Modeling of loops in protein structures,” *Protein Sci.*, vol. 9, no. 9, pp. 1753–1773, Jan. 2000.

[48] A. Fiser and A. Sali, “ModLoop: automated modeling of loops in protein structures,” *Bioinformatics*, vol. 19, no. 18, pp. 2500–2501, Dec. 2003.

[49] D. Marsh, *Handbook of Lipid Bilayers*, 2nd ed. CRC press, 2013.

[50] H. Koldsø and M. S. P. Sansom, “Organization and Dynamics of Receptor Proteins in a Plasma Membrane,” *J. Am. Chem. Soc.*, vol. 137, no. 46, pp. 14694–14704, Nov. 2015.

[51] H. I. Ingólfsson *et al.*, "Lipid Organization of the Plasma Membrane," *J. Am. Chem. Soc.*, vol. 136, no. 41, pp. 14554–14559, Oct. 2014.

[52] A. Y. Antipina and A. A. Gurtovenko, "Molecular Mechanism of Calcium-Induced Adsorption of DNA on Zwitterionic Phospholipid Membranes," *J. Phys. Chem. B*, vol. 119, no. 22, pp. 6638–6645, Jun. 2015.

[53] S. Jo, J. B. Lim, J. B. Klauda, and W. Im, "CHARMM-GUI Membrane Builder for Mixed Bilayers and Its Application to Yeast Membranes," *Biophys. J.*, vol. 97, no. 1, pp. 50–58, 2009.

[54] M. M. Ghahremanpour, S. S. Arab, S. B. Aghazadeh, J. Zhang, and D. van der Spoel, "MemBuilder: a web-based graphical interface to build heterogeneously mixed membrane bilayers for the GROMACS biomolecular simulation program," *Bioinformatics*, vol. 30, no. 3, pp. 439–441, Feb. 2014.

[55] C. Bovigny, G. Tamò, T. Lemmin, N. Maïno, and M. Dal Peraro, "LipidBuilder: A Framework To Build Realistic Models for Biological Membranes," *J. Chem. Inf. Model.*, vol. 55, no. 12, pp. 2491–2499, Dec. 2015.

[56] L. Martínez, R. Andrade, E. G. Birgin, and J. M. Martínez, "PACKMOL: A package for building initial configurations for molecular dynamics simulations," *J. Comput. Chem.*, vol. 30, no. 13, pp. 2157–2164, Oct. 2009.

[57] T. A. Wassenaar, H. I. Ingólfsson, R. A. Böckmann, D. P. Tieleman, and S. J. Marrink, "Computational Lipidomics with insane: A Versatile Tool for Generating Custom Membranes for Molecular Simulations," *J. Chem. Theory Comput.*, vol. 11, no. 5, pp. 2144–2155, May 2015.

[58] J. Lee *et al.*, "CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field," *J. Chem. Theory Comput.*, vol. 12, no. 1, pp. 405–413, Jan. 2016.

[59] J. B. Klauda *et al.*, "Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types," *J. Phys. Chem. B*, vol. 114, no. 23, pp. 7830–7843, Jun. 2010.

[60] J. Lee, M. Hitzenberger, M. Rieger, N. R. Kern, M. Zacharias, and W. Im, "CHARMM-GUI supports the Amber force fields," *J. Chem. Phys.*, vol. 153, no. 3, p. 35103, Jul. 2020.

[61] O. Edholm, O. Berger, and F. Jähnig, "Structure and Fluctuations of Bacteriorhodopsin in the Purple Membrane: A Molecular Dynamics Study," *J. Mol. Biol.*, vol. 250, no. 1, pp. 94–111, 1995.

[62] T. B. Woolf and B. Roux, "Molecular dynamics simulation of the gramicidin channel in a phospholipid bilayer," *Proc. Natl. Acad. Sci.*, vol. 91, no. 24, pp. 11631 LP – 11635, Nov. 1994.

