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**).

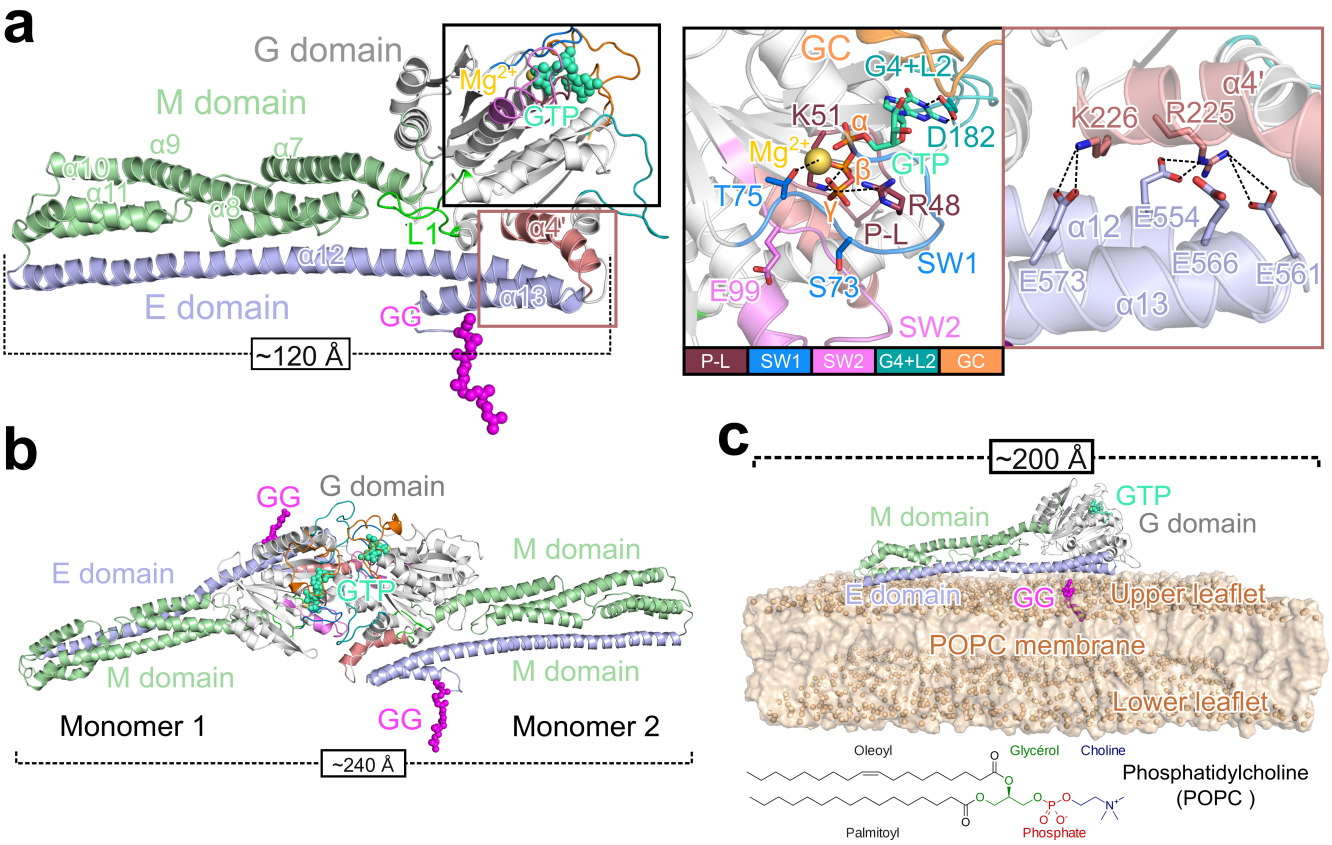

Figure 1. Model systems of mGBP2 investigated in this study. (a) The homology model of the entire mGBP2 molecule contains three different domains: (i) the G domain (gray cartoon) with GTP (green spheres) and Mg^{2+} (orange sphere); (ii) the M domain (green cartoon); (iii) the E domain (blue cartoon), 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 ($mGBP2_{apo}$), the protein with GTP being bound ($mGBP2_{GTP}$), and the protein that has GTP bound and the geranylgeranyl anchor attached, which we denote as the *holo*-state ($mGBP2_{holo}$). 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 α10 (residues 420–446) and α12 (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 $mGBP2_{GTP}$, 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].

