HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

JÜLICH
FORSCHUNGSZENTRUM

**Master's Thesis**

# Characterizing the murine Guanylate Binding Protein 2 by molecular modeling

**Matthias Henzgen**

**Heinrich-Heine-Universität**

# Characterizing
# the murine Guanylate Binding Protein 2
# by
# molecular modeling

Charakterisierung des murinen Guanylat bindenden Proteins 2
mittels Molekularmodellierung

28th of September, 2015

Düsseldorf

**Supervisor**
Jun.-Prof. Dr. Birgit Strodel
**Examiner**
Jun.-Prof. Dr. Birgit Strodel
**Co-Examiner**
Prof. Dr. Claus A. M. Seidel

# Acknowledgement

I want to thank Junior Professor Birgit Strodel for her kind and elaborate aid and guidance during my Master's Thesis. To Professor Claus A. M. Seidel, whom I would like to thank for his work as co-examiner and together with Thomas Peulen and Mykola Dimura for introducing me to Förster Resonance Energy Transfer and providing me with tools and experimental data to benchmark my insights from molecular modeling of mGBP2.
To the Computational Biochemistry Group: Thanks to my office mates Philipp and Michael for inspiring discussions, recreation time at ''Bochum Total'' and grilling some delicious steaks in Jülich; thanks to Qinghua for sharing his expertise in GROMACS and his knowledge on dummy models for ions; particularly to Dusan for his help in the tricky derivation of force field parameters and for drawing eye catching pictures in pymol; again above all to Philipp for always asking the right questions leading me forward and finding the snags in my codes.
Furthermore I want to thank Elisabeth Kravets from the group of Professor Klaus Pfeffer for providing me with the structural models of the mGBP2 protein. I thank Chunxia Gao from the Eriksson group at the University of Gothenburg for providing me with insights on the derivation of GTP force field parameters.

To Kathrin and my family: I am very grateful for the motivation and encouragement you gave me during my studies, but especially for the cheerful, light-hearted time we experienced together.
Thank you!

Matthias Henzgen

# Declaration

I, Matthias Henzgen, born on the 22nd of April, 1991, declare, that this thesis is the outcome of my own work. I did not use any other source than those mentioned and every citation is identifiable as such.

Düsseldorf, 28th of September, 2015
Matthias Henzgen

# Abstract

The murine Guanylate Binding Protein 2 (mGBP2) is an important effector molecule in host defense against intracellular pathogens. Its role in resistance against $Toxoplasma\ gondii$ ($T.\ gondii$) is subject to extensive experimental studies, as this may serve as a model system for other related parasites such as $Plasmodia$ – a pathogen that causes malaria.

In this work molecular modeling approaches, namely Elastic Network Models (ENM), classical Molecular Dynamics simulations (cMD) and conformational Flooding simulations (fMD) were used to gain insight into the dynamics of the mGBP2 protein. To the best of the author's knowledge, this is the first simulation study of mGBP2. Moreover, the new information regarding structure and dynamics of mGBP2 gained from this explorative study shall complement the existing knowledge previously acquired from experiments.

Molecular dynamics simulations of the $apo$- and $holo$-state of mGBP2, where $holo$-mGBP2 was explicitly described by the binding of mGBP2 to GTP, were performed. The results therefrom indicated that despite the considerable size of mGBP2 the protein can be handled in MD simulations at reasonable computational effort producing stable trajectories. Enhanced simulation methods such as conformational flooding may help to reduce the arising sampling problem. The therefore needed reaction coordinates are perfectly described by normal modes obtained from elastic network models at low computational costs.

Comparing the simulations of $apo$-mGBP2 with present experimental hpFRET data for hGBP1 evinced the consistent sampling of the experimentally described major state of hGBP1. The comparison of simulations of the $apo$- and $holo$-state of mGBP2 demonstrated the conformational changes and the reduction of the flexibility of the G domain caused by GTP binding. A detailed analysis of residue interactions within the active site of the protein revealed that hydrogen bonds can explain the high rigidity of the $holo$ G domain and the stability of the protein-ligand complex. Taken together this led to a model that incorporates all observed effects of GTP binding on structures and dynamics of mGBP2 to possibly explain resulting new dynamics – particularly of the $\alpha$13-helix – that were recently reported for the hGBP1 dimer.

# Contents

**Acknowledgement** . . . II
**Declaration** . . . III
**Abstract** . . . IV
**List of Tables** . . . VII
**List of Figures** . . . VIII
**List of Abbreviations** . . . XII

**1 Introduction** **1**
1.1 Why study the murine Guanylate Binding Protein 2? . . . 1
1.2 State of the art in mGBP2 research – Why apply Molecular Modeling to mGBP2? 1
1.3 The aim of this thesis . . . 2

**2 Background** **3**
2.1 The murine Guanylate Binding Protein 2 . . . 3
2.1.1 Characteristics of mGBP2 . . . 3
2.1.2 Biological Function of mGBP2 . . . 3
2.1.3 Structure of mGBP2 and Hydrolysis Model . . . 4
2.2 Elastic Network Models . . . 7
2.3 Molecular Dynamics Simulations . . . 8
2.3.1 Newton's equation of motion – Leap Frog algorithm . . . 8
2.3.2 Force Fields – Amber ff99SB-ILDN . . . 9
2.3.3 Enhanced Sampling Simulations – Flooding . . . 11

**3 Methods** **13**
3.1 mGBP2 structure models for molecular modeling . . . 13
3.1.1 Nucleotide-free model of mGBP2 . . . 13
3.1.2 GTP bound model of mGBP2 . . . 13
3.1.3 Derivation of GTP force field parameters . . . 14
3.2 Elastic Network Models . . . 14
3.2.1 Software for ENM calculations and employed parameters . . . 14
3.2.2 Verifying the choice of ENM software and ENM parameters . . . 15
3.3 Molecular Dynamics Simulations . . . 15
3.3.1 Free Simulations . . . 15
3.3.2 Enhanced Sampling Simulations – Flooding . . . 16
3.4 Analysis . . . 20
3.4.1 Essential Dynamics Analysis . . . 20
3.4.2 Eigenvector Overlap . . . 22
3.4.3 Root Mean Square Deviation . . . 22
3.4.4 Root Mean Square Fluctuation and B-Factors . . . 22
3.4.5 Distance Measurements . . . 23
3.4.6 Clustering . . . 24

3.4.7 Hydrogen Bonds and Salt Bridges . . . . . 25

**4 Results and Discussion** **26**

4.1 Examining ENM, cMD and fMD results for $apo$-mGBP2 . . . . . 26

4.1.1 cMD and fMD yield stable simulations . . . . . 26

4.1.2 ENM, cMD and fMD predict similar collective, essential dynamics . . . . 26

4.1.3 ENM, cMD and fMD predict similar protein flexibility . . . . . 29

4.1.4 cMD and fMD sample different conformational spaces . . . . . 32

4.1.5 cMD of $apo$-mGBP2 samples major state from FRET experiments . . . . 33

4.1.6 Interim Conclusion . . . . . 40

4.2 Comparing fMD results for the $apo$- and $holo$-mGBP2 . . . . . 40

4.2.1 fMD simulations of $apo$-mGBP2 show stable binding of GTP . . . . . 40

4.2.2 GTP binding stabilizes the G Domain of mGBP2 . . . . . 45

4.2.3 GTP binding causes great conformational changes in mGBP2 . . . . . 46

4.2.4 Hydrogen Bonds stabilize $holo$ G domain and ensure solid binding of GTP 52

**5 Conclusion and Outlook** **59**

**6 Appendix** **61**

6.1 Appendix 1 – Simulation protocols . . . . . 61

6.1.1 Energy Minimization . . . . . 61

6.1.2 Equilibration . . . . . 62

6.1.3 MD production runs . . . . . 65

6.1.4 Flooding input file . . . . . 68

6.2 Appendix 2 – Sequence Alignment of hGBP1 and mGBP2 . . . . . 68

**References** . . . . . XIV

# List of Tables

1 Positions of FRET labels in hGBP1 and corresponding positions in mGBP2. . . 24
2 Eigenvector overlaps between the first two collective modes for cMD, fMD and ENM of nucleotide-free mGBP2. The largest overlap is highlighted in green with ''bold'' writing. . . . . 29
3 Distances between the given FRET pairs (hGBP1 counting of residues) of distinct structures from the cMD simulation of $apo$-mGBP2 are compared to the experimental results from hpFRET measurements of hGBP1.[15] The structure at 45.39 ns was chosen as it shows the greatest deviation from the initial state of the cMD simulation of $apo$-mGBP2 (RMSD of 7.9 Å) to cover the entire sampling. . . . . 35
4 Distances between the given FRET pairs (hGBP1 counting of residues) of distinct structures from the cMD simulation of $apo$-mGBP2 are compared to the experimental results from hpFRET measurements of hGBP1.[15] . . . . . 36

# List of Figures

1 Structure model of mGBP2, which consists of three domains. The G domain is colored in green, the M domain in silver and the E domain in blue. . . . . 4
2 Salt bridges between the G domain and the $\alpha$12/13-helix of $apo$-mGBP2 are visualized, as well as possible changes in the $holo$-state (shown in red). The G domain of $apo$-mGBP2 is colored in green, the M domain in silver and the E domain in blue. . . . . 5
3 The G domain of $apo$-mGBP2 is displayed emphasizing the guanine cap, the P-loop, the switch I and II regions, as well as the $\alpha$4'-helix (in glossy green). The entire protein is shown in the background: the G domain is colored in transparent green, the M domain in transparent silver and the E domain in transparent blue. . . . . 6
4 The plot sketches the principle of conformational flooding. The red Gaussian shaped flooding potential $V_{\mathrm{fl}}$ is added to the system's potential along user defined reaction coordinates suggested to reflect the essential dynamics of the protein. The figure was adapted from Lange et al.[64] . . . . . 12
5 The work-flow of fMD simulations is delineated. . . . . 19
6 The RMSD in Å is plotted against the simulation time in ns for fMD simulations using different flooding strengths $E_{\mathrm{fl}}$ in kJ/mol. . . . . 21
7 The flooding potential $V_{\mathrm{fl}}$ in kJ/mol of an fMD simulation is plotted against the simulation time in ns for a flooding strength of 10 kJ/mol. . . . . 21
8 FRET labeling positions are visualized in the mGBP2 model as red spheres given the residue indices for hGBP1. The G domain of mGBP2 is shown in green, the M domain in silver and the E domain in blue. . . . . 24
9 The potential energy of $apo$-mGBP2 in kJ/mol (a) and the temperature of the system in K (b) are plotted against the time in ns. The cMD simulation data is shown in red, the fMD data in green. . . . . 27
10 The RMSD of $apo$-mGBP2 (in Å) is plotted against the time in ns. The cMD simulation is shown in red and the corresponding fMD simulation in green. . . 28
11 The first two collective modes of $apo$-mGBP2 are shown following the mode order from PCA. The orange arrows mark the main direction of the motion of each mode. The G domain of mGBP2 is colored in green, the middle domain in silver and the E Domain in blue. . . . . 30
12 In the bottom half of the figure the relative B-factors from ENM (red), cMD (green) and fMD (blue) of nucleotide-free mGBP2 are plotted against residue indices. The orange arrows relate the maxima to the corresponding flexible parts of the mGBP2 structure shown in the upper half of the figure. The G domain of mGBP2 is colored in green, the M domain in silver and the E domain in blue. Vertical gray lines separate the three domains in the B-factor plot. . . 31

13 The plot shows the projection of the structures of $apo$-mGBP2 from the cMD (red crosses) and fMD (green crosses) trajectories onto the first two principal components. The eigenvector factors are given in Å. . . . . 32
14 The figures illustrate the two states found for $apo$-hGBP1 from hpFRET experiments. The G domain is shown in blue, the M domain in silver and the E domain in green color. The figures were gratefully obtained from the Seidel group[15]. . . . . 34
15 The plots show the distance distribution of hpFRET pairs during the cMD simulation of $apo$-mGBP2. Distances are measured in Å with the bin width being 0.1 Åwhile the occurrence is given in %. The residue indices are based on the hGBP1 sequence. To obtain the corresponding amino acid in mGBP2, one has to subtract 2 starting from hGBP1 residue 165. . . . . 39
16 The figure visualizes the conformational space of the terminal part of the E domain as sampled in the cMD simulation of nucleotide-free mGBP2. To this end, the terminal E domain conformations from the middle structure of each cluster are overlaid. The E domain is shown in blue, the M domain in silver and the G domain in green. . . . . 39
17 The potential energy of $holo$-mGBP2 in kJ/mol (a) and the temperature of the system in K (b) are plotted against the time in ns. . . . . 41
18 The RMSD in Å is plotted against the time in ns for the fMD simulation of nucleotide-bound mGBP2. . . . . 42
19 Non-bonded interaction energies in kJ/mol are plotted against the time in ns for selected interactions of $holo$-mGBP2 during the fMD simulation. Non-bonded interactions consist of Coulomb interactions (red) and Lennard-Jones interactions (green). . . . . 44
20 The figures show the distributions for selected distances in $holo$-mGBP2 during the fMD simulation. Distances are measured in Å with a bin width of 0.1 Å, while the occurrence is given in %. . . . . 45
21 Absolute B-factors in $nm^2$ are plotted against the residue indices. Data from the fMD simulation of $apo$-mGBP2 is shown in red, the corresponding data of $holo$-mGBP2 in green. In the upper and lower parts of the figure B-factors are visualized in the mGBP2 structure. As underlying structure the middle structure of the biggest cluster of each simulation was chosen. Tube thickness and colors reflect the corresponding B-factors reaching from thin/blue (=0 $nm^2$, stiff) to thick/red (=1500 $nm^2$, flexible). . . . . 47
22 The conformational space of the G domain of $apo$- and $holo$-mGBP2 explored in fMD simulations. The middle structures of selected regions resulting from cluster analyses are overlaid onto the structure of the $apo$-mGBP2 model. For the $apo$-state the G domain is shown in green, the M domain in silver and the E domain in blue. The $holo$-state is represented in red. . . . . 49

23 The conformational changes upon nucleotide binding are visualized for different structural motifs. The motifs of the $apo$-state are shown in green and the $holo$-state is colored in red. Conformations were obtained from cluster analysis and the middle structure of the highest populated cluster are shown for both $apo$- and $holo$-mGBP2. The entire structure of $apo$-mGBP2 is shown in the background in transparent green for the G domain, silver for the middle and blue for the E domain. The $holo$-state is colored in transparent red. The nucleotide GTP and the $Mg^{2+}$ co-factor are colored in magenta. . . . . 50
24 The conformational space of the $\alpha$13-helix in the GTP-bound state of mGBP2 is shown in red by overlaying all middle structures from cluster analysis at once. They are superimposed onto the initial structure of $apo$-mGBP2. The G domain is colored in green, the M domain in silver and the E domain in blue. . 51
25 Distance distributions for three salt bridges postulated between the G domain and the terminal E domain for the $apo$-state are shown. $apo$-state data is given in red, $holo$ data in green. . . . . 52
26 The figure sketches the conformational changes upon nucleotide binding where the directions of resulting dynamics are indicated by orange arrows. The GTP binding site in the $apo$-state is shown in green and the $holo$-state is colored in red. Conformations were obtained from cluster analysis and the middle structure of the highest populated cluster of both states are overlaid. The entire structure of $apo$-mGBP2 is shown in the background in transparent green for the G domain, silver for the middle and blue for the E domain. The $holo$-state is colored in transparent red. The nucleotide GTP and the $Mg^{2+}$ co-factor are colored in magenta. . . . . 53
27 The results of the H-bond analysis between the P-loop and the G4 motif are shown. The upper part reflects the $apo$-state, the lower part the $holo$-state. The strongest H-bonds are found between GLU50 and ARG181. . . . . 54
28 The H-bond between SER69 and SER246 in the $holo$-state is visualized. The named residues are shown in red for the $holo$-state and in green for the $apo$-state. In the background the structure of the entire mGBP2 is displayed with red coloring for the nucleotide-bound protein, in green the G domain of $apo$-mGBP2, in silver the M domain and in blue the E domain. GTP and $Mg^{2+}$ are visualized in purple. . . . . 55
29 Visualization of H-bonds, where the residues involved are shown in red for the $holo$-state and in green for the $apo$-state. In the background the structure of the entire mGBP2 is displayed with red coloring for the nucleotide-bound protein, green for the G domain of $apo$-mGBP2, silver for the M domain and blue for the E domain. GTP and $Mg^{2+}$ are visualized in purple. . . . . 56