[63] C. Kandt, W. L. Ash, and D. P. Tieleman, "Setting up and running molecular dynamics simulations of membrane proteins.," *Methods*, vol. 41, no. 4, pp. 475–488, Apr. 2007.

[64] M. G. Wolf, M. Hoefling, C. Aponte-Santamaría, H. Grubmüller, and G. Groenhof, "g_membed: Efficient insertion of a membrane protein into an equilibrated lipid bilayer with minimal perturbation.," *J. Comput. Chem.*, vol. 31, no. 11, pp. 2169–2174, Aug. 2010.

[65] J. Lemkul, "From Proteins to Perturbed Hamiltonians: A Suite of Tutorials for the GROMACS-2018 Molecular Simulation Package [Article v1.0]," *Living J. Comput. Mol. Sci.*, vol. 1, no. 1, p. 5068, 2018.

[66] 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, pp. e880–e880, Sep. 2007.

[67] L. Shen, D. Bassolino, and T. Stouch, "Transmembrane helix structure, dynamics, and interactions: multi-nanosecond molecular dynamics simulations.," *Biophys. J.*, vol. 73, no. 1, pp. 3–20, Jul. 1997.

[68] R. Staritzbichler, C. Anselmi, L. R. Forrest, and J. D. Faraldo-Gómez, "GRIFFIN: A versatile methodology for optimization of protein-lipid interfaces for membrane protein simulations.," *J. Chem. Theory Comput.*, vol. 7, no. 4, pp. 1167–1176, Apr. 2011.

[69] M. U. Rahman, A. U. Rehman, H. Liu, and H. Chen, "Comparison and Evaluation of Force Fields for Intrinsically Disordered Proteins," 2020.

[70] O. Guvench and A. D. MacKerell, "Comparison of protein force fields for molecular dynamics simulations," *Methods Mol. Biol.*, vol. 443, pp. 63–88, 2008.

[71] A. P. Lyubartsev and A. L. Rabinovich, "Force Field Development for Lipid Membrane Simulations," *Biochim. Biophys. Acta - Biomembr.*, vol. 1858, no. 10, pp. 2483–2497, 2016.

[72] K. Lindorff-Larsen *et al.*, "Improved side-chain torsion potentials for the Amber ff99SB protein force field," *Proteins Struct. Funct. Bioinforma.*, vol. 78, no. 8, pp. 1950–1958, Jun. 2010.

[73] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, and C. Simmerling, "ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB," *J. Chem. Theory Comput.*, vol. 11, no. 8, pp. 3696–3713, Aug. 2015.

[74] Å. A. Skjevik, B. D. Madej, R. C. Walker, and K. Teigen, "LIPID11: a modular framework for lipid simulations using amber.," *J. Phys. Chem. B*, vol. 116, no. 36, pp. 11124–11136, Sep. 2012.

[75] C. J. Dickson *et al.*, "Lipid14: The Amber Lipid Force Field," *J. Chem. Theory Comput.*, vol. 10, no. 2, pp. 865–879, Feb. 2014.

[76] J. P. M. Jämbeck and A. P. Lyubartsev, "Derivation and Systematic Validation of a Refined All-Atom Force Field for Phosphatidylcholine Lipids," *J. Phys. Chem. B*, vol. 116, no. 10, pp. 3164–3179, Mar. 2012.

[77] J. P. M. Jämbeck and A. P. Lyubartsev, "An Extension and Further Validation of an All-Atomistic Force Field for Biological Membranes," *J. Chem. Theory Comput.*, vol. 8, no. 8, pp. 2938–2948, Aug. 2012.

[78] J. P. M. Jämbeck and A. P. Lyubartsev, "Another Piece of the Membrane Puzzle: Extending Slipids Further," *J. Chem. Theory Comput.*, vol. 9, no. 1, pp. 774–784, Jan. 2013.

[79] I. Ermilova and A. P. Lyubartsev, "Extension of the Slipids Force Field to Polyunsaturated Lipids," *J. Phys. Chem. B*, vol. 120, no. 50, pp. 12826–12842, Dec. 2016.