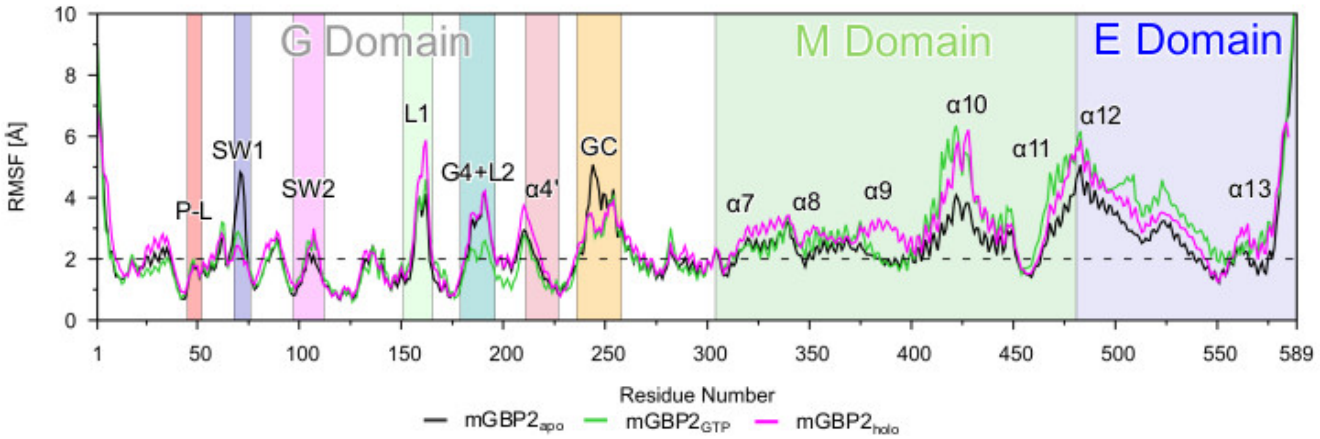

Figure 2. The root mean square fluctuations per residue of monomeric mGBP2 obtained from HREMD simulations of $mGBP2_{apo}$ (black), $mGBP2_{GTP}$ (green), and $mGBP2_{holo}$ (magenta), as sampled in the target replica. All important motifs, loops, and helices are labeled and a background color added 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.

	$mGBP2_{apo}$			$mGBP2_{GTP}$			$mGBP2_{holo}$		
(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.