30 Visualization of the H-bonds between GTP and residues SER246, ARG238, ASP182, SER73, ARG48 and LYS51. These residues are shown in red. In the background the structure of the entire mGBP2 is displayed with red coloring for the nucleotide-bound protein, green for the G domain of $apo$-mGBP2, silver for the M domain and blue for the E domain. GTP and $Mg^{2+}$ are visualized in purple. . . . . 57
31 Visualization of the coordination of $Mg^{2+}$ by SER52. Residues are shown in red. In the background the structure of the entire mGBP2 is displayed with red coloring for the nucleotide-bound protein, green for the G domain of $apo$-mGBP2, silver for the M domain and blue for the E domain. GTP and $Mg^{2+}$ are visualized in purple. . . . . 57
32 The sequence alignment of hGBP1 and mGBP2 using the EMBOSS stretcher tool for global pairwise alignment is shown. Green indicates identical sequence parts, yellow marks biochemically similar residues and red means no similarity. Gray parts show gaps between the sequences. . . . . 69

# List of Abbreviations

| | |
|---|---|
| **A** | acceptor |
| **ARG** | arginine |
| **ASP** | aspartic acid |
| **CYS** | cysteine |
| **D** | donor |
| **ED** | essential dynamics |
| **ENM** | elastic network model |
| **EV** | eigenvector |
| **FPS** | FRET-restrained positioning and screening |
| **FRET** | Förster Resonance Energy Transfer |
| **GBP** | guanylate binding protein |
| **GDP** | guanosine-diphosphate |
| **GLU** | glutamic acid |
| **GLY** | glycine |
| **GMP** | guanosine-monophosphate |
| **GTP** | guanosine-triphosphate |
| **H-bonds** | hydrogen bonds |
| **hGBP1** | human guanylate binding protein 1 |
| **hpFRET** | high performance Förster Resonance Energy Transfer |
| **H-REMD** | Hamiltonian replica exchange molecular dynamics |
| **IFN-$\gamma$** | Interferon-$\gamma$ |
| **kDa** | kilo dalton |
| **LYS** | lysine |
| **MD** | molecular dynamics |
| **cMD** | classical molecular dynamics |
| **fMD** | conformational flooding molecular dynamics |
| **Mg** | magnesium |
| **mGBP2** | murine guanylate binding protein 2 |
| **NMA** | normal mode analysis |
| **PBC** | periodic boundary conditions |
| **PCA** | principal component analysis |
| **PDB** | protein data bank |
| **PV** | parasitophorous vacuole |
| **QM/MM** | quantum mechanics/molecular mechanics |
| **RESP** | restrained electrostatic potential |
| **REMD** | replica exchange molecular dynamics |

| | |
|---|---|
| **RMSD** | root mean square deviation |
| **RMSF** | root mean square fluctuation |
| **RTB** | rotational-translational-block |
| **SAXS** | small angle X-ray scattering |
| **SER** | serine |
| *T. gondii* | *Toxoplasma gondii* |
| **THR** | threonin |
| **VAL** | valine |
| **VSV** | *vesicular stomatitis virus* |

# 1 Introduction

## 1.1 Why study the murine Guanylate Binding Protein 2?

The murine Guanylate Binding Protein 2 (mGBP2) belongs to the family of Guanylate Binding Proteins (GBP) and Atlastins of the Dynamin Superfamily of large GTPases[1]. GBPs can make up up to 20% of the total cell proteome[2]. This family is more and more discovered to play an important role in the host defense against pathogenic infections[3]. Therefore there is great interest in unraveling the variety of processes, in which these proteins are involved, to determine the interactions between GBPs, signaling and effector molecules and thereby understand their importance during immune response. Furthermore, GBPs are seen as a model system to gain insights in the functioning of large GTPases[4] and thereby of our immune system.

Highly up-regulated by interferon-$\gamma$ (IFN-$\gamma$) GBPs are expressed through classical secondary response genes[5]. Among others particularly mGBP2 is induced upon infection of mice with $Listeria\ monocytogenes$ or $Toxoplasma\ gondii$ ($T.\ gondii$)[6,7].
$T.\ gondii$ is an obligate, intracellular parasite causing Toxoplasmosis in various warm blooded vertebrates[3,8]. The importance of its understanding arises from its serious course of disease in immune deficient humans[9,10] and its ability to act as a model system for other related parasites like $Plasmodia$ – a pathogen causing malaria[11]. Via developing an infection model of mGBP2 after $T.\ gondii$ contamination one aims at a better understanding of the immune response against $T.\ gondii$ and thereby gaining new insights in the general functioning of the mammalian immune system.

## 1.2 State of the art in mGBP2 research – Why apply Molecular Modeling to mGBP2?

Currently there is already some information on the structure and properties of mGBP2 available. As the sequence identity with the human Guanylate Binding Protein 1 (hGBP1) is 68%[12--14], a structure model of mGBP2 was built based on the crystal structure of hGBP1 via homology modeling using SWISS model[3]. The nucleotide affinity of mGBP2 for the substrates Guanosine monophosphate (GMP), Guanosine diphosphate (GDP) and Guanosine triphosphate (GTP) was studied and a mutant test series revealed important residues for nucleotide binding and hydrolysis, as well as for dimerization and tetramerization of mGBP2. Truncated mGBP2 models gave first insights in the specific functioning of the different protein domains and their interactions. Moreover, the localization of mGBP2 in vertebrates and the influence of biochemical processes like GTP hydrolysis and oligomerization were examined[3,8,14].
In addition, the structure and dynamics of hGBP1 were investigated via high-precision Förster Resonance Energy Transfer (hpFRET) and small angle X-ray scattering (SAXS) on $apo$-hGBP1 monomers[15] and via electron paramagnetic resonance (EPR) and FRET on

$holo$-hGBP1 dimers[16]. This revealed the stable conformations of the monomeric protein and, together with X-ray structures of $holo$-hGBP1[17], gave insights on structural changes during nucleotide binding, hydrolysis and dimerization. Given the high sequence identity between mGBP2 and hGBP1 these findings can be probably applied to mGBP2, which would be in agreement with the findings from extensive biochemical studies[3].

Taken together, there is already knowledge about the structure of mGBP2, expected conformations, important residues for GTP binding, hydrolysis and oligomerization, but yet there is little known about the dynamics of mGBP2. Questions yet to be answered are, e.g., are there differences in the stable conformations of $apo$- and $holo$-mGBP2? How do the already revealed stable conformations interchange? What does the conformational space of functionally important residues look like? What is the dynamics upon nucleotide binding, which structural changes do lead to oligomerization and what happens during hydrolysis? Answering these questions is difficult to achieve but yet important to get a deeper understanding of the biologic functioning of mGBP2 and to develop models for immunological processes upon pathogenic infection. One possible approach is studying the murine Guanylate Binding Protein via molecular modeling as these tools are especially designed to explore the dynamics of a protein system.

## 1.3 The aim of this thesis

As the 67 kDa murine Guanylate Binding Protein 2[3,14] is rather large for molecular modeling, a first aim is to get a feeling for the handling of mGBP2 in modeling approaches; i.e., to determine appropriate modeling tools, test them for their ability to deliver new insights in the protein dynamics, and to discover whether there are particular difficulties with mGBP2. A second aim is to unveil what kind of mGBP2 dynamics is predicted by molecular modeling and how does it compare to experimental results? Can the simulations reveal structural changes and pathways between stable conformations and what insights can be gained from looking in detail at the amino acids predicted to be important by experimental studies? Finally, this thesis shall give a sound recommendation on how future work on mGBP2 via molecular modeling could proceed.

# 2 Background

## 2.1 The murine Guanylate Binding Protein 2

### 2.1.1 Characteristics of mGBP2

The murine Guanylate Binding Protein 2 is a 67 kDa[3,14], highly IFN-$\gamma$ induced protein[5,7,8,18--20] representing one of eleven murine isoforms[6,13]. It belongs to the family of Guanylate Binding Proteins (GBPs) within the Dynamin Superfamily[1]. Located in clusters on chromosome 3[6], it is encoded in a 17.4 kb long gene via 1767 base pairs on 11 separate exons coding for 589 amino acids[3].
mGBP2 has 68% sequence identity with hGBP1[12,14]. It shows a cooperative mechanism of GTP hydrolysis, which can be characterized as self-activation by oligomerization[14]. Binding to all guanosine phosphates (GTP, GDP and GMP)[12,21] it hydrolyzes GTP to GDP and GDP to GMP in two consecutive steps[14]. Whereas its nucleotide binding affinity is lower than that of the Ras family, the hydrolysis activity is much higher[3]. $Mg^{2+}$ is an essential co-factor for hydrolysis[3]. The $apo$-state is stable as monomer and does not show any aggregation[14]. Instead, mGBP2 multimerizes in a nucleotide-dependent manner: it dimerizes upon GTP binding and tetramerizes during the transitional state of hydrolysis[14]. Under cellular conditions this GTPase occurs as GTP-bound dimer[14].

### 2.1.2 Biological Function of mGBP2

To date, there is little known about the biologic functions of GBPs[14]. mGBP2 predominantly localizes in vesicle like structures[14] and is an important effector molecule in host defense against intracellular pathogens[14]. The interplay between GTP binding, oligomerization and GTP hydrolysis, as well as the isoprenylation motif of the C-terminus (see section 2.1.3) are prerequisites for the correct intracellular localization of mGBP2[3]. GBPs are assumed to have regulatory effects in cell adhesion and proliferation[22,23]. They are expected to be important in host resistance against bacteria such as $Listeria\ monocytogenes$ and $Mycobacteria$[20,24]. mGBP2 is thought to have antiviral activities in NIH/3T3 fibroblasts after infection with $vesicular\ stomatitis\ virus$ (VSV) or $encephalomyocarditis\ virus$[25,26]. It is supposed to have an inhibiting role for the cellular mobility of fibroblasts[27].
Pfeffer and co-workers studied in detail the role of mGBP2 upon infection with $Toxoplasma\ gondii$ ($T.\ gondii$) in mice[3,8,14]. It is up-regulated via IFN-$\gamma$ through several pathways including B- and T-cells[8]. After recruiting to the parasitophorous vacuole (PV) of $T.\ gondii$[14] its activity is assumed to be of reducing $T.\ gondii$ replication[8]. To do so, it needs to act in synergism with other mGBPs[8]. It is suggested to form a kind of scaffold at the PV via oligomerization upon GTP hydrolysis to enable the assembly of other binding partners and co-factors, which then fulfill the antimicrobial functions[14] and destroy the PV[8].

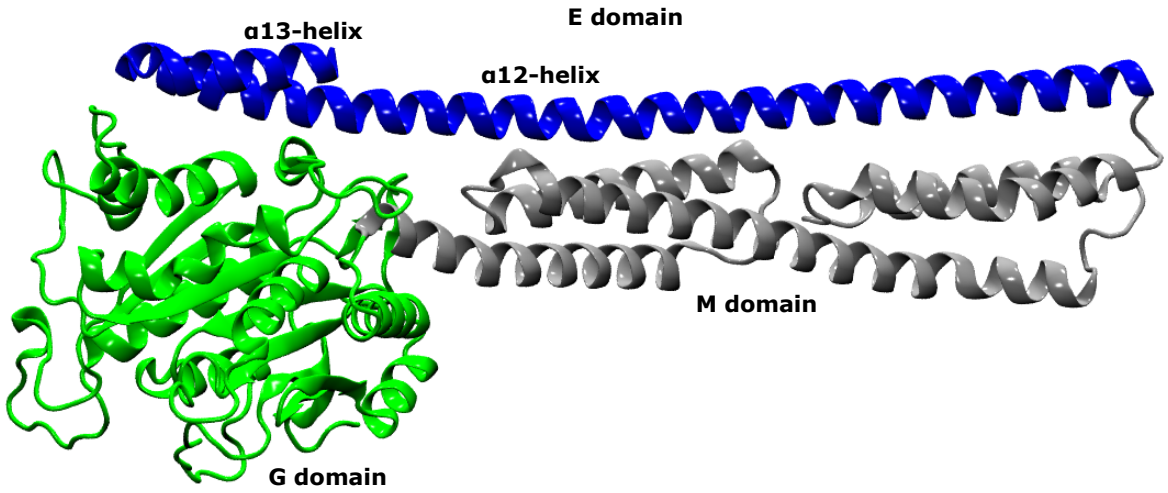



Figure 1: Structure model of mGBP2, which consists of three domains. The G domain is colored in green, the M domain in silver and the E domain in blue.

### 2.1.3 Structure of mGBP2 and Hydrolysis Model

Since up to now only the crystal structure of hGBP1 has been solved, the structure of mGBP2 is not known yet. However, given the high sequence identity between hGBP1 and mGBP2 it was possible to predict the mGBP2 structure via homology modeling using the SWISS-MODEL server[28,29]. The mGBP2 overall structure is expected to be a three domain structure consisting of a Ras-like G domain, which is the catalytic GTPase domain, a middle domain (M domain) built of two bundles of $\alpha$-helices and the long $\alpha$-helical E domain spanning the complete backside of the protein[3] (see figure 1).

The E domain consists of the protein spanning $\alpha$12-helix and, after a turn, the $\alpha$13-helix near the C-terminus of the protein. At the C-terminus the CAAX sequence is located (amino acids 586-589), which is a recognition motif for farnelylization or geranylgeranylization enzymes. This enzyme will attach an isoprene at the X position. Thereby the sub-cellular localization of mGBP2 is determined and interactions with a membrane are enabled[3]. The E domain is stabilized via salt bridges between the G domain and the $\alpha$12/13-helices. They are visualized in figure 2.

Apart from some deviations the G domain of mGBP2 shown in figure 3 is mainly conserved from the canonical Ras structure[14]. It includes the four motifs (G1 –G4) where G1 and G3 are fully conserved[3]. The G1 motif (P-loop) is responsible for binding GTP via hydrogen bonds (H-bonds) to $\alpha$-, $\beta$- and $\gamma$-phosphate, the hydroxy groups of serine (SER) and threonin (THR) coordinate $Mg^{2+}$ in the $holo$-state. The so-called switch I region (G2 motif) is important during hydrolysis as it coordinates the reacting water molecule and $Mg^{2+}$, and shields the binding pocket from the solvent. Just as the switch I region the switch II region (G3 motif) undergoes large conformational changes upon hydrolysis. Its THR and glycine (GLY) residues coordinate the GTP $\gamma$-phosphate[3]. The G4 motif contains the recognition

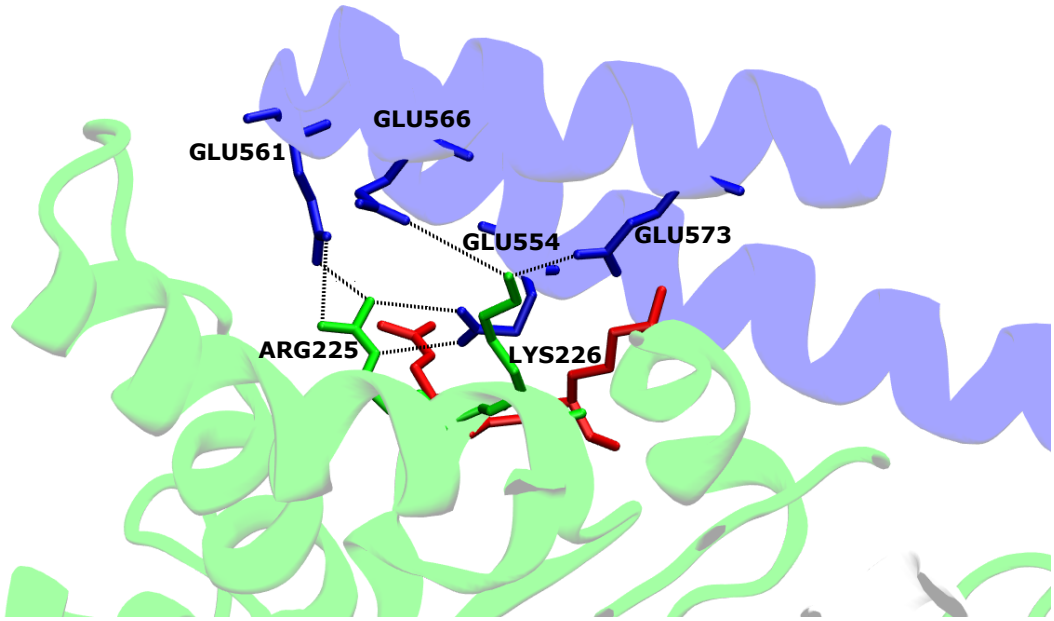



Figure 2: Salt bridges between the G domain and the $\alpha$12/13-helix of $apo$-mGBP2 are visualized, as well as possible changes in the $holo$-state (shown in red). The G domain of $apo$-mGBP2 is colored in green, the M domain in silver and the E domain in blue.

sequence for the guanine base and is only conserved in form of a reduced RD motif from the canonical Ras structure[3]. Aspartic acid (ASP) stabilizes GTP and GDP via H-bonds, two more amino acids preserve the structure of the nucleotide binding pocket via contacting the P-loop[3]. There is one more exclusive motif found for hGBP1 called the guanine cap. It shows large conformational changes upon nucleotide binding and seems to be important for the stable binding of GTP/GDP, hydrolysis and dimerization[30]. This structural element is also predicted for mGBP2[3]. Furthermore, from crystal structures of $holo$-hGBP1[17] it is seen, that the $\alpha$4'-helix also severely changes its conformation during GTP binding[31].
From mutation studies of arginine (ARG) 48, lysine (LYS) 51, ASP 182 and glutamic acid (GLU) 99 the important amino acids carrying the specific functions of the G1-G4 motifs were resolved[3]. ARG 48 acts as the so-called "arginine finger" which is important for the hydrolysis activity of mGBP2 and stabilizes the mGBP2 dimer[14]. With its positively charged side chain LYS 51 supports the binding of GTP via $\beta$- and $\gamma$-phosphate interactions and is essential for a cooperative hydrolysis mechanism. It preserves the P-loop structure by forming H-bonds to THR 98 (switch I)[3]. The specificity for guanine bases is ensured by ASP 182 in the G4 motif[3,14]. GLU 99 (switch II) might stabilize the transition state of hydrolysis and thereby accelerate the phosphate cleavage by conformational changes[3,14].