[80] F. Grote and A. P. Lyubartsev, "Optimization of Slipids Force Field Parameters Describing Headgroups of Phospholipids," *J. Phys. Chem. B*, vol. 124, no. 40, pp. 8784–8793, Oct. 2020.

[81] C. J. Dickson, L. Rosso, R. M. Betz, R. C. Walker, and I. R. Gould, "GAFFlipid: a General Amber Force Field for the accurate molecular dynamics simulation of phospholipid," *Soft Matter*, vol. 8, no. 37, pp. 9617–9627, 2012.

[82] B. D. Madej, I. R. Gould, and R. C. Walker, "A Parameterization of Cholesterol for Mixed Lipid Bilayer Simulation within the Amber Lipid14 Force Field," *J. Phys. Chem. B*, vol. 119, no. 38, pp. 12424–12435, Sep. 2015.

[83] K. N. Kirschner *et al.*, “GLYCAM06: A generalizable biomolecular force field. Carbohydrates,” *J. Comput. Chem.*, vol. 29, no. 4, pp. 622–655, Mar. 2008.

[84] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, “Development and testing of a general amber force field,” *J. Comput. Chem.*, vol. 25, no. 9, pp. 1157–1174, Jul. 2004.

[85] J. Wang, W. Wang, P. A. Kollman, and D. A. Case, “Automatic atom type and bond type perception in molecular mechanical calculations,” *J. Mol. Graph. Model.*, vol. 25, no. 2, pp. 247–260, 2006.

[86] L. Huang and B. Roux, “Automated Force Field Parameterization for Nonpolarizable and Polarizable Atomic Models Based on Ab Initio Target Data,” *J. Chem. Theory Comput.*, vol. 9, no. 8, pp. 3543–3556, Aug. 2013.

[87] R. M. Betz and R. C. Walker, “Paramfit: Automated optimization of force field parameters for molecular dynamics simulations,” *J. Comput. Chem.*, vol. 36, no. 2, pp. 79–87, Jan. 2015.

[88] J. Wang, W. Wang, P. A. Kollman, and D. A. Case, “Antechamber: an accessory software package for molecular mechanical calculations,” *J. Am. Chem. Soc*, vol. 222, p. U403, 2001.

[89] D. van der Spoel, P. J. van Maaren, and C. Caleman, “GROMACS molecule & liquid database,” *Bioinformatics*, vol. 28, no. 5, pp. 752–753, Mar. 2012.

[90] R. M. Venable *et al.*, “CHARMM All-Atom Additive Force Field for Sphingomyelin: Elucidation of Hydrogen Bonding and of Positive Curvature,” *Biophys. J.*, vol. 107, no. 1, pp. 134–145, 2014.

[91] R. B. Best *et al.*, “Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ, ψ and Side-Chain $\chi1$ and $\chi2$ Dihedral Angles,” *J. Chem. Theory Comput.*, vol. 8, no. 9, pp. 3257–3273, Sep. 2012.

[92] O. Guvench *et al.*, “CHARMM Additive All-Atom Force Field for Carbohydrate Derivatives and Its Utility in Polysaccharide and Carbohydrate–Protein Modeling,” *J. Chem. Theory Comput.*, vol. 7, no. 10, pp. 3162–3180, Oct. 2011.

[93] K. Hart, N. Foloppe, C. M. Baker, E. J. Denning, L. Nilsson, and A. D. MacKerell, “Optimization of the CHARMM Additive Force Field for DNA: Improved Treatment of the BI/BII Conformational Equilibrium,” *J. Chem. Theory Comput.*, vol. 8, no. 1, pp. 348–362, Jan. 2012.

[94] J. D. Yesselman, D. J. Price, J. L. Knight, and C. L. 3rd Brooks, “MATCH: an atom-typing toolset for molecular mechanics force fields.,” *J. Comput. Chem.*, vol. 33, no. 2, pp. 189–202, Jan. 2012.

[95] K. Vanommeslaeghe *et al.*, “CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields.,” *J. Comput. Chem.*, vol. 31, no. 4, pp. 671–690, Mar. 2010.