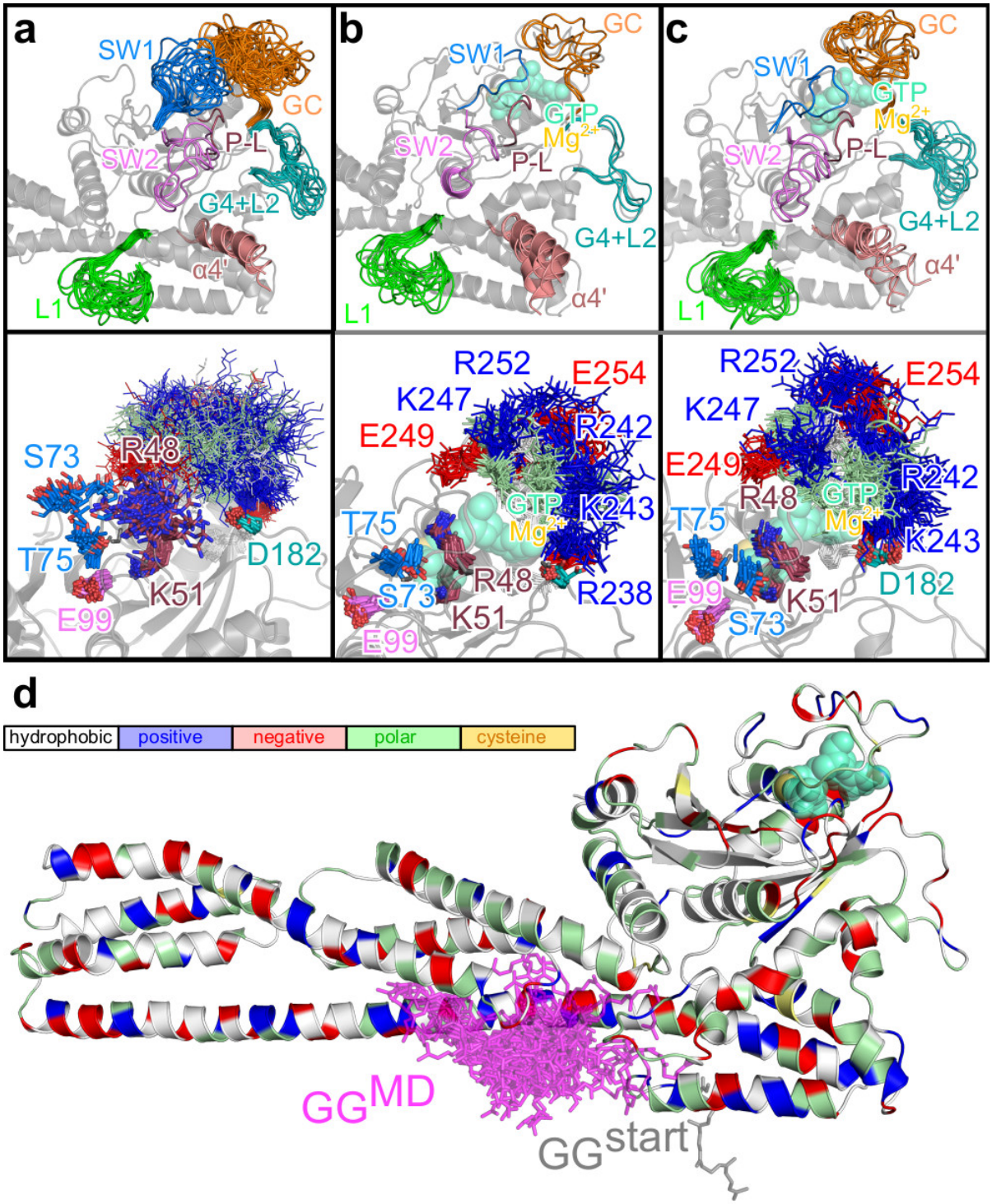

Figure 3. Conformational clusters of the G domain loops of (a) $mGBP2_{apo}$, (b) $mGBP2_{GTP}$, and (c) $mGBP2_{holo}$ as well as (d) for the geranylgeranyl anchor. (a-c) In the upper panels, each central cluster conformation is shown, using different colors for the different loops as indicated in the 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 $mGBP2_{holo}$, most loops are slightly more flexible than those in $mGBP2_{GTP}$ (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 $mGBP2_{holo}$ 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 $mGBP2_{GTP}$ and $mGBP2_{holo}$ 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 $mGBP2_{GTP}$ and $mGBP2_{holo}$ is observed for the G4 motif/L2 loop. Only in the case of $mGBP2_{GTP}$, 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 $mGBP2_{holo}$ and $mGBP2_{apo}$, 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 $mGBP2_{holo}$, 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 $mGBP2_{holo}$, loop L1 is the most flexible, as the number of loop L1 clusters is 25 in $mGBP2_{holo}$, 23 in $mGBP2_{apo}$, and 17 in $mGBP2_{GTP}$. Moreover in $mGBP2_{holo}$, 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 $mGBP2_{apo}$ and $mGBP2_{GTP}$.

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 $mGBP2_{GTP}$ and $mGBP2_{holo}$ 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 $mGBP2_{apo}$ to 3 and 8 in $mGBP2_{GTP}$ and $mGBP2_{holo}$, 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, $mGBP2_{GTP}$ seems the most stable monomeric protein state compared to the two other states, $mGBP2_{apo}$ and $mGBP2_{holo}$.