The mechanism of hydrolysis is not entirely solved yet but differs from the mechanisms in the Ras and $G_\alpha$ family[3]. Upon GTP binding and subsequent hydrolysis to GDP, large conformational changes occur in hGBP1. The closure of the binding pocket by a large conformational change of the guanine cap enables stable nucleotide binding and initiates the dimerization process through polar interactions with the switch regions forming head

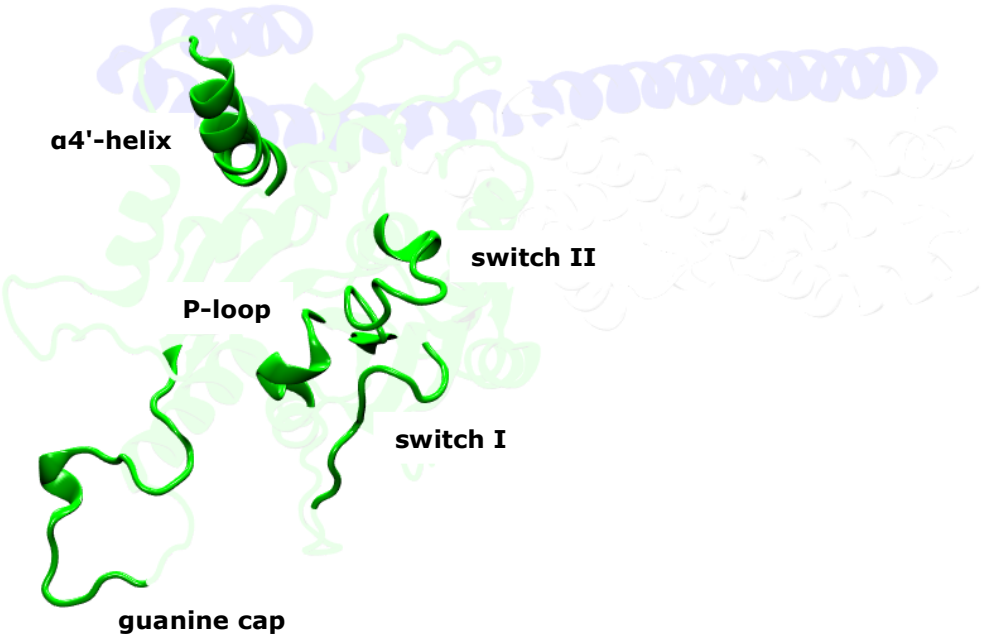



Figure 3: The G domain of $apo$-mGBP2 is displayed emphasizing the guanine cap, the P-loop, the switch I and II regions, as well as the $\alpha$4'-helix (in glossy green). The entire protein is shown in the background: the G domain is colored in transparent green, the M domain in transparent silver and the E domain in transparent blue.

to head dimers[30]. The $\alpha$6-helix is assumed to play a critical role during this process[32]. It is at 80% identical between hGBP1 and mGBP2 and is therefore likely to be involved in the dimerization of mGBP2, too[3]. Dimerization leads to a self-activation of the hydrolysis making it a cooperative mechanism.
Like for hGBP1[31], during hydrolysis of GTP to GDP the $\alpha$4'-helix of the G domain seems to mediate a force to the E domain of mGBP2 resulting from structural changes in the active site[3]. This could unveil a surface for further interactions with other proteins, e.g., causing tetramerization by dimerization of two mGBP2 dimers[3]. This tetramerization is only possible during the transitional state of GTP hydrolysis[3]. Because of the high sequence identity between hGBP1 and mGBP2, similar mechanisms are expected for both proteins.

Via truncation studies it was shown that for its biological functions all three mGBP2 domains are essential[3]. The G and E domain have direct contacts[3] and it is proposed that the E domain inhibits hydrolysis of externally bound GDP because conformational changes from the first cleavage step are missing[3]. Whereas the E domain is important for the recruitment of mGBP2 to the PV of $T.\ gondii$ and membrane attachment, the G domain regulates this process. It ensures that mGBP2 is only activated after IFN-$\gamma$ stimulation by unmasking a binding motif in the E domain during hydrolysis as explained above[8]. This is in line with the finding that GTPase activity is needed for assembly at the PV[14].

## 2.2 Elastic Network Models

Elastic Network Models (ENM) provide a simplified, low-resolution model of a biomolecule (in particular proteins) intending to describe its overall large-scale, low-frequency motions[33]. The biomolecule is modeled via a three-dimensional elastic network build up of beads representing a user-defined group of atoms and connecting beads below a certain predefined cutoff via springs[34].
Taking a protein as example, these beads are often represented by the $C_\alpha$-atoms and all $C_\alpha$-atoms closer than, e.g., 10 Å from each other are connected via springs. This leads to a very simple energy function, which is only dependent on these elastic connections. There are different approaches to the spring descriptions, for example, either using a single-parameter Hookean potential like suggested by Tirion[33] or a more sophisticated description like the Hinsen model[35,36]. Equation (1) shows the potential $V_p(r_a, r_b)$ as proposed by Tirion[33]:

$$V_p(r_a, r_b) = \sum_{|r^0_{a,b}|<r_c} \left[ \frac{C}{2} \left( |r_{a,b}| - |r^0_{a,b}| \right)^2 \right] \tag{1}$$

with parameter $C$ being a definable force constant describing all interactions, $\mathbf{r_{a,b}} \equiv r_a - r_b$ denoting the distance vector between atoms $a$ and $b$, the superscript $0$ indicates the starting configuration, and the sum runs only over atom pairs being closer than a predefined cutoff $r_c$.

By applying Normal Mode Analysis (NMA) to equation (1) (meaning diagonalizing the matrix of second derivatives of the potential energy, i.e., the Hesse matrix)[33] the system's eigenvectors (EV) and corresponding eigenvalues are obtained. The eigenvalues are directly related to the frequencies of the normal modes. The resulting eigenvectors are sorted according to growing frequency, where for molecules with three degrees of rotational freedom (valid for all biomolecules) the first six eigenvectors represent the translational and rotational so-called trivial modes. The interesting part are the non-trivial low-frequency modes representing the system's large-scale conformational changes. It was shown, that the temperature factors (so-called B-factors) predicted from ENMs are in good correlation with crystallographic data and that specific conformations are recovered[37].

One has to be aware that applying NMA to the ENM does only make sense, if the ENM describes an equilibrium state of the represented biomolecule meaning that it reflects an energy minimum[38]. That is because the simple ENMs are basically a harmonic approximation of the biomolecule's potential function and applying NMA defines the provided structure as the energy minimum of this harmonic potential. This already delineates the only requirement needed as input to build an ENM: A structure of the biomolecule at an equilibrium state must be available (e.g., a crystal structure). ENMs save the costly energy minimization step which would be necessary when doing NMA with a standard biomolecular

force field. Furthermore, these models speed up the matrix diagonalization remarkably by only considering a subset of atoms, e.g., the $C_\alpha$-atoms[34]. Taken together this leads to the clear advantage of ENMs: They provide key insights into large-scale conformational, often function-related motions of proteins which are often computationally demanding to obtain via classical Molecular Dynamics (cMD) simulations, especially for large biomolecules. And as explained, this is achieved at almost no computational expense: The mGBP2 protein studied in this work consists of 573 residues. Building a $C_\alpha$-model leads to an ENM with 573 beads, providing NMA results in less than 5 minutes on four 2.53 GHz cores.

## 2.3 Molecular Dynamics Simulations

### 2.3.1 Newton's equation of motion – Leap Frog algorithm

The underlying foundation of molecular dynamics (MD) simulations is Newton's Equation of motion (2)[39,40]:

$$F_i(t) = m_i \dot{v}_i(t) \tag{2}$$

It says that the force $F_i$ currently acting on particle $i$ and its mass $m_i$ solely determine the acceleration of particle $i$. Knowing the current position of particle $i$ and its velocity $v_i$ one can obtain its future position via integrating over a time step $\Delta t$. In MD simulations this is implemented by iteratively updating the forces on particle $i$ and then calculating the newly resulting velocities and positions for a defined time step $\Delta t$.
The leap-frog algorithm[41] is one widely used algorithm to iteratively integrate Newton's differential equation (2). Its name arose due to its characteristic that positions and velocities always jump over each other like leapfrogging. Velocities are updated at each half time step $\Delta t/2$ and positions at each time step $\Delta t$.

$$v_i\left(t - \frac{\Delta t}{2}\right) = \frac{x_i(t) - x_i\left(t - \Delta t\right)}{\Delta t} \tag{3}$$

$$v_i\left(t + \frac{\Delta t}{2}\right) = \frac{x_i\left(t + \Delta t\right) - x_i(t)}{\Delta t} \tag{4}$$

Equation (3) describes the latest known velocity, equation (4) the future velocity, where $v_i$ denotes the velocity and $x_i$ the position of particle $i$. In case of all-atom MD simulations particle $i$ represents one atom. $\Delta t$ defines the time step between iterations and $t$ indicates the current time. By expanding equation (4) in a Taylor series, the future velocity can be obtained:

$$v_i\left(t + \frac{\Delta t}{2}\right) \approx v_i\left(t - \frac{\Delta t}{2}\right) + \Delta t \frac{F_i(t)}{m_i} \tag{5}$$

with $m_i$ being the mass of and $F_i$ the force acting on atom $i$. Solving equation (4) for $x_i(t + \Delta t)$ yields the future position of atom $i$:

$$x_i\left(t + \Delta t\right) = x_i\left(t\right) + \Delta t v_i\left(t + \frac{\Delta t}{2}\right) \tag{6}$$

One has to be aware that the resulting position leads to an error of order $O$ ($\Delta t^4$) and the velocity to an error of order $O$ ($\Delta t^2$) because of stopping the Taylor expansion (5) after the second term.

### 2.3.2 Force Fields – Amber ff99SB-ILDN

The leap-frog algorithm (see section 2.3.1) itself does not yet allow us to compute new positions and velocities for a given atom. The force acting on each atom has to be determined first to solve equation (5). It is calculated via the first derivative of the potential energy:

$$F_i = \frac{\partial V(R)}{\partial x_i} \tag{7}$$

with $V(R)$ denoting the potential energy which depends on $R$, i.e., all positions of the $N$ atoms composing the system and $x_i$ is one of the three Cartesian coordinates of atom $i$.

In MD simulations force fields are used to describe the potential energy of a simulation system. Various implementations exist to account for this need, which can be divided into knowledge-based and physics-based force fields[42]. The first derive the potential energy from statistics and experimental data, whereas the latter account for the interactions between particles via a physical description. Physics-based force fields can be further divided into all-atom and coarse-grained force fields. Whereas all-atom force fields represent the highest level of theory explicitly describing each atom of the system, coarse-grained models group several atoms into one bead to reduce the computational effort.
Within these categories one can choose between many different models, each having individual strengths and weaknesses, which, e.g., can be overstabilization of $\alpha$-helices, salt-bridges or protein oligomers. This is due to the fact that force fields represent empirical descriptions of the biomolecules, which involve many parameters which where obtained in a complex parametrization procedure[43--46]. Well known all-atom force fields are the CHARMM[47--49], GROMOS[50,51] and AMBER[52--54] force fields. Choosing the correct level of theory and force field is a very sensitive decision immensely depending on the simulation problem one wants to describe. When analyzing simulation results one has to carefully consider the impact of the force field applied to the problem of interest[42].

The mutual foundation of all physics-based all-atom force fields is the systematic derivation

of the system's potential energy based on equation (8):

$$V = V^{bond} + V^{elec} + V^{vdW} \tag{8}$$

$V^{bond}$ accounts for the bonded interactions, while the non-bonded interactions arise from electrostatic $V^{elec}$ and van der Waals $V^{vdW}$ interactions[45]. One example for such a physics-based all-atom force field is the Amber ff99 force field given in equation (9)[52]:

$$\begin{aligned} V^{\text{ff99}} = & \sum_{i \in bonds} K_{r,i}(r_i - r_i^0)^2 \\ & + \sum_{i \in angles} K_{\theta,i}(\theta_i - \theta_i^0)^2 + \sum_{i \in dihedrals} \frac{V_n}{2}\left[1 + cos(n\phi_i - \gamma)\right] \\ & + \sum_{i \in atoms} \sum_{j>i} \frac{q_i q_j}{\epsilon r_{ij}} + \sum_{i \in atoms} \sum_{j>i} \left[\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}}\right] \end{aligned} \tag{9}$$

Here, the first three terms account for the bonded interactions $V^{bond}$ whereas the last two terms describe the non-bonded electrostatic $V^{elec}$ and van der Waals $V^{vdW}$ interactions. $K_{r,i}$, $K_{\theta,i}$ and $V_n$ denote the force constants for bond vibrations, angle bending and dihedral torsions, the subscript $0$ marks the corresponding equilibrium geometry. For the non-bonded interactions, $r_{ij}$ represents the distance between atoms $i$ and $j$ with $q_i$ and $q_j$ being their partial charges. $A_{ij}$ and $B_{ij}$ correlate to their van der Waals radii. The last term describing the van der Waals interactions has the form of a Lennard-Jones potential. The parameter $\epsilon$ denotes the dielectric constant.

The Amber ff99 force field has been re-parametrized and thereby improved twice over time: First it was replaced by the Amber ff99SB force field introduced by Hornak et al.[53], which improved the backbone $\phi/\psi$ dihedral terms in order to obtain a better balance between $\alpha$-helical and $\beta$-sheet structures. The last update was by Lindorff-Larsen et al.[54] to improve amino acid side-chain torsion potentials of isoleucine, leucine, aspartate and asparagine (I,L,D,N) of the Amber ff99SB force field leading to the Amber ff99SB-ILDN force field.

For the calculation of the potential energy as defined in equation (9) two ingredients are needed: Atom coordinates which are being provided during the MD simulation and force field parameters. The latter are of utmost importance and the core problem when developing a new force field. As a force field is generally not intended to describe a special atom or molecule but, for example, any protein, these parameters have to be carefully derived from ab-initio calculations at the quantum-mechanics level and fitting to experimental target data like vibrational spectra and solvation free energies[45]. They highly determine the quality of the force field and the achievable results in MD simulations.

### 2.3.3 Enhanced Sampling Simulations – Flooding

A main concern of molecular dynamics simulations is to achieve a sufficient sampling of the conformational space to derive statistically significant results[55,56], which becomes more critical with growing degrees of freedom. Therefore a vast effort is taken to develop advanced techniques which enhance the sampling of MD simulations, which widely differ in their theoretical foundations and approaches.
Adcock and McCammon proposed to summarize the developments into three, partially overlapping classes: methods explicitly modifying the potential energy surface, methods using non-Boltzmann sampling, and methods accelerating the sampling of particular degrees of freedom while slowing down others[57]. Into the first class they count, for example, the deflation method[58], umbrella sampling[59], local elevation[60], potential smoothing[61], conformational flooding[62--64], hyperdynamics[65,66], and accelerated MD[67]. Following Bernardi et al. also metadynamics[68] should be added here[56]. The main characteristic of this group is to enlarge the sampling efficiency by reducing the simulation time spent at local minima and thereby to prevent the simulation from being stuck at a local minimum. The second group, in principle, follows the same idea. It includes different sampling methods like simulated annealing[69] and replica exchange MD (REMD)[70,71]. They differ from the first group in that they leave the energy surface unchanged. One of the most established methods from the third group is the SHAKE algorithm[72], which is widely used to constrain bond lengths during MD simulations.