[96] K. Vanommeslaeghe, E. P. Raman, and A. D. J. MacKerell, “Automation of the CHARMM General Force Field (CGenFF) II: assignment of bonded parameters and partial atomic charges.,” *J. Chem. Inf. Model.*, vol. 52, no. 12, pp. 3155–3168, Dec. 2012.

[97] K. Vanommeslaeghe and A. D. J. MacKerell, “Automation of the CHARMM General Force Field (CGenFF) I: bond perception and atom typing.,” *J. Chem. Inf. Model.*, vol. 52, no. 12, pp. 3144–3154, Dec. 2012.

[98] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, “Comparison of simple potential functions for simulating liquid water,” *J. Chem. Phys.*, vol. 79, no. 2, pp. 926–935, 1983.

[99] W. L. Jorgensen and J. D. Madura, “Temperature and size dependence for Monte Carlo simulations of TIP4P water,” *Mol. Phys.*, vol. 56, no. 6, pp. 1381–1392, 1985.

[100] M. W. Mahoney and W. L. Jorgensen, “A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions,” *J. Chem. Phys.*, vol. 112, no. 20, pp. 8910–8922, 2000.

[101] S. M. Tschampel, M. R. Kennerty, and R. J. Woods, “TIP5P-Consistent Treatment of Electrostatics for Biomolecular Simulations.,” *J. Chem. Theory Comput.*, vol. 3, no. 5, pp. 1721–1733, Sep. 2007.

[102] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma, “The Missing Term in Effective Pair Potentialst,” 2001.

[103] P. G. Kusalik and I. M. Svishchev, “The spatial structure in liquid water.,” *Science*, vol. 265, no. 5176, pp. 1219–1221, Aug. 1994.

[104] A. Glättli, X. Daura, and W. F. Van Gunsteren, “A novel approach for designing simple point charge models for liquid water with three interaction sites,” *J. Comput. Chem.*, vol. 24, no. 9, pp. 1087–1096, 2003.

[105] D. J. Price and C. L. 3rd Brooks, “A modified TIP3P water potential for simulation with Ewald summation.,” *J. Chem. Phys.*, vol. 121, no. 20, pp. 10096–10103, Nov. 2004.

[106] H. W. Horn, W. C. Swope, and J. W. Pitera, “Characterization of the TIP4P-Ew water model: vapor pressure and boiling point.,” *J. Chem. Phys.*, vol. 123, no. 19, p. 194504, Nov. 2005.

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

[108] C. Vega, J. L. F. Abascal, M. M. Conde, and J. L. Aragones, “What ice can teach us about water interactions: a critical comparison of the performance of different water models,” *Faraday Discuss.*, vol. 141, no. 0, pp. 251–276, 2009.

[109] J. Zielkiewicz, “Structural properties of water: comparison of the SPC, SPCE, TIP4P, and TIP5P models of water.,” *J. Chem. Phys.*, vol. 123, no. 10, p. 104501, Sep. 2005.

[110] M. A. González and J. L. F. Abascal, “The shear viscosity of rigid water models,” *J. Chem. Phys.*, vol. 132, no. 9, p. 96101, 2010.

[111] M. Deserno, “Mesoscopic Membrane Physics: Concepts, Simulations, and Selected Applications,” *Macromol. Rapid Commun.*, vol. 30, no. 9–10, pp. 752–771, May 2009.

[112] A. Davtyan, M. Simunovic, and G. A. Voth, “The mesoscopic membrane with proteins (MesM-P) model,” *J. Chem. Phys.*, vol. 147, no. 4, p. 44101, Jul. 2017.

[113] M. Chavent, A. L. Duncan, and M. S. P. Sansom, “Molecular dynamics simulations of membrane proteins and their interactions: from nanoscale to mesoscale,” *Curr. Opin. Struct. Biol.*, vol. 40, pp. 8–16, 2016.