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 $mGBP2_{holo}$, 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 $mGBP2_{holo}$ (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., $mGBP2_{apo}$, $mGBP2_{GTP}$, and $mGBP2_{holo}$, 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 $mGBP2_{apo}$, the value range is between -12.1 nm^2 ($PC1_{min}$) and 8.5 nm^2 ($PC1_{max}$) for PC1 as well as -8.1 nm^2 ($PC2_{min}$) and 10.9 nm^2 ($PC2_{max}$) for PC2, which serve as reference. Upon GTP binding, the deflection along PC1 increased, with $PC1_{max}$ reaching 12.7 nm^2 ($\Delta PC1_{max} = 4.2$ nm^2), whereas all other values lie in the same area as for $mGBP2_{apo}$. When the geranylgeranyl anchor is further added to the protein, this main motion became even more pronounced: the $mGBP2_{holo}$ has an almost two-fold higher $PC1_{max}$ value of 16.9 nm^2 ($\Delta PC1_{max} = 8.4$ nm^2) compared to $mGBP2_{apo}$. On the other hand, the motion along PC2 is somewhat reduced in this protein, as $PC2_{max}$ only reaches a value of 9.4 nm^2 ($\Delta PC2_{max} = -1.5$ 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 $mGBP2_{GTP}$ and $mGBP2_{holo}$ compared to the $mGBP2_{apo}$ (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 $mGBP2_{apo}$ vs. > 42 Å for $mGBP2_{GTP}$ and $mGBP2_{holo}$).

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 $mGBP2_{apo}$ 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