In this work the conformational flooding technique[64] as implemented in the GROMACS software package[73--78] is applied. Its basic principle is to destabilize a local minimum structure by adding a Gaussian shaped flooding potential $V_{\mathrm{fl}}$ to the potential energy surface along user-defined so-called collective coordinates (see figure 4). When simulating proteins, these collective coordinates are typically chosen such that they reflect the Essential Dynamics (ED) of the protein. ED represent the overall large-scale motions of a system emphasizing the function-related dynamics. They may be obtained via Principal Component Analysis (PCA)[79] from a preceding MD simulation or also via NMA from an ENM (as introduced in section 2.2).
The destabilization of the local minimum structure is achieved through adding a flooding potential $V_{\mathrm{fl}}$ (equation (10)) along the $a\ priori$ defined $m$ collective coordinates.

$$V_{\mathrm{fl}}(c_1, \ldots c_m) = E_{\mathrm{fl}} \exp\left[-\frac{k_B T}{2E_{\mathrm{fl}}} \sum_{j=1}^{m} \lambda_j \left(c_j - c_j^{(0)}\right)^2\right] \tag{10}$$

Here, $k_B$ denotes the Boltzmann-constant, $T$ the simulation temperature, and $\lambda_j$ designates a curvature reflecting the eigenvalue corresponding to $c_j$, the $j$-th eigenvector of the $m$ user-defined collective coordinates. $c_j^{(0)}$ describes the center of the flooding potential of eigenvector $j$. $V_{\mathrm{fl}}$ can be chosen to be constant throughout the simulation via a user-defined fixed flooding strength $E_{\mathrm{fl}}$ or be adapted by defining a target destabilization free

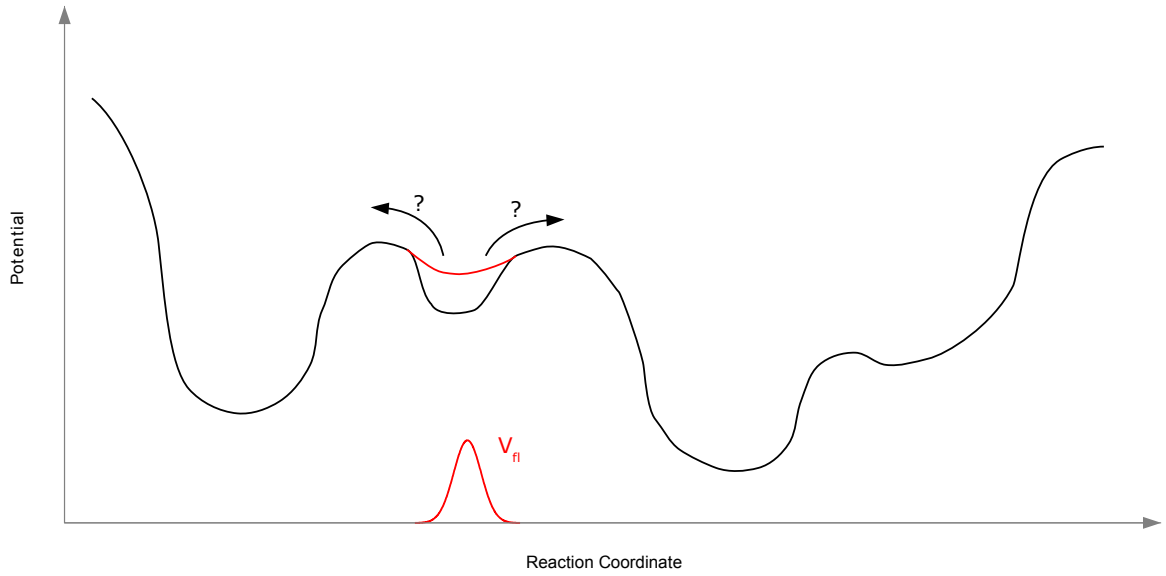



Figure 4: The plot sketches the principle of conformational flooding. The red Gaussian shaped flooding potential $V_{\mathrm{fl}}$ is added to the system's potential along user defined reaction coordinates suggested to reflect the essential dynamics of the protein. The figure was adapted from Lange et al.[64]

energy $\Delta F_0$ and updating the corresponding flooding strength $E_{\mathrm{fl}}^{(i+1)}$ at each time step $i$ (equation (11)).

$$E_{\mathrm{fl}}^{(i+1)} \equiv E_{\mathrm{fl}}^{(i)} + \frac{\Delta t}{\tau}\left[\Delta F_0 - \Delta F^{(i)}\right] \tag{11}$$

$\Delta t$ is the step size used for the MD simulation and $\tau$ the user-defined time constant, which controls the update frequency of $E_{\mathrm{fl}}$. $\Delta F^{(i)}$ and $E_{\mathrm{fl}}^{(i)}$ are the destabilization free energy and flooding strength at the current time step $i$, respectively.

The advantage of the flooding technique is that it allows an unbiased sampling of the conformational space towards any direction by only driving the system away from local minimum states. Apart from the required collective coordinates from PCA or NMA there is no further prior information, like transition pathways or product states, needed.[64]

# 3 Methods

## 3.1 mGBP2 structure models for molecular modeling

### 3.1.1 Nucleotide-free model of mGBP2

To date, the only GBP structure that has been solved is the one of hGBP1[80] (protein data bank (PDB)[81] search as of 17th of September, 2015). As mGBP2 shows a sequence identity of 68% with hGBP1[12,14], the crystal structure of the $apo$-state of the latter (PDB entry 1DG3)[80] was used by Kravets[3] to build a model of $apo$-mGBP2 via homology modeling using the SWISS-MODEL server available at *http://swissmodel.expasy.org/*[28,29]. Thereby a structural model of mGBP2 covering amino acids 6 to 578 was obtained. The last eight amino acids including the CAAX sequence are missing in this model as the structure of the corresponding hGBP1 residues could not be resolved[3]. The structure model obtained from Kravets[3] was the basis for all modeling of the $apo$-state of mGBP2 done in this work.

### 3.1.2 GTP bound model of mGBP2

The crystal structure of the G domain of $holo$-hGBP1 bound to the non-hydrolyzable GTP analogue GppNHp (PDB entry 2BC9)[17] served as precursor to obtain a homology model of the $holo$-state G domain of mGBP2 using the same method as explained above (section 3.1.1). To be able to simulate the complete $holo$-state of mGBP2, the M and E domains had to be added to the nucleotide-bound G domain. To this end, the $holo$-state G domain model (also obtained from Kravets[3]) was used to replace the G domain of the existing $apo$-state model. To do so, the latter was cut at the $\psi$-angle of valine (VAL) 286 and amino acids 6 to 286 were replaced by the $holo$-state G domain model. To prevent the loop from residue 155 to 170 from being wrapped around the $\alpha$12-helix, the bond between the C and $C_\alpha$-atoms of residue 481 was slightly stretched to move the $\alpha$12/13-helices a little apart from the G domain. All these transformations were done using the VMD software[82].
Furthermore, the nucleotide GTP and the $Mg^{2+}$ co-factor had to be added. To do so, the resulting structure was further processed with the free version of the Maestro program[83]. The $holo$-state model was loaded together with the coordinates of $Mg^{2+}$ and GDPxAlF$_3$ obtained from the crystal structure of the GDPxAlF$_3$ bound hGBP1 G domain (PDB entry 2B92)[17]. AlF$_3$ was removed and a phosphate group was attached to the $\beta$-phosphate of the existing GDP. The $\gamma$-phosphate was added in a way to avoid atom clashes and to be in a reasonable position relative to important protein residues, in particular to LYS51 and the $Mg^{2+}$ co-factor as described by Kravets (see figure 1.13 of Kravets' dissertation[3] and section 2.1.3 of this work). The GTP structure was protonated to mimic the physiological pH value of 7.4 (the protonation state was derived from the GTP structure with PDB code 2KSQ, which was obtained using solution NMR at a pH value of 7[84]) giving it an overall charge of $-4$. It was then pre-optimized according to the protein and co-factor environment within Maestro, which makes use of the OPLS_2005 force field[85].
The here described procedure is in line with previous work from others[86]. The obtained

coordinates defined the input structure for the simulations of the $holo$-state which was first energy minimized in order to optimize the geometry, especially to relax the artificially enlarged bond in residue 481, using the GROMACS package. Details on the setup and minimization and MD simulation protocols are given in section 3.3.1.

### 3.1.3 Derivation of GTP force field parameters

As there are no standard parameters implemented in the Amber ff99SB-ILDN force field for GTP in GROMACS, they were taken from the work of Meagher et al.[87]. A similar procedure for the incorporation of the named GTP parameters in GROMACS was recently published by Gao and Eriksson[86]. The still missing partial atomic charges of GTP were derived following the procedure of the Amber ff99SB-ILDN force field parametrization using the restrained electrostatic potential (RESP) method[88] at the Hartree-Fock theory level applying the 6-31G* basis set after prior geometry optimization of GTP using the B3LYP functional with the 6-31G* basis set. These quantum chemical calculations were done using the Gaussian 09 program[89] following the recommendations of the Department of Theoretical Chemistry of the Lund University (available at *http://www.teokem.lu.se/ ulf/Methods/resp.html*) and are in line with procedures described elsewhere[90,91]. Finally a complete parameter file for use in GROMACS was built by combining the derived RESP charges and the GTP parameters published by Meagher et al.[87] in Amber *frcmod* file format via the antechamber[92,93] and tleap[94] tools available within the AmberTools15 software package[95]. The resulting *prmtop* and *prmcrd* files were converted to GROMACS topology and coordinate files using the ACpype tool[96].

## 3.2 Elastic Network Models

### 3.2.1 Software for ENM calculations and employed parameters

All elastic network models applied in this work were build using the $\Delta\Delta$PT toolbox developed by Rodgers et al.[97] The source code is available at *http://sourceforge.net/ projects/durham-ddpt/* and can be installed at a local machine free of charge. For setting up the ENMs, structure files in *pdb* file format were parsed to the toolbox and coarse grained models only considering $C_\alpha$-atoms without mass-weighting were built, using a distance cutoff of 10 Å for spring connections applying a single-parameter Hookean potential with a force constant of 1 kcal/mol/Å$^2$. These settings reflect a standard procedure as described elsewhere[34]. Normal mode analysis was fully performed with the DIAGSTD procedure of $\Delta\Delta$PT not applying the rotational-translational-block (RTB) approach[98]. The results were further analyzed using the FREQ/EN tool at a temperature of 298 K, the RMS/COL tool considering all non-trivial mode eigenvectors and the NMWIZWT program for normal mode visualization.

### 3.2.2 Verifying the choice of ENM software and ENM parameters

To check the influence of different parameters and levels of coarse graining, tests of different ENMs were performed. Here, all-atom ENMs and $C_\alpha$-ENMs with residue mass weighting were considered. Visual inspection of the resulting three non-trivial normal modes with smallest eigenvalues using the VMD software, comparison of the overlap between corresponding eigenvectors (using the OVERLAP procedure of $\Delta\Delta$PT described in section 3.4.2) and the mode collectivity did not show any major differences. But it was seen that the amplitude of motions declined with increasing model accuracy, which is not favored.

As many different ENM software like the elNèmo[99], WebNMA[100] and ADENM[101] web servers and others are available, it was further tested whether they produce different results. Building standard $C_\alpha$-ENMs and doing NMA with these web servers showed in general a high agreement in results. These were examined following the same procedure as mentioned above. Only the WebNMA web server differed from the others, yielding a reduced eigenvector overlap of 93-96% compared to almost 100% for the others.

The findings of the tests performed are in line with the results from Kitao et al.: The choice of ENM and the level of theory (e.g., describing the potential via all-atom force fields vs. a simple single-parameter Hookean potential) is seen to be uncritical for the low frequency collective motions[102,103]. Therefore, the software choice for ENMs can be done according to user convenience. Here the $\Delta\Delta$PT has clear advantages compared to the others tested, because it allows to control any single parameter and setting when building the ENM. Furthermore, the calculations are all easy to follow, while it turned out to be rather difficult for the web servers to follow, what was actually done during building the ENM and performing the NMA, giving the impression of being a ''black box''. Finally, the $\Delta\Delta$PT toolbox comes with many additional features like visualization of normal modes, it is easy to install and use and it is distributed with a detailed documentation.

## 3.3 Molecular Dynamics Simulations

### 3.3.1 Free Simulations

Classical Molecular Dynamics (cMD) simulations were performed using the GROMACS[73--78] software package – available as free download at *http://www.gromacs.org/* – and applying the Amber ff99SB-ILDN force field[52--54]. The mGBP2 $apo$-state model was solvated with the explicit TIP3P water model[104] in a dodecahedron box employing periodic boundary conditions (PBC) with a 10 Å minimum distance from any mGBP2 atom to the closest box edge. $Na^+$ was added to neutralize the system and the temperature was set to 298 K. Missing hydrogens were added using the interactive *pdb2gmx* command to mimic the physiological pH of 7.4. To decide on the protonation states of amino acids the propka31 tool[105,106] was used to predict pKa values. Amino acids with a pKa value greater than pH=7.4 were assigned to be protonated, in turn those with a pKa lower than pH were decided to be not protonated.

After setting up the system, an energy minimization was performed using the steepest descent algorithm with a maximum force tolerance of 500 kJ/mol/nm on any atom and a maximum number of steps of 10,000. Further details can be seen in the $apo$-state minimization *mdp* file in the appendix (section 6.1.1). Next, the system was equilibrated for 100 ps using the leap-frog integrator[41] (compare section 2.3.1), the Berendsen barostat[107] ensuring a pressure of 1 bar and the velocity-rescale thermostat[108] maintaining a temperature of 298 K. Protein heavy atom coordinates were restrained. Velocities were generated according to a Maxwell-Boltzmann distribution at 298 K. Finally, the MD production run was performed for 100 ns applying the Nose-Hoover thermostat[109,110] and the Parinello-Rahman barostat[111,112] at the before mentioned temperature and pressure. Further equilibration and production run specifications are given in the appendix (section 6.1.2 and 6.1.3).

To set up the $holo$-state simulation system for the GTP bound model of mGBP2 described in section 3.1.2, the aforementioned procedures for the $apo$-state system were followed including a new pKa value prediction with propka31. However, slight changes to the settings described above had to be made for the $holo$-state simulations. For a successful energy minimization, GROMACS had to be run in double precision mode. The initial energy step size was enlarged to 0.1 Å. Two equilibration runs were used, the first for velocity generation while restraining protein heavy atom coordinates, as well as GTP and $Mg^{2+}$, the second only restraining protein heavy atoms, which allowed GTP and $Mg^{2+}$ to find their optimal position in the binding pocket. For temperature coupling, two groups were used: one including protein, GTP and $Mg^{2+}$, and the other one with water and ions. The MD production runs were performed with the same temperature coupling groups as during equilibration. For later analysis the energy terms were split into two groups, protein and GTP. All specifications can be viewed in the appendix in sections 6.1.1, 6.1.2 and 6.1.3.

### 3.3.2 Enhanced Sampling Simulations – Flooding

Enhanced sampling simulations were performed using the flooding tool[64] implemented in GROMACS for both the $apo$- and the $holo$-state. When specified in the *mdrun* command this tool is called by GROMACS' main MD engine within its *edsam* module and accounts for the force calculations due to the additional flooding potential. The needed input parameters are parsed to GROMACS via the *edi* input file which can be generated with GROMACS' *make_edi* program[64].

Enhanced sampling simulations shall speed up the sampling of collective essential dynamics. Because the simulation time needed to receive a converged principal component analysis and thereby a sufficient description of reaction coordinates for flooding is unclear, a coordinate derivation via NMA was preferred. Here, elastic network models provide an efficient possibility and an appropriate level of theory for the calculation of normal modes and the determination of flooding coordinates. As an ENM and subsequent NMA defines the input geometry as the energy minimum, the reaction coordinates of the flooding potential

need to be updated during a flooding MD (fMD) simulation. The fact that ENMs are not implemented in GROMACS and therefore were built with the ΔΔPT program as described in section 3.2.1 rendered the *make_edi* program of GROMACS useless. Therefore, a new tool was developed to convert the information from ΔΔPT into GROMACS format, to define the flooding parameters, and to generate a file in *edi* format to start fMD simulations.