[114] L. Monticelli, S. K. Kandasamy, X. Periole, R. G. Larson, D. P. Tieleman, and S.-J. Marrink, “The MARTINI Coarse-Grained Force Field: Extension to Proteins.,” *J. Chem. Theory Comput.*, vol. 4, no. 5, pp. 819–834, May 2008.

[115] C. Czaplewski, A. Karczyńska, A. K. Sieradzan, and A. Liwo, “UNRES server for physics-based coarse-grained simulations and prediction of protein structure, dynamics and thermodynamics,” *Nucleic Acids Res.*, vol. 46, no. W1, pp. W304–W309, Jul. 2018.

[116] A. Kolinski, “Protein modeling and structure prediction with a reduced representation.,” *Acta Biochim. Pol.*, vol. 51, no. 2, pp. 349–371, 2004.

[117] S. M. Gopal, S. Mukherjee, Y.-M. Cheng, and M. Feig, “PRIMO/PRIMONA: a coarse-grained model for proteins and nucleic acids that preserves near-atomistic accuracy.,” *Proteins*, vol. 78, no. 5, pp. 1266–1281, Apr. 2010.

[118] P. Kar, S. M. Gopal, Y.-M. Cheng, A. Predeus, and M. Feig, “PRIMO: A Transferable Coarse-grained Force Field for Proteins,” *J. Chem. Theory Comput.*, vol. 9, no. 8, pp. 3769–3788, Aug. 2013.

[119] P. Derreumaux, “From polypeptide sequences to structures using Monte Carlo simulations and an optimized potential,” *J. Chem. Phys.*, vol. 111, no. 5, pp. 2301–2310, Jul. 1999.

[120] M. Cheon, I. Chang, and C. K. Hall, “Extending the PRIME model for protein aggregation to all 20 amino acids.,” *Proteins*, vol. 78, no. 14, pp. 2950–2960, Nov. 2010.

[121] X. Periole and S.-J. Marrink, “The Martini coarse-grained force field.,” *Methods Mol. Biol.*, vol. 924, pp. 533–565, 2013.

[122] D. H. de Jong *et al.*, “Improved Parameters for the Martini Coarse-Grained Protein Force Field.,” *J. Chem. Theory Comput.*, vol. 9, no. 1, pp. 687–697, Jan. 2013.

[123] C. A. López, A. J. Rzepiela, A. H. de Vries, L. Dijkhuizen, P. H. Hünenberger, and S. J. Marrink, “Martini Coarse-Grained Force Field: Extension to Carbohydrates.,” *J. Chem. Theory Comput.*, vol. 5, no. 12, pp. 3195–3210, Dec. 2009.

[124] J. J. Uusitalo, H. I. Ingólfsson, P. Akhshi, D. P. Tieleman, and S. J. Marrink, “Martini Coarse-Grained Force Field: Extension to DNA,” *J. Chem. Theory Comput.*, vol. 11, no. 8, pp. 3932–3945, Aug. 2015.

[125] S. O. Yesylevskyy, L. V. Schäfer, D. Sengupta, and S. J. Marrink, “Polarizable water model for the coarse-grained MARTINI force field,” *PLoS Comput. Biol.*, vol. 6, no. 6, pp. 1–17, 2010.

[126] M. N. Rosenbluth and A. W. Rosenbluth, “Further results on Monte Carlo equations of state,” *J. Chem. Phys.*, vol. 22, no. 5, pp. 881–884, 1954.

[127] W. W. Wood and J. D. Jacobson, “Preliminary results from a recalculation of the Monte Carlo equation of state of hard spheres,” *J. Chem. Phys.*, vol. 27, no. 5, pp. 1207–1208, 1957.

[128] B. J. Alder and T. E. Wainwright, “Phase transition for a hard sphere system,” *J. Chem. Phys.*, vol. 27, no. 5, pp. 1208–1209, 1957.

[129] P. Boisseau and M. Lahmani, Eds., *Nanoscience : Nanobiotechnology and Nanobiology*, 1st ed. Springer-Verlag Berlin Heidelberg, 2009.