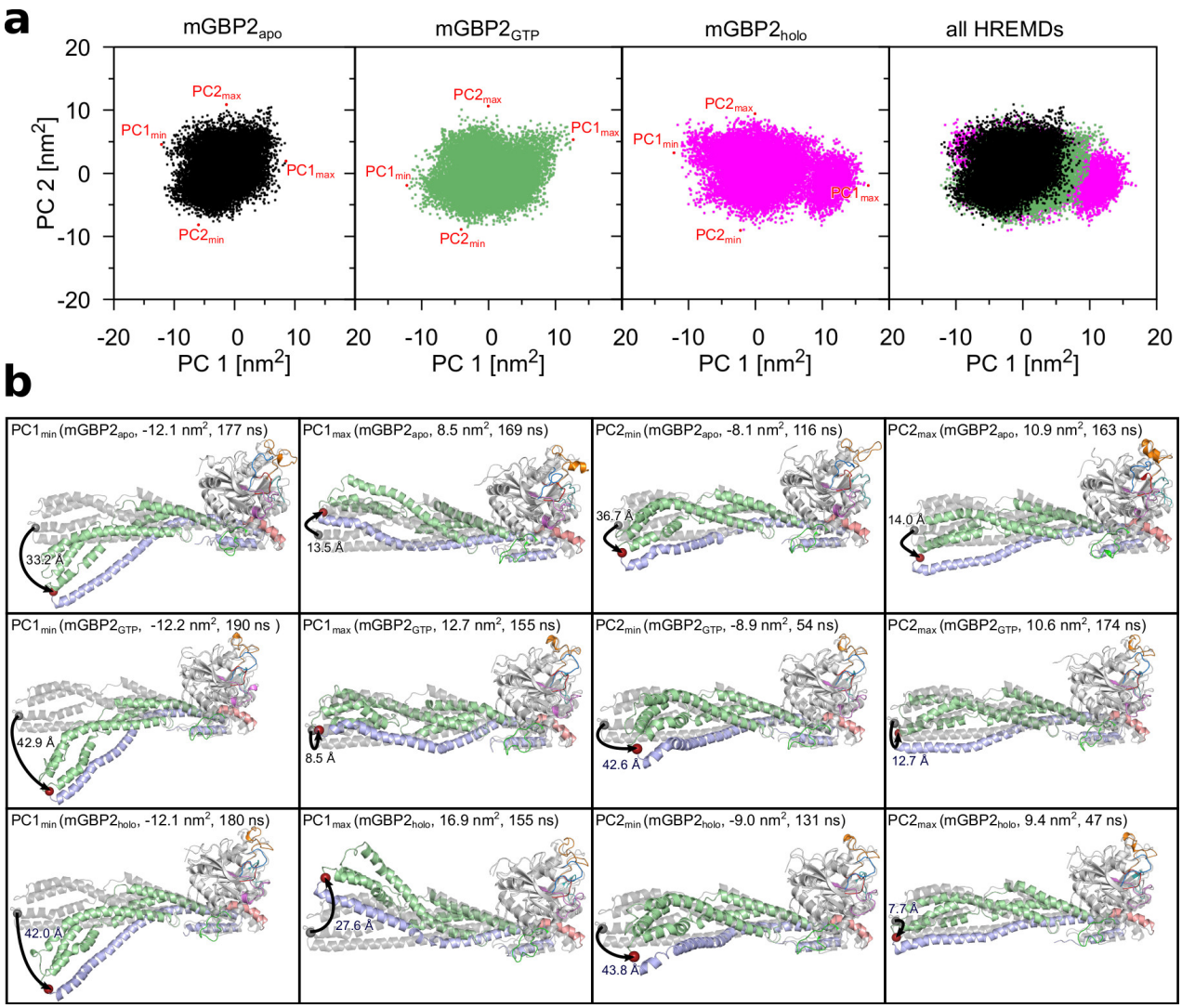

Figure 4. Main motions of the M and E domain in monomeric mGBP2 identified by PCA of the target replica of the HREMD simulations. (a) The projection of the conformations collected in the target replica onto the first two principal components (PC1 and PC2). The results for GTP-free mGBP2 (black), mGBP2 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 $mGBP2_{apo}$ (top), $mGBP2_{GTP}$ (middle) and $mGBP2_{holo}$ (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 $mGBP2_{GTP}$, 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 $mGBP2_{apo}$, 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 $\alpha 13$. With the hinge motion taking place into the opposite direction, the larger PC1 space as seen for $mGBP2_{GTP}$ compared to $mGBP2_{apo}$ can be explained. Moreover, in $mGBP2_{GTP}$ 