The interface of the developed "*edi file generator*" allows the user to set the desired flooding parameters (e.g., the temperature $T$ and the flooding strength $E_{\mathrm{fl}}$ (see equation (10) and (11)), determine the number of $m$ eigenvectors to be included for defining the flooding potential, and parse the output files from ΔΔPT containing the eigenvectors and eigenvalues, as well as the file containing the coordinates used to build the ENM together into the needed *edi* file. Furthermore, the user can include the coordinates of a structure to define the origin ($c_j^{(0)}$) and the target of the flooding potential. All structure files need to be in *pdb* file format.
As eigenvectors of normal modes form an ortho-normal system and are therefore dimensionless they are taken from ΔΔPT without any conversion. The corresponding eigenvalues $\tilde{\nu}_j$ obtained from ΔΔPT in cm$^{-1}$ were converted into eigenfrequencies $\nu_j^{GR}$ in GROMACS units according to equation (12) and curvatures $\lambda_j$ (see equation (10)) for flooding needed were calculated by equation (13).

$$\nu_j^{GR} \equiv \frac{(2\pi c)^2 N_A m_u}{10^{11}} \tilde{\nu_j}^2 \tag{12}$$

$$\lambda_j = \frac{\nu_j^{GR}}{kT} \tag{13}$$

Here, $c$ denotes the speed of light in GROMACS units (nm/ps), $N_A$ the Avogadro constant in mol$^{-1}$, $m_u$ the atomic mass unit in kg, $k \equiv 2.5/300.0$ the Boltzmann constant as defined for flooding[64] and $T$ the desired temperature.
Because the structure files parsed to the *edi file generator* are in *pdb* file format, coordinates are given in Å. They therefore need to be converted to nm to be used within GROMACS.

Figure 5 visualizes the work-flow of the fMD simulations performed in this work. Because of the need to update the flooding potential after a short simulation time, an fMD simulation is split into loops consisting of the following steps:

1. Build an ENM and perform NMA using ΔΔPT.
2. Generate the flooding input file (*edi* file) using the *edi file generator* and define flooding parameters.

3. Run the fMD simulation using GROMACS.
4. Save the coordinates of the last frame of the fMD simulation to a *pdb* file and with that start again at step 1.

Before starting fMD simulations the systems were set up, energy minimized and equilibrated as described in section 3.3.1. To start fMD simulations, the last frame of the preceding equilibration was saved to a *pdb* file to build the first ENM. Then the work-flow of figure 5 was followed. For building the ENMs and performing NMA the standard parameters as described in section 3.2.1 were used. The structure based on which the ENM was built was further used to define the origin of the flooding potential. No target state was required. Eigenvectors 7 to 16 were chosen to define the flooding potential – this means the first ten non-trivial, lowest frequency modes from NMA, as the first six eigenvectors describe the translational and rotational transformations. The flooding potential was chosen to be constant with a flooding strength $E_{\mathrm{fl}}$ of 10 kJ/mol and a temperature of 298 K. Further details are provided in the the parameter definition part of a sample *edi* input file of fMD simulations shown in the appendix (see section 6.1.4). The length of one fMD simulation loop was set to 250 ps. Any other MD options remained unchanged from what was described previously (see section 3.3.1). To cover a total simulation time of 100 ns, 400 update loops had to be performed. This was implemented via the ''*fMD*'' tool, which autonomously performs the work-flow depicted in figure 5 on a computing cluster.

The choice of parameters for fMD simulations is critical, as up to now their influence on simulations and results is untested and no reference work was found using the flooding technique in MD studies except the publication of the flooding methodology itself[64]. The pivotal parameters are expected to be the flooding mode, i.e., constant or adaptive flooding, the strength of the flooding potential, the number of reaction coordinates to which a flooding potential is to be added and the duration of one fMD simulation loop.
For testing the plausibility of chosen parameters the basic principle of ''as much as necessary and as little as possible'' was assumed: To maintain a physically meaningful sampling during fMD simulations the impact of the flooding potential on the protein's native potential energy surface should be as small as possible.
Determining the appropriate number of eigenvectors from NMA to define the flooding potential a procedure similar to what was reported by Costa et al.[113], as well as Peng and Zhang[114] was followed. As an fMD simulation aims at enhancing global conformational changes only the first few highly collective modes should be included as we do not wish to enhance local and thermal fluctuations. Still the number of eigenvectors has to be high enough to reproduce the overall flexibility of the protein. Therefore, B-factors were calculated from an ENM of the $apo$-model of mGBP2 using all non-trivial mode eigenvectors. In addition, eigenvector subspaces were defined consisting of only the first, the first two, the first three, the first ten and the first twenty eigenvectors and B-factors were calculated for these subspaces. Correlation of the eigenvector subspace B-factors with the complete

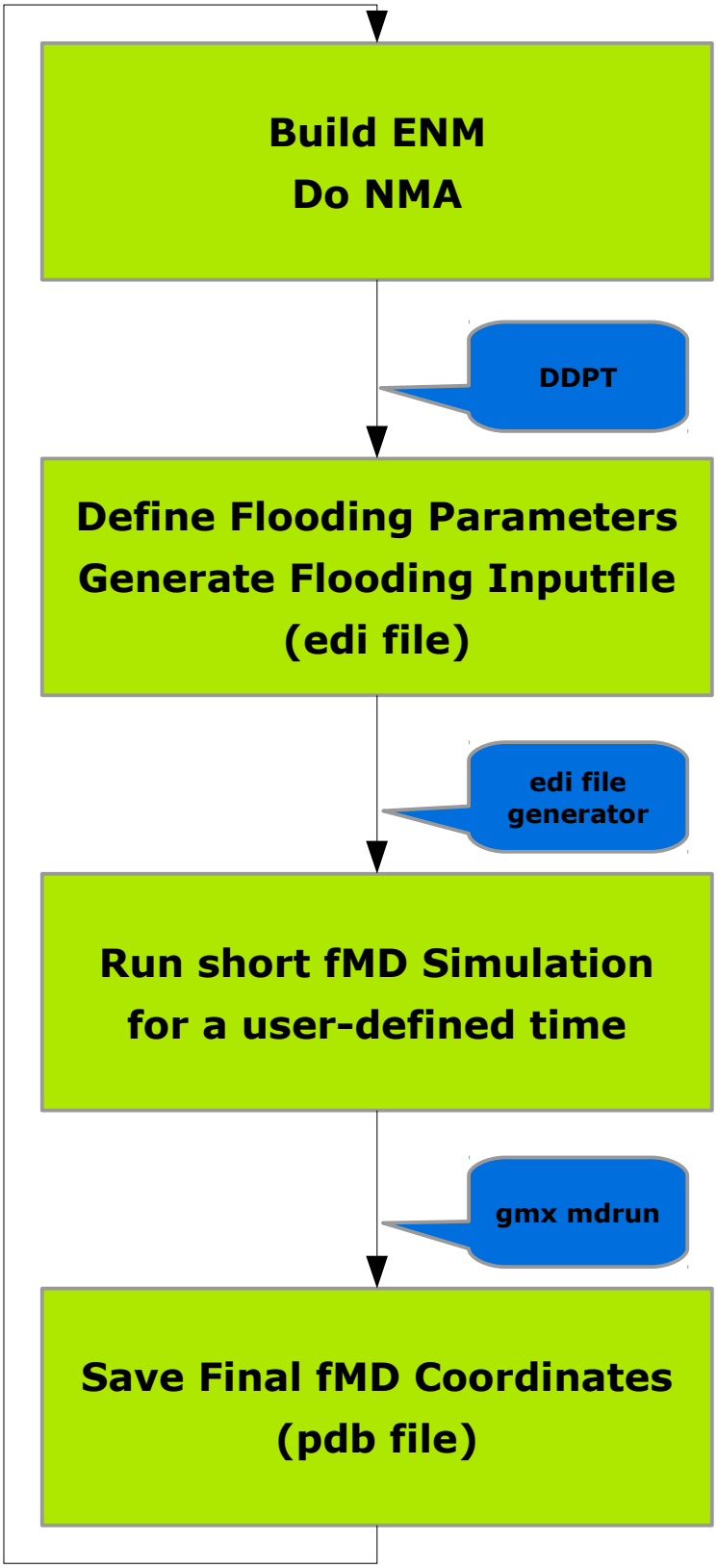



Figure 5: The work-flow of fMD simulations is delineated.

eigenvector space showed that the first eigenvector covers 84% of the protein's flexibility. As the first ten eigenvectors account for already 98% this led to the choice of the first ten non-trivial eigenvectors for defining the flooding potential.
Regarding the fact that applying the adaptive flooding mode for fMD simulations causes even more parameters to be defined, the constant flooding scheme was chosen for the current very first work applying the flooding technique to mGBP2. In this mode the constant flooding strength $E_{\mathrm{fl}}$ and the duration of one fMD simulation loop have to be further defined. To test the influence of the flooding strength $E_{\mathrm{fl}}$ eleven 5 ns fMD simulations were run at values for $E_{\mathrm{fl}}$ of 5, 8, 10, 12, 15, 20, 25, 35, 50, 75, 100 and 150 kJ/mol, using standard *mdp* file specifications as described in section 3.3.1. The $apo$-state of mGBP2 after equilibration served as start structure for the ENM and fMD simulations. From these simulations the flooding potential $V_{\mathrm{fl}}$ was plotted against time. As the flooding simulation causes the system to be driven away from its initial state, the Gaussian shaped flooding potential should drop remarkably after a while indicating that the conformation has changed. This characteristic is observed for the first few hundred ps for all flooding strengths. However, with growing $E_{\mathrm{fl}}$ the flooding potential does not approach zero and for longer simulation times it is seen that the flooding potential increases again. This indicates the return of the system to its initial conformation, which is not desired. Using the same test simulations, the appropriate duration of one fMD simulation loop has been determined. For a chosen flooding strength the simulation time has to be long enough to allow the structure to move away from its initial state – meaning the flooding potential to drop – yet has to be short enough to prevent the system from returning to the starting structure.
To identify the best choice for the flooding strength, the speed of departing from the initial state was estimated by calculating the root mean square deviation (RMSD) from the start structure for the first 250 ps for each flooding strength. The first 250 ps were chosen, because in that time the flooding potential dropped considerably for all test simulations. The fastest departures were observed for $E_{\mathrm{fl}}$ values of 10, 12, 15, 50 and 100 kJ/mol as shown in figure 6.
In conclusion, the insights from the RMSD and flooding potential analysis combined with the aim to alter the energy landscape as little as necessary, a flooding strength of 10 kJ/mol and 250 ps loops were identified as the most reasonable choice for the fMD simulations. This flooding strength leads to a considerable acceleration for moving away from the starting structure while the flooding potential rapidly dropps close to zero within 250 ps (figure 7).

## 3.4 Analysis

### 3.4.1 Essential Dynamics Analysis

The modes describing the essential dynamics of mGBP2 were either retrieved as eigenvectors from ENMs using the $\Delta\Delta$PT toolbox as described in section 3.2.1 or via PCA from MD simulations. PCA was done using the GROMACS program *gmx covar* with fitting of

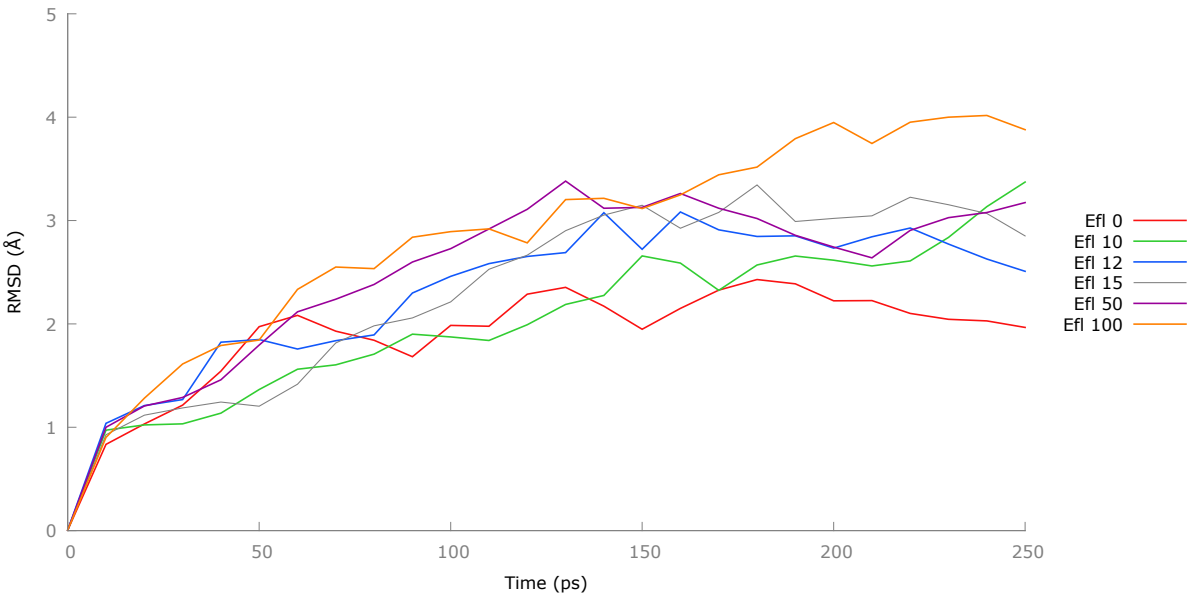



Figure 6: The RMSD in Å is plotted against the simulation time in ns for fMD simulations using different flooding strengths $E_{\mathrm{fl}}$ in kJ/mol.

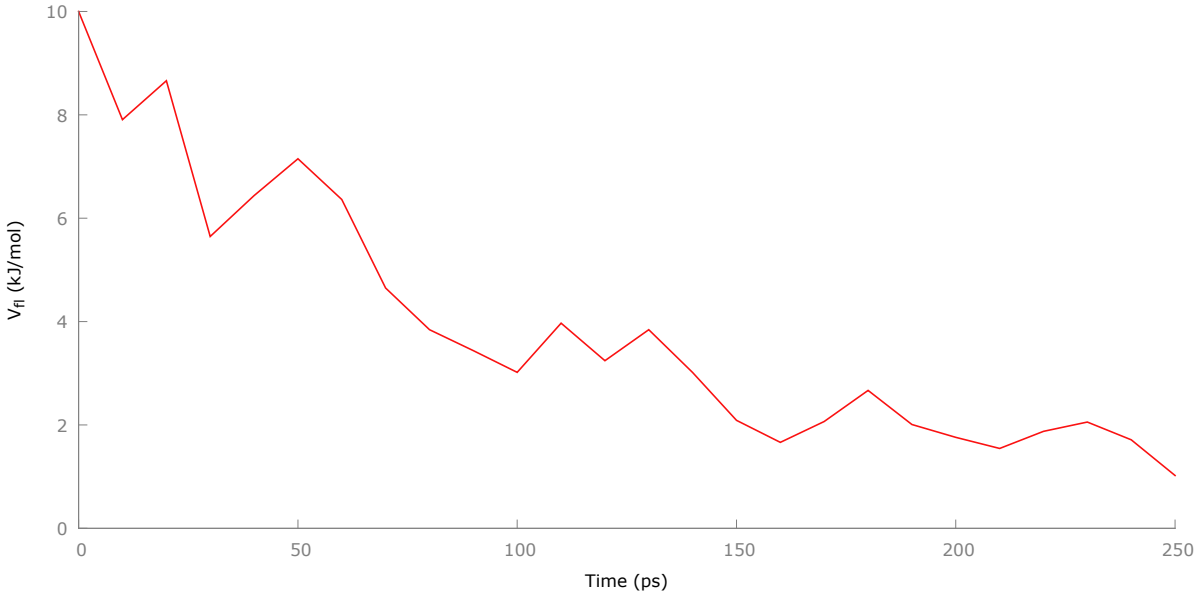



Figure 7: The flooding potential $V_{\mathrm{fl}}$ in kJ/mol of an fMD simulation is plotted against the simulation time in ns for a flooding strength of 10 kJ/mol.

$C_\alpha$-atoms to the start structure of each simulation and performing covariance analysis for $C_\alpha$-atoms without mass weighting to maintain comparability with ENM results. Principal components were visualized via the GROMACS tool *gmx anaeig* using the "*extr*" option.

### 3.4.2 Eigenvector Overlap

Overlaps between eigenvectors were calculated using the ΔΔPT OVERLAP tool. It calculates the overlap between two normalized eigenvectors via equation (14):[97]

$$I(v_1, v_2) = \frac{\left| \sum_{i=1}^{N} \frac{e_i(v_1) \times e_i(v_2)}{|e_i(v_1)| \times |e_i(v_2)|} \right|}{N} \tag{14}$$

with $I(v_1, v_2)$ denoting the overlap between eigenvectors $v_1$ and $v_2$, $N$ giving the number of atoms, and $e_i$ containing the three Cartesian coordinates $(x, y, z)$ of atom $i$ of the given eigenvectors.