[130] J. A. McCammon, B. R. Gelin, and M. Karplus, “Dynamics of folded proteins,” *Nature*, vol. 267, no. 5612, pp. 585–590, 1977.

[131] L. Verlet, “Computer ‘Experiments’ on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules,” *Phys. Rev.*, vol. 159, no. 1, pp. 98–103, Jul. 1967.

[132] R. W. Hockney, “The potential calculation and some applications,” 1970.

[133] D. Beeman, “Some multistep methods for use in molecular dynamics calculations,” *J. Comput. Phys.*, vol. 20, no. 2, pp. 130–139, 1976.

[134] R. A. Willoughby, “Numerical Initial Value Problems in Ordinary Differential Equations (C. William Gear),” *SIAM Rev.*, vol. 15, no. 3, pp. 676–678, 1973.

[135] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, “A smooth particle mesh Ewald method,” *J. Chem. Phys.*, vol. 103, no. 19, pp. 8577–8593, Nov. 1995.

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

[137]S. Nosé, “A molecular dynamics method for simulations in the canonical ensemble,” *Mol. Phys.*, vol. 52, no. 2, pp. 255–268, Jun. 1984.

[138] W. G. Hoover, “Canonical dynamics: Equilibrium phase-space distributions,” *Phys. Rev. A*, vol. 31, no. 3, pp. 1695–1697, 1985.

[139] M. Parrinello and A. Rahman, “Polymorphic transitions in single crystals: A new molecular dynamics method,” *J. Appl. Phys.*, vol. 52, no. 12, pp. 7182–7190, Dec. 1981.

[140] S. Nosé and M. L. Klein, “Constant pressure molecular dynamics for molecular systems,” *Mol. Phys.*, vol. 50, no. 5, pp. 1055–1076, Dec. 1983.

[141] G. J. Martyna, M. E. Tuckerman, D. J. Tobias, and M. L. Klein, “Explicit reversible integrators for extended systems dynamics,” *Mol. Phys.*, vol. 87, no. 5, pp. 1117–1157, Apr. 1996.

[142] T. Ackermann, “C. L. Brooks III, M. Karplus, B. M. Pettitt. Proteins: A Theoretical Perspective of Dynamics, Structure and Thermodynamics, Volume LXXI, in: Advances in Chemical Physics, John Wiley & Sons, New York 1988. 259 Seiten,

Preis: US $ 65.25," *Berichte der Bunsengesellschaft für Phys. Chemie*, vol. 94, no. 1, p. 96, Jan. 1990.

[143] M. Leitgeb, C. Schröder, and S. Boresch, "Alchemical free energy calculations and multiple conformational substates.," *J. Chem. Phys.*, vol. 122, no. 8, p. 84109, Feb. 2005.

[144] J. P. Valleau and S. G. Whittington, "Monte Carlo in Statistical mechanics: Choosing between alternative transition matrices," *J. Comput. Phys.*, vol. 24, no. 2, pp. 150–157, 1977.

[145] Y. Sugita and Y. Okamoto, "Replica-exchange molecular dynamics method for protein folding," *Chem. Phys. Lett.*, vol. 314, no. 1–2, pp. 141–151, 1999.

[146] B. Barz, J. Loschwitz, and B. Strodel, "Large-scale, dynamin-like motions of the human guanylate binding protein 1 revealed by multi-resolution simulations," *bioRxiv*, no. 5030732, 2019.

[147] "Alphafold highlighted-research." [Online]. Available: https://www.deepmind.com/research/highlighted-research/alphafold.

[148] "AlphaFold Protein Structure Database." [Online]. Available: https://alphafold.com/.

[149] E. Rosta and G. Hummer, "Error and efficiency of replica exchange molecular dynamics simulations.," *J. Chem. Phys.*, vol. 131, no. 16, p. 165102, Oct. 2009.

[150] A. Khot, S. B. Shiring, and B. M. Savoie, "Evidence of information limitations in coarse-grained models," *J. Chem. Phys.*, vol. 151, no. 24, p. 244105, Dec. 2019.