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 $\alpha 9$ and $\alpha 12$ 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 $mGBP2_{apo}$ (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.

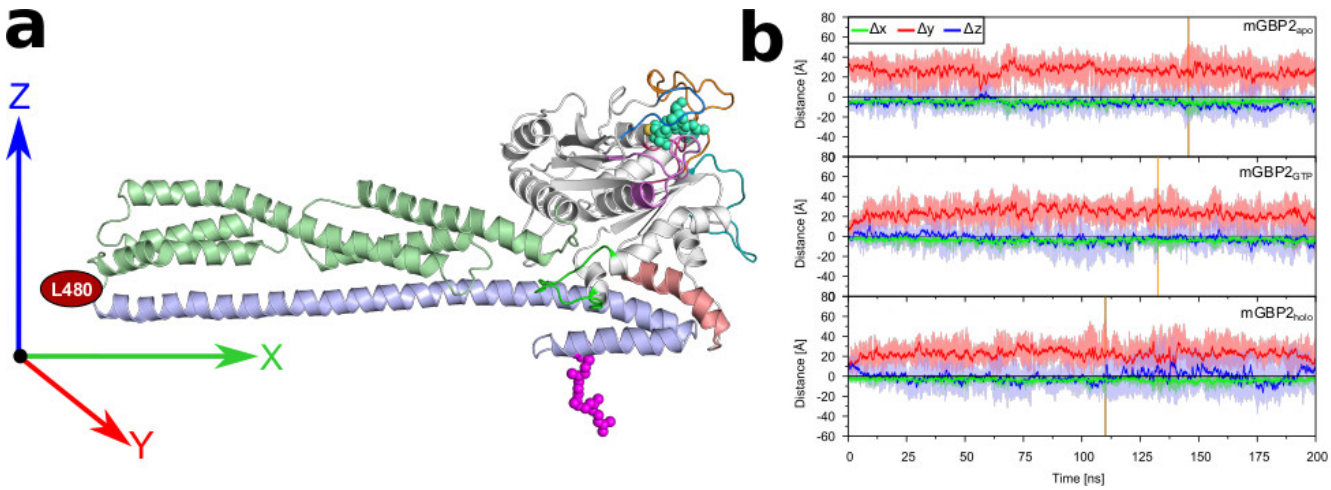

Figure 5. Motions of residue L480 in the HREMD simulations of monomeric mGBP2. **(a)** Reference conformation of the mGBP2 is shown for the calculation of the coordinate distances difference (Δx, Δy, Δz) of L480. **(b)** The evolution of these distances (Δx: green, Δy: red, Δz: blue) is shown for $mGBP2_{apo}$ (top), $mGBP2_{GTP}$ (middle), and $mGBP2_{holo}$ (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 $mGBP2_{holo}$ 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 $mGBP2_{holo}$ 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, $mGBP2_{holo}$ 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