### 3.4.3 Root Mean Square Deviation

Root mean square deviations (RMSD) are calculated for MD simulations via the program *gmx rms* of the GROMACS package with least squares fitting of $C_\alpha$-atoms to the start structure and RMSD calculation for $C_\alpha$-atoms. The calculation is based on the following equation (15):[115]

$$RMSD(t) = \sqrt{\frac{\sum_{i=1}^{N} m_i \left| r_i(t) - r_i^{ref} \right|^2}{M}} \tag{15}$$

where $m_i$ denotes the mass of atom $i$, $M = \sum_i m_i$, $r_i$ its position at time $t$, and $r_i^{ref}$ its corresponding position in the reference structure (chosen to be the start structure as explained above).

### 3.4.4 Root Mean Square Fluctuation and B-Factors

Root mean square fluctuations (RMSF) were calculated using the *gmx rmsf* tool from the GROMACS software after least squares fitting of $C_\alpha$-atoms to the start structure of each simulation choosing $C_\alpha$-atoms for output. Results were assigned to the corresponding residue using the "*-res*" flag. RMSFs are based on equation (16):[116]

$$RMSF^i = \sqrt{\frac{\sum_{t=1}^{t_f} \left| r_t^i - r_{ref}^i \right|^2}{t_f}} \tag{16}$$

Here, $t_f$ gives the total time over which one wants to average, $r_t^i$ defines the position of atom $i$ at time $t$ and $r_{ref}^i$ marks its reference position (chosen to be the start structure as explained above).

Temperature or so-called B-factors $B_i$ are derived from RMSFs via equation (17):[117]

$$B_i = \frac{8\pi^2}{3} RMSF_i^2 \tag{17}$$

For ENMs the B-factors were obtained from the RMS/COL program of $\Delta\Delta$PT as described in section 3.2.1.

### 3.4.5 Distance Measurements

Distances were measured using the GROMACS *gmx distance* tool after defining desired distance pairs via the *gmx make_ndx* program. For distance histograms the bin width was set to 0.1 Å.

To compare mGBP2 simulation results to data from FRET experiments of hGBP1, distances had to be further computed using the FRET-restrained positioning and screening (FPS) toolkit from Seidel and co-workers[118]. This was done to consider the fact that FRET experiments do not measure the distance between atoms of the protein itself, but between the donor (D) and acceptor (A) dyes which are attached to the protein via linker molecules. If a structure from MD simulations is parsed to the FPS tool, the latter allows to simulate the spatial distribution of donor and acceptor dyes after defining the attachment point of the dyes to the protein and the parameters for linker length and diameter, as well as the dye radius to account for the specific conditions of the FRET experiment. This results in distance distributions and by calculating the mean distance between donor and acceptor, $\langle R_{DA} \rangle$ according to equation (18)[15] the distance from MD simulations is comparable to the experimental results. A more detailed description of the tool and the underlying theory can be found in the FPS publication[118] and at *http://www.mpc.hhu.de/software/fps.html*.

$$\langle R_{DA} \rangle = \left| \langle \vec{R}_D^{(i)} - \vec{R}_A^{(j)} \rangle \right| = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} \left| \vec{R}_D^{(i)} - \vec{R}_A^{(j)} \right| \tag{18}$$

Here $\vec{R}_D^{(i)}$ and $\vec{R}_A^{(j)}$ denote the simulated positions of donor and acceptor, respectively. Donor and acceptor distances are calculated for each data point $i$ and $j$.

For the FRET experiments with hGBP1, ten double labeled species were designed by mutating pairs of amino acids to cysteines (CYS) and labeling them with dyes through attachment via linker molecules at the $C_\beta$-sulfur[15]. Table 1 shows the position of the FRET labels in hGBP1, as well as the corresponding positions in mGBP2 while in figure 8 these positions are visualized for mGBP2. Corresponding amino acids were decided based on the sequence alignment of mGBP2 with hGBP1. The results of this alignment can be found in the appendix (section 6.2).

Table 1: Positions of FRET labels in hGBP1 and corresponding positions in mGBP2.

| hGBP1 | mGBP2 |
|---|---|
| 496 - 540 | 494 - 538 |
| 18 - 344 | 18 - 342 |
| 344 - 496 | 342 - 494 |
| 344 - 525 | 342 - 523 |
| 344 - 540 | 342 - 538 |
| 18 - 577 | 18 - 575 |
| 254 - 540 | 252 - 538 |
| 18 - 540 | 18 - 538 |
| 254 - 344 | 252 - 342 |
| 254 - 577 | 252 - 575 |

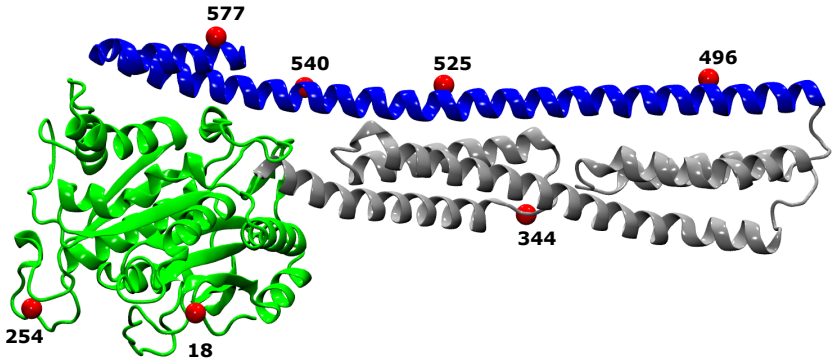



Figure 8: FRET labeling positions are visualized in the mGBP2 model as red spheres given the residue indices for hGBP1. The G domain of mGBP2 is shown in green, the M domain in silver and the E domain in blue.

As MD simulations were done using unmutated mGBP2, the $C_\beta$-atoms of the corresponding amino acids were defined as attachment point in the FPS software. After labeling hGBP1 for FRET it was not checked, which of the two cysteines carried the donor and which one the acceptor. Therefore, it is not known whether the order of dyes is "DA" or "AD". Thus, in the FPS software simplified parameters had to be defined: The linker length was set to 20.5 Å, the linker diameter to 4.5 Å and the dye radius was set to 5.0 Å for both donor and acceptor.

### 3.4.6 Clustering

Clustering of MD simulations was done using the GROMACS cluster tool applying the *gromos* method. Trajectories were fitted to the start structure with respect to $C_\alpha$-atoms, the same is valid for RMSD calculations. The RMSD cutoff for clustering was set to 2 Å and output structures were written including only $C_\alpha$-atoms.

### 3.4.7 Hydrogen Bonds and Salt Bridges

To check whether non-bonded interactions influence or stabilize the conformations of mGBP2, salt bridges and hydrogen bonds (H-bonds) were evaluated.
Salt bridges were defined to be existent if the distance between two non-bonded charge groups was smaller than 4.5 Å. H-bonds were examined using the GROMACS *gmx hbond* tool. The cutoff distance was set to 3.5 Å and the angle between donor and acceptor was restricted to a maximum of 30° using the GROMACS convention of a stretched donor acceptor angle being 0°. Nitrogens were included to act as H-bond acceptors.

# 4 Results and Discussion

## 4.1 Examining ENM, cMD and fMD results for $apo$-mGBP2

Before turning to the main focus of this work – gaining new insights in the dynamics and structural details of the $apo$- and $holo$-state of mGBP2 and its implications for future experimental studies – in section 4.2, the applied modeling tools are examined to identify their potential in coping with the aforementioned aim. The different approaches are compared to scrutinize their influence on results and their consistency. The cMD simulation results are furthermore compared to experimental data from high performance Förster Resonance Energy Transfer (hpFRET) of the hGBP1 homologue to validate simulation results. All analyses in this section concentrate on results for $apo$-mGBP2.

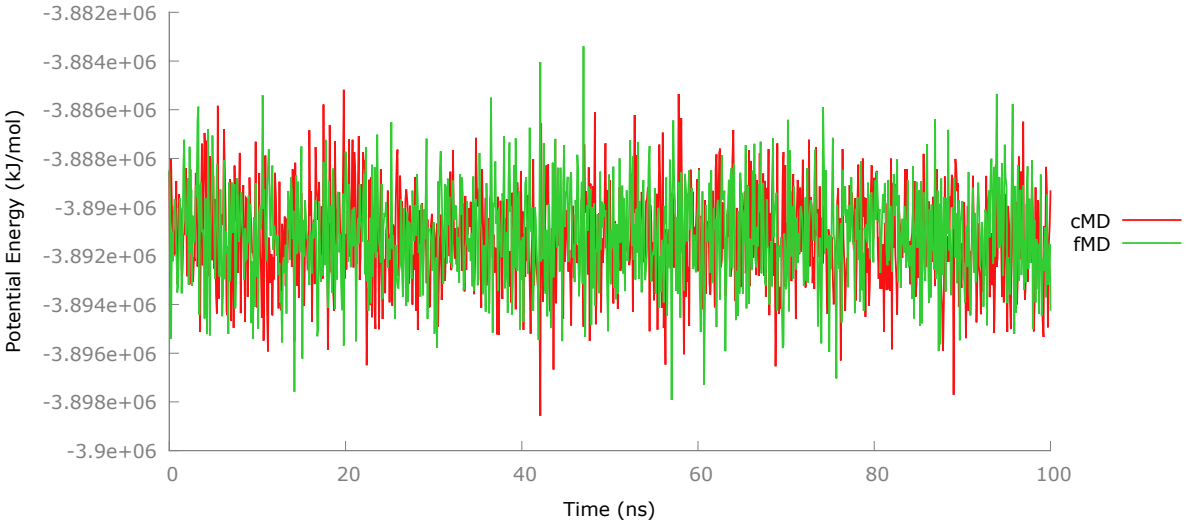
### 4.1.1 cMD and fMD yield stable simulations

Before starting with any detailed analysis of molecular modeling results, the stability of the MD simulations has to be examined. First insights can be gained from visual inspection of the MD trajectories in VMD: for both the cMD and fMD simulation of $apo$-mGBP2 the resulting trajectories look reasonable. To go beyond the visualization, the potential energy of mGBP2 (figure 9(a)) and the system's temperature (figure 9(b)) are plotted against the time. The temperature fluctuates $\pm 2$ K around the desired value of 298 K. Together with the potential energy, which does not drift with time, this assures the simulations to be stable.

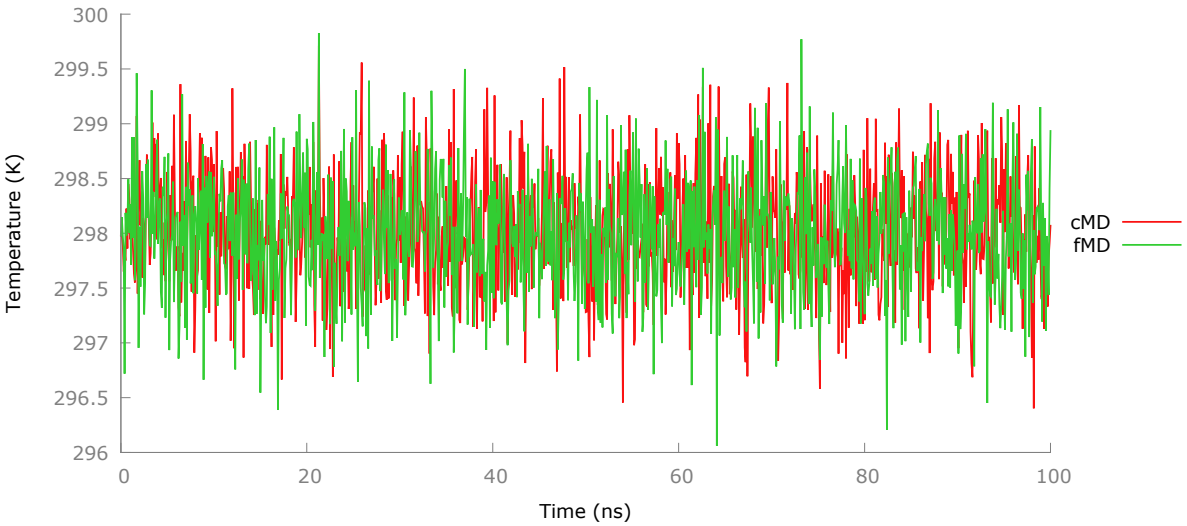
A further indicator for the soundness of a protein structure during an MD simulation is the RMSD. Figure 10 shows the RMSD against the time. The RMSD was calculated as explained in section 3.4.3.
For both the cMD and fMD simulation, the initial state is left rapidly, which is indicated by a rise of RMSD to about 4 Å after around 2 ns. Then the sampled conformations fluctuate between RMSDs of 3 to 7 Å for the fMD simulation and between 3 to 8 Å for the cMD simulation. This indicates that the overall tertiary protein structure is stable and unfolding processes are not likely. While this is to be expected for a folded protein of this size with the simulation being started from a homology model of a crystal structure, it also confirms the aforementioned conclusion that the set-up of the mGBP2 simulation system was successful leading to stable molecular dynamics simulations.

### 4.1.2 ENM, cMD and fMD predict similar collective, essential dynamics

One important objective of the MD simulations of mGBP2 is to unravel its global, function-related motions to identify distinct structural states of the protein and their transitions, which may lead to a deeper understanding of the biological role and functionality of mGBP2. Delivering first insights on the shape of these motions is furthermore the predominant purpose of ENMs. Therefore, having confirmed the stability of MD simulations, a first




(a) Potential energy of cMD and fMD simulations.




(b) Temperature of cMD and fMD simulations.

Figure 9: The potential energy of $apo$-mGBP2 in kJ/mol (a) and the temperature of the system in K (b) are plotted against the time in ns. The cMD simulation data is shown in red, the fMD data in green.

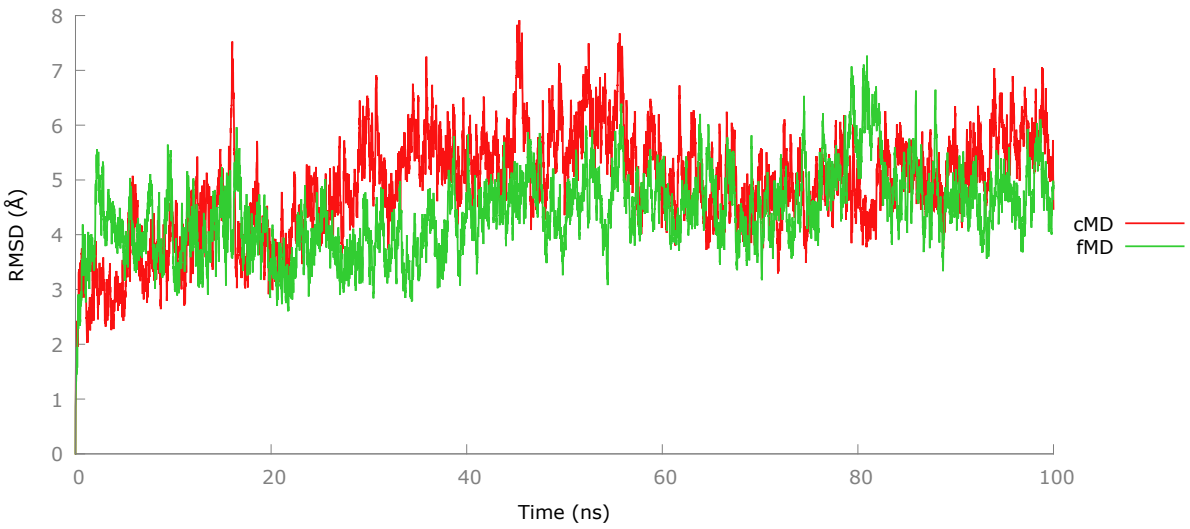



Figure 10: The RMSD of $apo$-mGBP2 (in Å) is plotted against the time in ns. The cMD simulation is shown in red and the corresponding fMD simulation in green.

important analysis is to compare the essential dynamics predicted by ENM and MD, which are presented by the corresponding eigenvectors (see section 3.4.1). Comparison is done by calculating the overlap between the eigenvectors as described in section 3.4.2 and via visualization in VMD.

Table 2 shows the overlaps between the first two non-trivial eigenvectors (EV) from ENM, cMD and fMD for nucleotide-free mGBP2. Regarding these overlaps cMD and fMD simulations deliver similar eigenvectors for the first two principal components, which represent the most collective ones. This can be confirmed by visual inspection in VMD. The first collective mode describes a twisting motion of the beginning of the $\alpha$12-helix and M domain helices, including a countermotion in the G domain with the biggest movement occurring in the guanine cap and the switch I and II regions (figure 11(a)). The second collective mode shows a kinking of the beginning part of the $\alpha$12-helix together with the M domain helices towards the G domain where again the largest countermotion comes from the guanine cap (figure 11(b)).
The VMD visualization conveys the impression that the fMD simulation enhances the motion along the collective modes, as local fluctuations and movements of single loops are barely present compared to the cMD simulation. In the latter, however, the overall motion is higher. The RMSD in figure 10 shows that the cMD simulation leads to larger structural changes than the fMD simulation marked by the former's larger RMSD range. The fMD simulation seems to stay a bit closer to the initial state in the beginning moving away from it later on, whereas the cMD simulation rapidly leaves the starting conformation but then stays at the same RMSD level as the fMD simulation in the second half of the trajectory. These findings could support the hypothesis of a more narrow, more directed sampling along the collective modes in the fMD simulation. But yet the differences in RMSD are small!