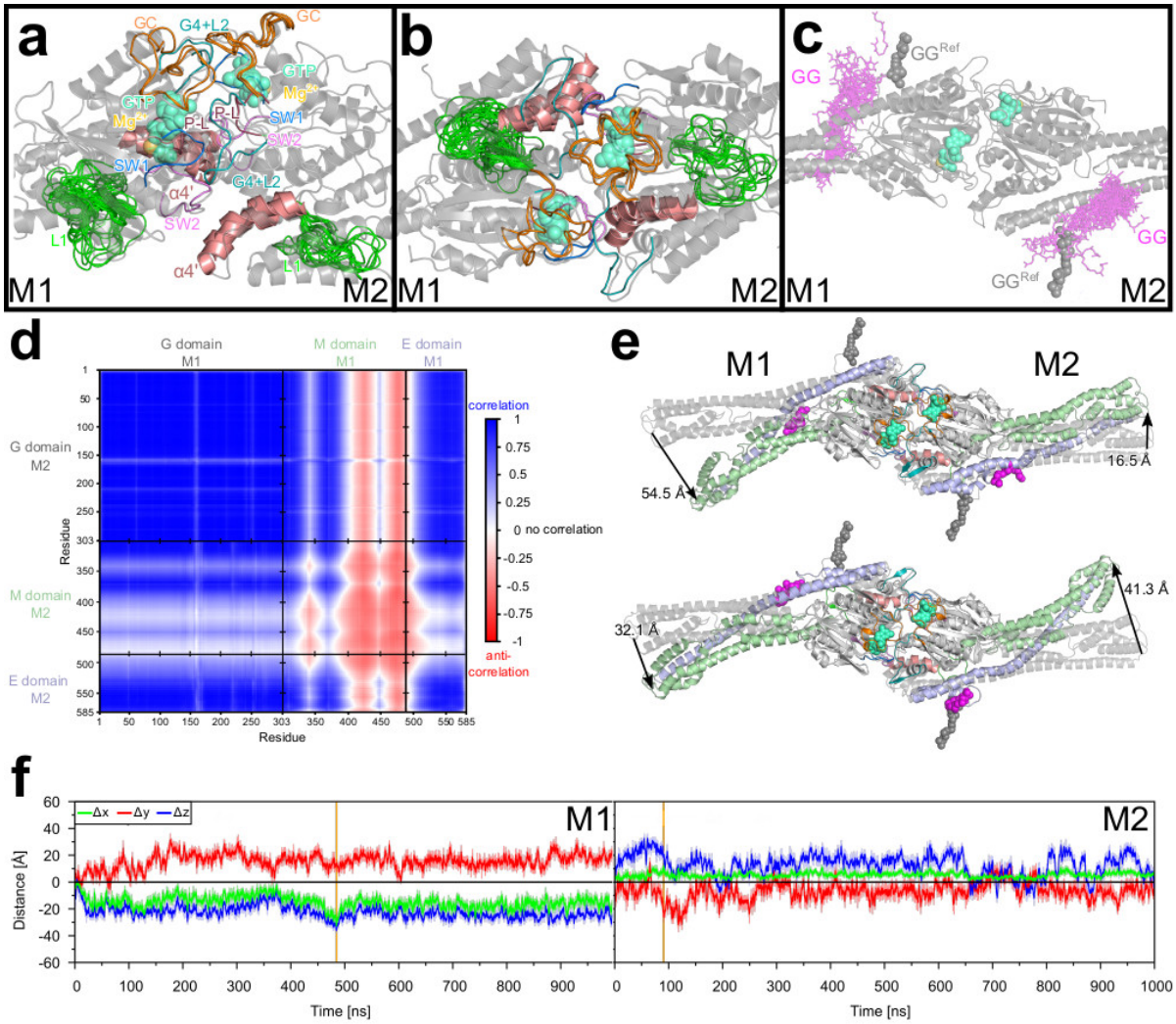

Figure 6. Molecular motions in the mGBP2$_{holo}$ dimer obtained from a 1 μs MD simulation. Conformational clusters of the G-domain motifs, the L1, and the guanine cap in the side view **(a)** and front view **(b)**. In **(c)** the clusters of the geranylgeranyl lipid anchor (GG) are shown. The same coloring scheme as in Fig. 3 was used. **(d)** The dynamic cross-correlation between the motions of the C_α 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 $\text{mGBP2}^{\text{noGTP}}_{\text{mem}}$, while $\text{mGBP2}^{\text{GTP}}_{\text{mem}}$ is used for mGBP2 featuring both bound GTP and the geranylgeranyl anchor. We are aware that the $\text{mGBP2}^{\text{noGTP}}_{\text{mem}}$ system does not reflect reality as GTP is needed for mGBP2 to bind to a membrane[12]. However, by comparing $\text{mGBP2}^{\text{noGTP}}_{\text{mem}}$ with $\text{mGBP2}^{\text{GTP}}_{\text{mem}}$ 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 $\text{mGBP2}^{\text{noGTP}}_{\text{mem}}$ system, with 34, 62, and 101 clusters, respectively. These numbers reduce to 4, 11, and 5 clusters upon GTP binding in the $\text{mGBP2}^{\text{GTP}}_{\text{mem}}$, which are similar, albeit somewhat higher, as in the monomeric $\text{mGBP2}_{\text{holo}}$ 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 $\text{mGBP2}^{\text{noGTP}}_{\text{mem}}$ and 59 Å for $\text{mGBP2}^{\text{GTP}}_{\text{mem}}$, and is around 25 Å on average in both systems. In the 1 μs MD simulation of $\text{mGBP2}_{\text{holo}}$ 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 $\text{mGBP2}^{\text{noGTP}}_{\text{mem}}$ and in (Supplementary Fig. S6**c**) for $\text{mGBP2}^{\text{GTP}}_{\text{mem}}$. A statistical analysis is summarized in Supplementary Tab. S5. The first observation is that $\text{mGBP2}^{\text{noGTP}}_{\text{mem}}$ and $\text{mGBP2}^{\text{GTP}}_{\text{mem}}$ 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 $\text{mGBP2}^{\text{noGTP}}_{\text{mem}}$ and $\text{mGBP2}^{\text{GTP}}_{\text{mem}}$ relative to the membrane suggests that GTP has no direct effect on the mGBP2-membrane interactions.

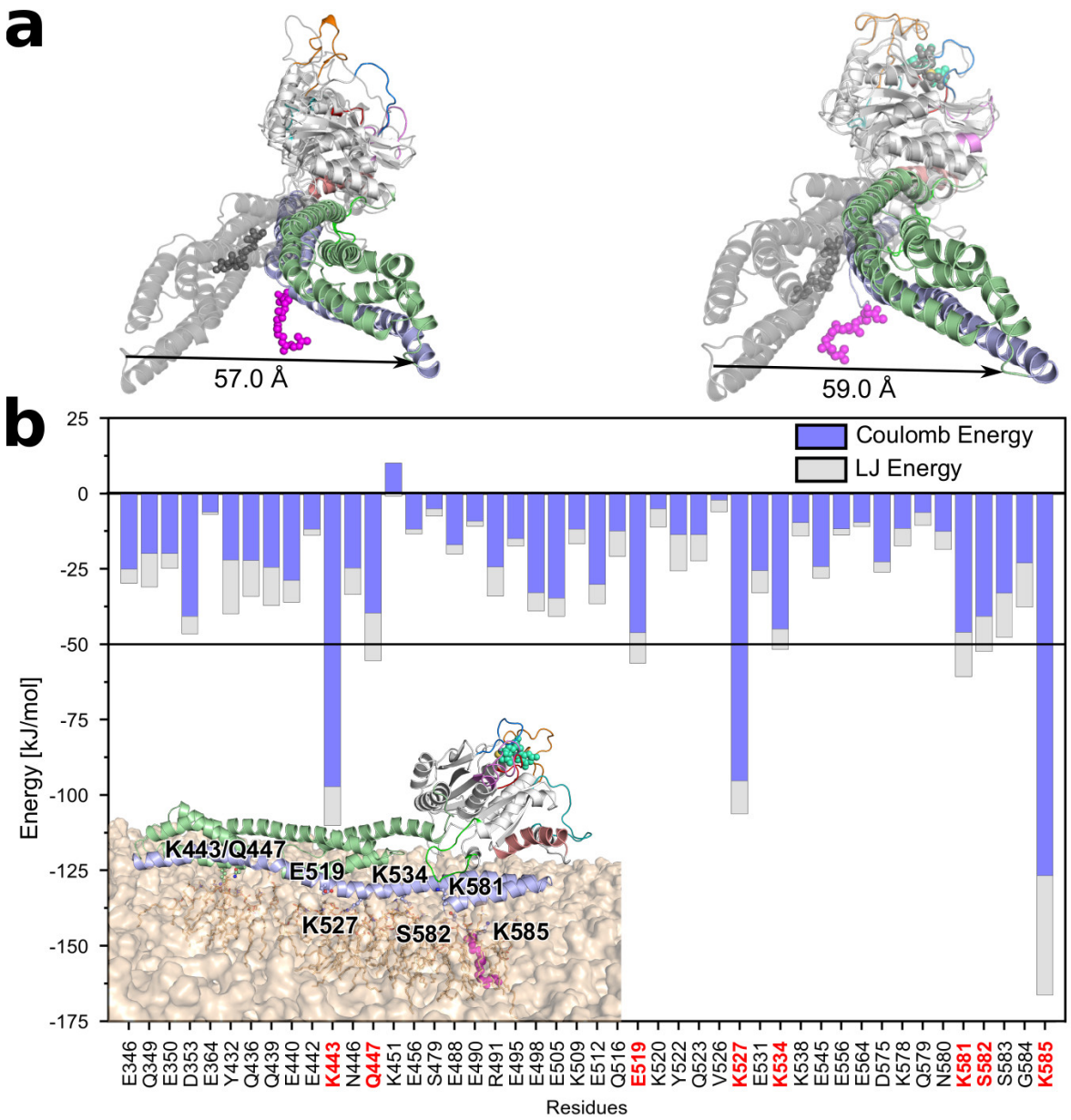

Figure 7. Hinge motion and energies of the strongest interactions between mGBP2 residues and the POPC membrane. **(a)** The conformation with the highest distance to the start structure (L480–L480 for E domain distance, in gray) are shown for $\text{mGBP2}^{\text{noGTP}}_{\text{mem}}$ (left) and $\text{mGBP2}^{\text{GTP}}_{\text{mem}}$ (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 $\text{mGBP2}^{\text{GTP}}_{\text{mem}}$ 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 $\times$Length	Cumulated time
HREMD	$mGBP2_{apo}$	131,493	16×200 ns	3.2 μs
HREMD	$mGBP2_{GTP}$	129,116	16×200 ns	3.2 μs
HREMD	$mGBP2_{holo}$	130,665	16×200 ns	3.2 μs
HREMD	$mGBP2_{holo}$	130,665	40×400 ns	16 μs
MD	$mGBP2_{holo}$	130,665	1×1 μs	1 μs
MD	$mGBP2_{holo}$ dimer	573,673	1×1 μs	1 μs
MD	$mGBP2_{mem}^{noGTP}$	659,946	1×1 μs	1 μs
MD	$mGBP2_{mem}^{GTP}$	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 ($\geq$ 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 $mGBP2_{apo}$, $mGBP2_{GTP}$, and $mGBP2_{holo}$. This step involved the removal of not only the solvent and ions, but also of GTP and Mg^{2+} from both $mGBP2_{GTP}$ and $mGBP2_{holo}$ 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 $mGBP2_{apo}$, 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_α 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.