Moreover, from overlapping the cMD principal components with the normal modes from ENM one finds the same motions among the first two modes but the order being reversed. The same is true for fMD and ENM overlaps, which is further approved by visual inspection in VMD.

In summary regarding the eigenvector overlaps and the type of motions, the three models show a very good agreement concerning the predicted essential dynamics of mGBP2.

Table 2: Eigenvector overlaps between the first two collective modes for cMD, fMD and ENM of nucleotide-free mGBP2. The largest overlap is highlighted in green with "bold" writing.

| fMD \ cMD | EV 1 | EV 2 |
|---|---|---|
| EV 1 | **68** | 16 |
| EV 2 | 27 | **80** |

(a) cMD-fMD overlaps.

| ENM \ cMD | EV 1 | EV 2 |
|---|---|---|
| EV 1 | 22 | **85** |
| EV 2 | **71** | 19 |

(b) cMD-ENM overlaps.

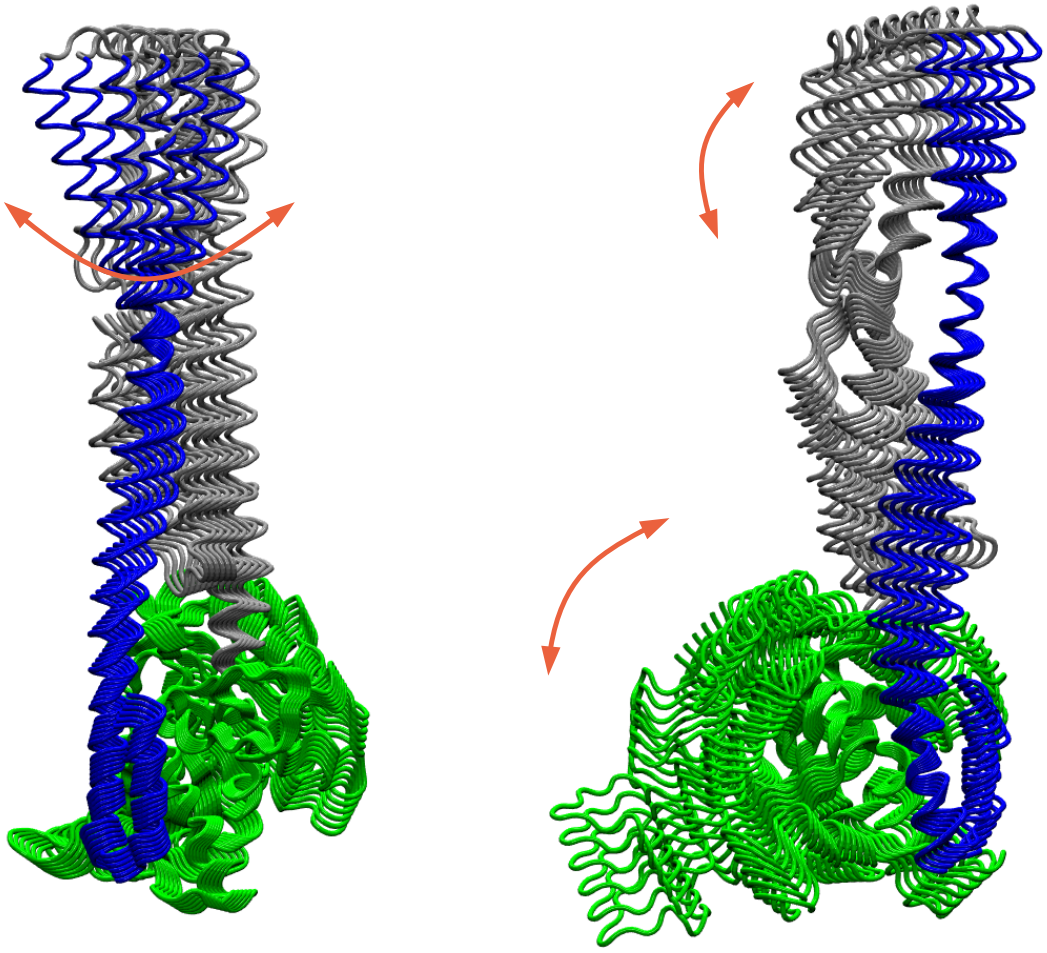
| ENM \ fMD | EV 1 | EV 2 |
|---|---|---|
| EV 1 | 0 | **94** |
| EV 2 | **63** | 17 |

(c) fMD-ENM overlaps.

### 4.1.3 ENM, cMD and fMD predict similar protein flexibility

First insights on the essential dynamics of $apo$-mGBP2 were gained by NMA of an elastic network model and PCA of cMD and fMD simulations. Discovering the flexible parts of the protein can further detail the knowledge about dynamic aspects of mGBP2. To this end, B-factors were calculated as explained in section 3.4.4.

Figure 12 shows the relative B-factors for the elastic network model and the cMD and fMD simulations of $apo$-mGBP2. Relative B-factors were obtained by scaling them to the maximum B-factor of each model to ensure comparability.
The characteristics of the protein flexibility – meaning the maxima positions – are in good agreement for all three models. cMD and fMD predict almost the same flexibility, which is supported by the high correlation of B-factors (89%). This is a first important result as this excludes the chance of fMD suffering from the "tip effect" caused by an overestimation of the flexibility of surface exposed loops by ENM[119]. Comparing ENM flexibility to the MD simulations one still finds a correlation of 68%. All models find higher B-factors in the G domain and the maximum flexibility is found for the guanine cap (residues 236 to 253). Furthermore, the switch I and II regions, the $\alpha$4'-region, the loop between M and E domain, and the $\alpha$13-helix are discovered to be very flexible. Besides that two more regions are noteworthy: The loops around amino acids 159 to 162 and 188 have not been of interest in previous mGBP2 studies. However, they show a very high mobility and possible relations to protein function will be discussed later (see section 4.2).
Two main differences are observed between MD and ENM results:



(a) First collective mode.

(b) Second collective mode.

Figure 11: The first two collective modes of *apo*-mGBP2 are shown following the mode order from PCA. The orange arrows mark the main direction of the motion of each mode. The G domain of mGBP2 is colored in green, the middle domain in silver and the E Domain in blue.

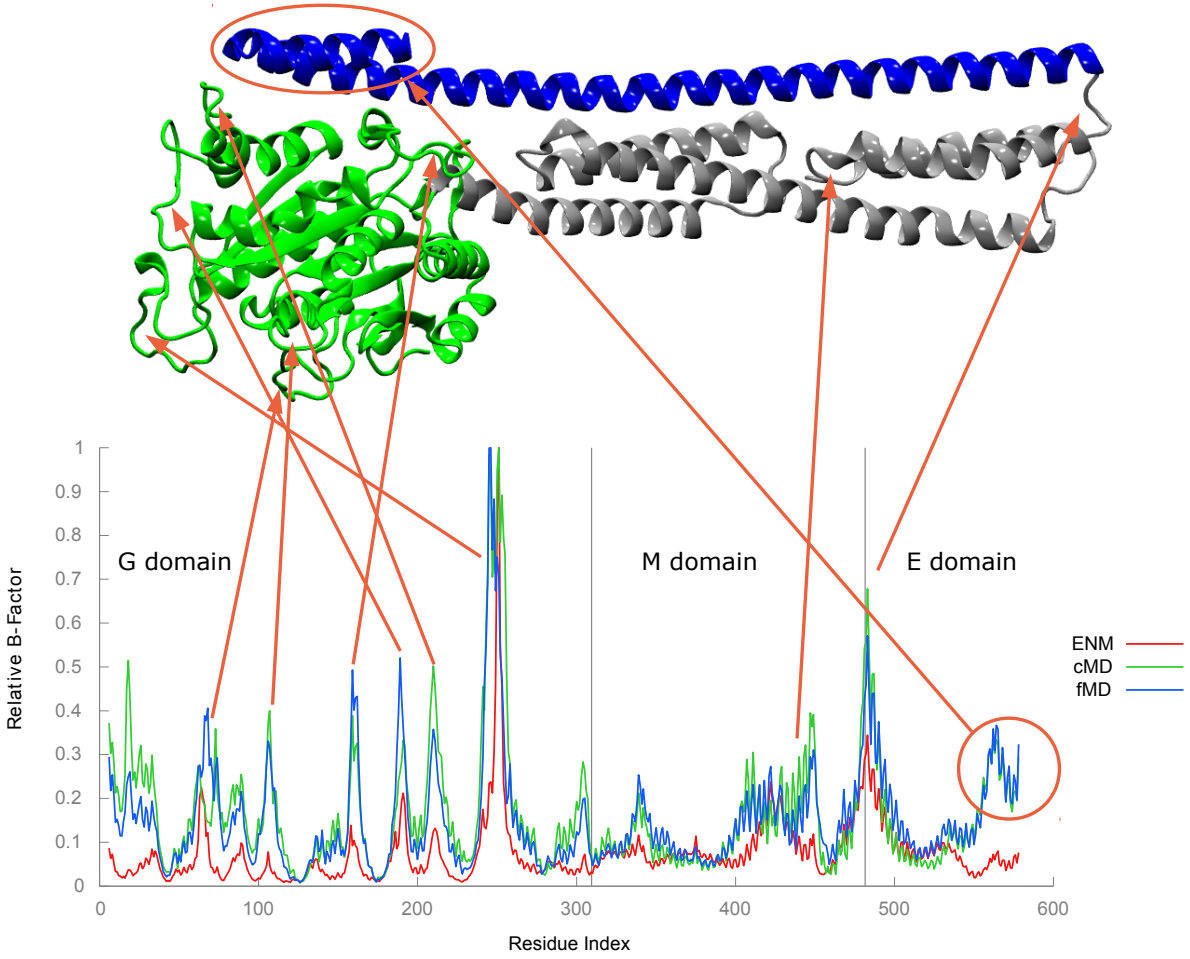



Figure 12: In the bottom half of the figure the relative B-factors from ENM (red), cMD (green) and fMD (blue) of nucleotide-free mGBP2 are plotted against residue indices. The orange arrows relate the maxima to the corresponding flexible parts of the mGBP2 structure shown in the upper half of the figure. The G domain of mGBP2 is colored in green, the M domain in silver and the E domain in blue. Vertical gray lines separate the three domains in the B-factor plot.

1. The ENM firmly concentrates the overall flexibility on the guanine cap whereas in MD simulations flexibility is widely spread over the protein.
2. The ENM predicts a remarkably lower flexibility of the $\alpha$13-helix than MD simulations do.

Summarizing the results obtained thus far, the ENM, cMD and fMD predictions on the protein mobility are in good agreement. Yet the conclusion drawn from the essential dynamics and RMSD analysis that the fMD simulation emphasizes protein flexibility along collective modes at the expense of local motions is not supported by B-factors. From the B-factors it is found that the flexibility spreads over the protein to a similar extend in both, cMD and fMD simulations.

### 4.1.4 cMD and fMD sample different conformational spaces

To gain deeper insights in the conformational sampling of the cMD and fMD simulations of $apo$-mGBP2 the trajectories are projected on the first two principal components from PCA (section 4.1.2). The results are shown in figure 13. The fMD simulation leads to an approximately circular sampling, while the cMD simulation reflects a broader sampling. This supports the former observations (section 4.1.2) of a more directed sampling along the collective modes in the fMD simulation.

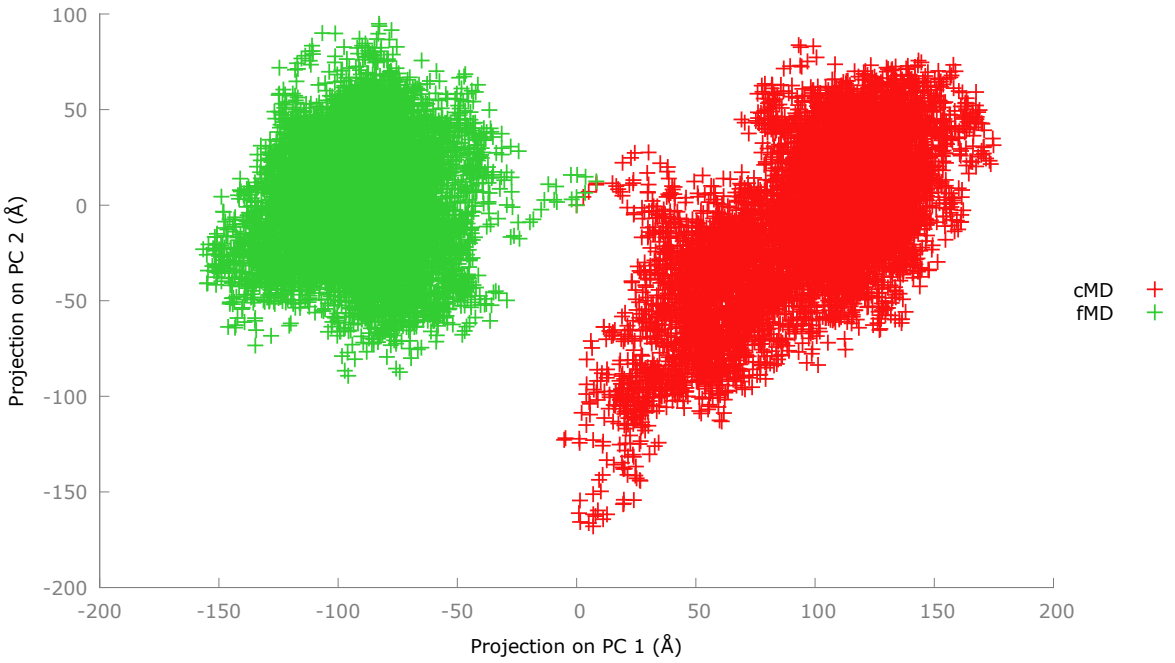



Figure 13: The plot shows the projection of the structures of $apo$-mGBP2 from the cMD (red crosses) and fMD (green crosses) trajectories onto the first two principal components. The eigenvector factors are given in Å.

Apparently the sampled conformational spaces differ: There is hardly any overlap between the cMD and fMD simulation. One reason could be that the PCs applied for the projection

are not the same. However, table 3(a) shows that there is considerable overlap between the PCs from cMD and fMD simulations. Another explanation might be that the sampling is insufficient as this is a great concern about MD simulations for systems as big as the one studied in this work. This possibility has to be taken into account in following analyses.

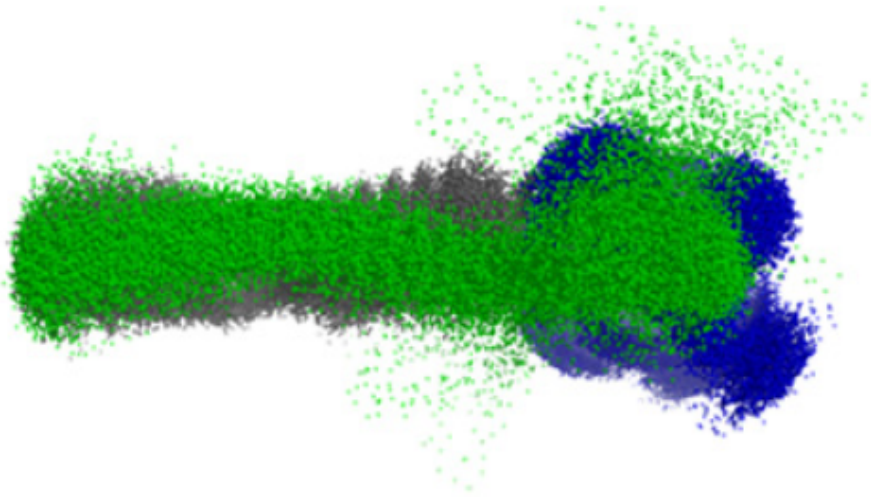
In summary, the results indicate different conformational samplings for cMD and fMD simulations. However, no conclusion can be drawn about whether the flooding simulation increases the sampling efficiency. Yet it conveys the impression of sampling the collective modes more firmly.

### 4.1.5 cMD of $apo$-mGBP2 samples major state from FRET experiments

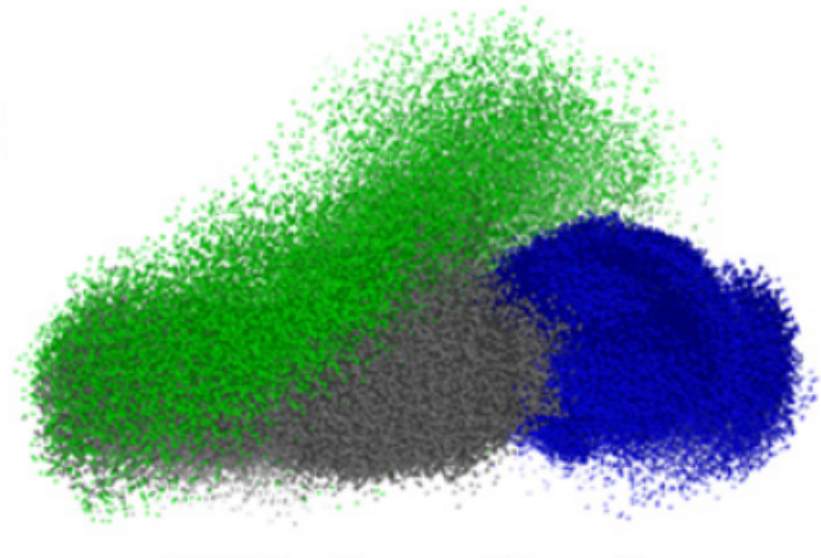
To finally validate the quality of the molecular modeling of $apo$-mGBP2, the cMD simulation results are compared to experimental hpFRET data for the hGBP1 homologue.
The positions of FRET labels in hGBP1 were shown in figure 8 (section 3.4.5), which defined the distance pairs in cMD analysis. The corresponding distances from cMD simulations were measured as explained in section 3.4.5 with the FPS software[118] determining the related mGBP2 residues to hGBP1 via sequence alignment as shown in the appendix (section 6.2). For easier comparison, all residue indices mentioned in this section refer to the hGBP1 sequence and the mGBP2 residue indices were translated accordingly. This is only valid for this section, all other sections use the original mGBP2 sequence numbering.
Single mGBP2 structures were parsed to the FPS software containing snapshots of the cMD simulation taken from the beginning, after 25, 50, 75 and 100 ns, as well as after 45.390 ns, which is the frame with the highest RMSD to the initial structure (around 7.9 Å).

Tables 3 and 4 contrast the findings from cMD distance analyses with the hpFRET data. The FRET data implicate that the crystal structure of hGBP1 cannot fully describe the protein in solution as two stable states are found: A major state resembling the crystal structure covers about two thirds of the population, whereas a minor state covering about one third of the population gives rise to a rotated G domain and a kinked $\alpha$12-helix (figure 14). This is supported by findings from SAXS.[15]

Based on the results from the distance measurements in the $apo$-mGBP2 structures taken from the MD simulation there are no indications that two conformational states were sampled. The average deviations of the cMD data from the hpFRET distances suggest that only the corresponding major state was sampled (figure 14(a)). Average deviations from the hpFRET major state are between 10% and 14%, which indicates a good agreement as hpFRET data describe hGBP1 and therefore a 100% agreement is not to be expected. This finding moreover confirms the stable and appropriate description of $apo$-mGBP2 by means of molecular modeling.
To ensure that the reason for not finding the minor state is due to not sampling it in the cMD simulation and not due to missing it because only distances in single snapshots



(a) Major state.



(b) Minor state.

Figure 14: The figures illustrate the two states found for $apo$-hGBP1 from hpFRET experiments. The G domain is shown in blue, the M domain in silver and the E domain in green color. The figures were gratefully obtained from the Seidel group[15].

Table 3: Distances between the given FRET pairs (hGBP1 counting of residues) of distinct structures from the cMD simulation of $apo$-mGBP2 are compared to the experimental results from hpFRET measurements of hGBP1.[15] The structure at 45.39 ns was chosen as it shows the greatest deviation from the initial state of the cMD simulation of $apo$-mGBP2 (RMSD of 7.9 Å) to cover the entire sampling.

| | | hpFRET distance $\langle R_{DA} \rangle$ [Å] | | deviation of cMD distances $\langle R_{DA} \rangle$ [%] | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | start structure | | at 25 ns | | at 45.39 ns | |
| hpFRET pair | | major | minor | Δ major | Δ minor | Δ major | Δ minor | Δ major | Δ minor |
| 496 | 540 | 60 | | 12 | | 26 | | 32 | |
| 18 | 344 | 70 | | 8 | | 13 | | 12 | |
| 344 | 496 | 43 | 36 | 17 | 40 | 21 | 44 | 22 | 45 |
| 344 | 525 | 48 | 33 | 14 | 25 | 18 | 19 | 10 | 30 |
| 344 | 540 | 57 | 45 | 14 | 10 | 19 | 2 | 10 | 14 |
| 18 | 577 | 58 | 30 | 6 | 104 | 4 | 85 | 6 | 104 |
| 254 | 540 | 58 | 43 | 18 | 60 | 11 | 20 | 1 | 34 |
| 18 | 540 | 65 | 47 | 2 | 41 | 1 | 39 | 1 | 37 |
| 254 | 344 | 90 | 61 | 5 | 40 | 26 | 10 | 6 | 39 |
| 254 | 577 | 70 | 55 | 11 | 41 | 2 | 30 | 1 | 29 |
| average deviation | | | | 11 | 38 | 14 | 29 | 10 | 38 |

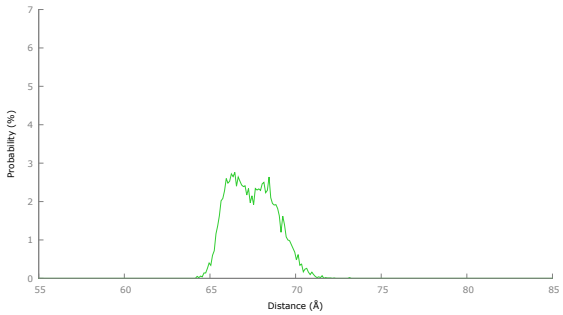
Table 4: Distances between the given FRET pairs (hGBP1 counting of residues) of distinct structures from the cMD simulation of $apo$-mGBP2 are compared to the experimental results from hpFRET measurements of hGBP1.[15]

| hpFRET pair | | hpFRET distance $\langle R_{DA} \rangle$ [Å] | | deviation of cMD distances $\langle R_{DA} \rangle$ [%] | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | at 50 ns | | at 75 ns | | at 100 ns | |
| | | major | minor | Δ major | Δ minor | Δ major | Δ minor | Δ major | Δ minor |
| 496 | 540 | 60 | | 22 | | 20 | | 25 | |
| 18 | 344 | 70 | | 10 | | 12 | | 8 | |
| 344 | 496 | 43 | 36 | 19 | 42 | 22 | 45 | 23 | 47 |
| 344 | 525 | 48 | 33 | 14 | 25 | 17 | 20 | 14 | 25 |
| 344 | 540 | 57 | 45 | 13 | 10 | 12 | 11 | 11 | 12 |
| 18 | 577 | 58 | 30 | 2 | 89 | 7 | 80 | 1 | 92 |
| 254 | 540 | 58 | 43 | 1 | 33 | 1 | 37 | 0 | 34 |
| 18 | 540 | 65 | 47 | 1 | 40 | 1 | 36 | 1 | 37 |
| 254 | 344 | 90 | 61 | 18 | 21 | 24 | 12 | 27 | 8 |
| 254 | 577 | 70 | 55 | 6 | 35 | 5 | 33 | 0 | 27 |
| average deviation | | | | 11 | 33 | 12 | 31 | 11 | 32 |

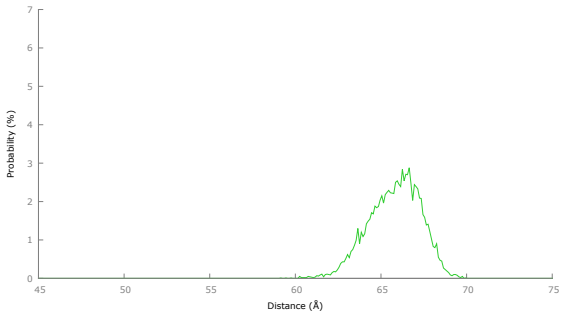
were determined, further analyses were done. First, the distance between residue pairs corresponding to the FRET pairs are analyzed for the whole cMD trajectory using the GROMACS *gmx distance* tool (see section 3.4.5). Additionally, a cluster analysis as described in section 3.4.6 was performed to categorize the sampled conformational space. Figure 15 shows the distance distributions of these residue pairs during the cMD simulation. The width of the distance distributions are mainly around 5 Å. Broader distributions of 10 Å are only found for pairs 254-540 and 18-577 (hGBP1 residue indices) and the broadest distribution is observed for the distance between amino acids 254 and 344 with 20 Å. This is in line with the hpFRET results, which also find the broadest distributions for these pairs. The distance distribution for FRET pairs 254-344 nicely reflects the kinked motion of the $\alpha$12- and M domain helices towards the G domain, which is reflected by the second principal component (figure 11(b)). However, all distribution plots are mainly centro-symmetric and therefore do not indicate the existence of two distinct states. The major and minor state predicted from hpFRET experiments mainly differ in the orientation of the $\alpha$12/13-helices (figure 14). The conformational space of the $\alpha$12/13-helices sampled in the cMD simulation is visualized in figure 16. It indeed reveals shows a motion in the direction of hpFRET findings and explains, why the distributions of the pairs 254-540 and 18-577 are broader. However, the motion is much smaller than predicted from experiment and together with the centro-symmetric distribution plot it rather depicts fluctuations around the major FRET state than the existence of two distinct equilibrium states. These new examinations are in line with the previous finding that the cMD simulation only samples the corresponding major state. Two possible reasons could explain this observation:

1. The cMD simulation of $apo$-mGBP2 is not converged and the sampling thus far only covers the major state.
2. The observation of a minor state in hpFRET data might be an artifact due to unexpectedly high interference of FRET labels with the protein structure.

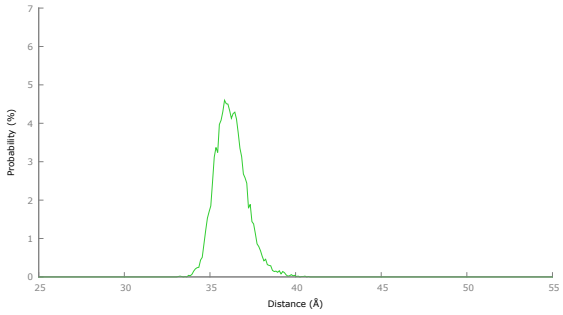
The first explanation would be in accordance with concerns already arising from trajectory projections on principle components (section 4.1.4) and has to be considered in future simulation studies. Yet the second explanation should also be evaluated in detail. FRET labels are attached to the protein after amino acid mutation to CYS including a linker molecule. These labels can spread over more than 25 Å, which is huge compared to local protein structures and the observed RMSDs. In addition, the residue mutation might change the local physico-chemical environment and lead to local changes of the protein structure. Thus, possible interference with protein structure should not be neglected. In particular in the case of $apo$-hGBP1 the labeling at residue 577 might have such effects: In this part of hGBP1 the $\alpha$12/13-helix is stabilized at the G domain surface via salt bridges (see figure 2). It is not impossible that the FRET label interferes with the protein structure here and reduces the salt bridge contacts, which would result in an increased flexibility of this part of the structure. Furthermore, the observation of a minor state for $apo$-hGBP1 would question



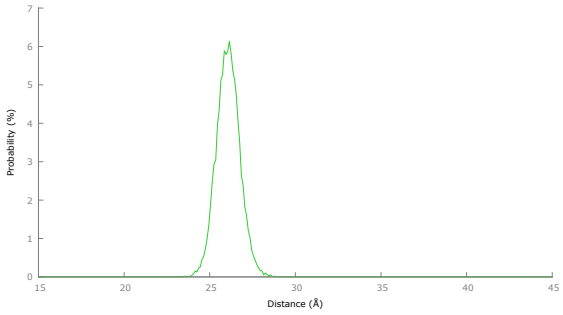

(a) Distance distribution of residue pair 496–540.



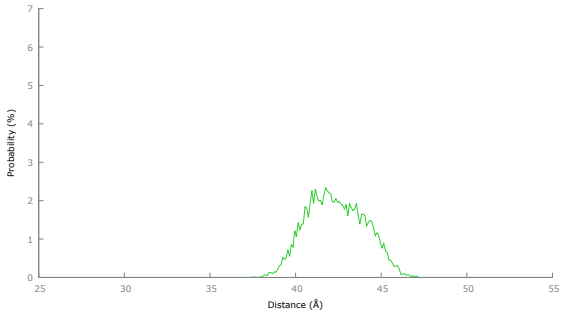

(b) Distance distribution of residue pair 18–344.



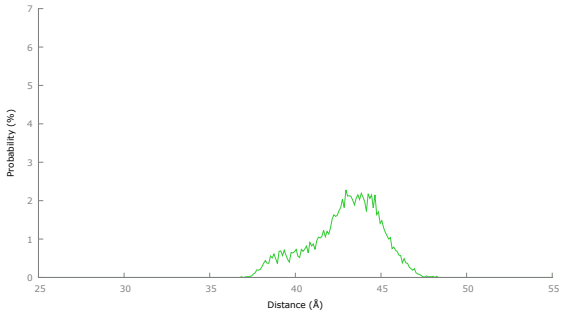

(c) Distance distribution of residue pair 344–496.



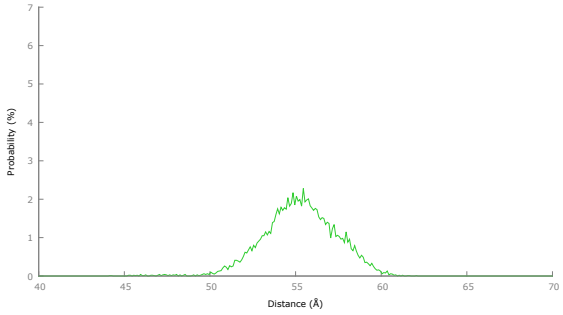

(d) Distance distribution of residue pair 344–525.



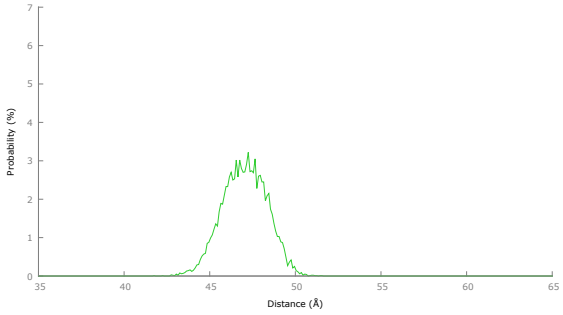

(e) Distance distribution of residue pair 344–540.



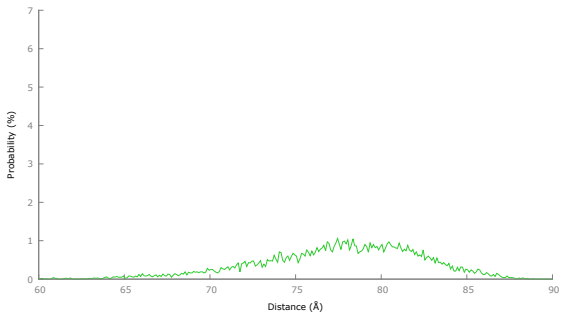

(f) Distance distribution of residue pair 18–577.



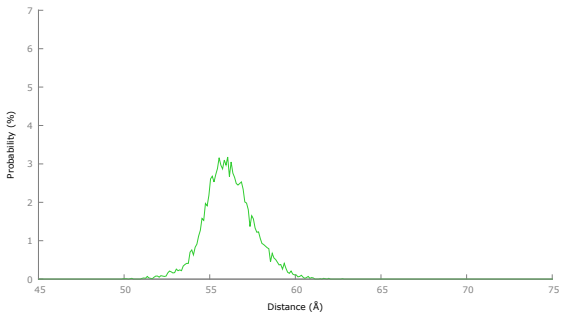

(g) Distance distribution of residue pair 254–540.




(h) Distance distribution of residue pair 18–540.




(i) Distance distribution of residue pair 254–344.




(j) Distance distribution of residue pair 254–577.

Figure 15: The plots show the distance distribution of hpFRET pairs during the cMD simulation of $apo$-mGBP2. Distances are measured in Å with the bin width being 0.1 Åwhile the occurrence is given in %. The residue indices are based on the hGBP1 sequence. To obtain the corresponding amino acid in mGBP2, one has to subtract 2 starting from hGBP1 residue 165.

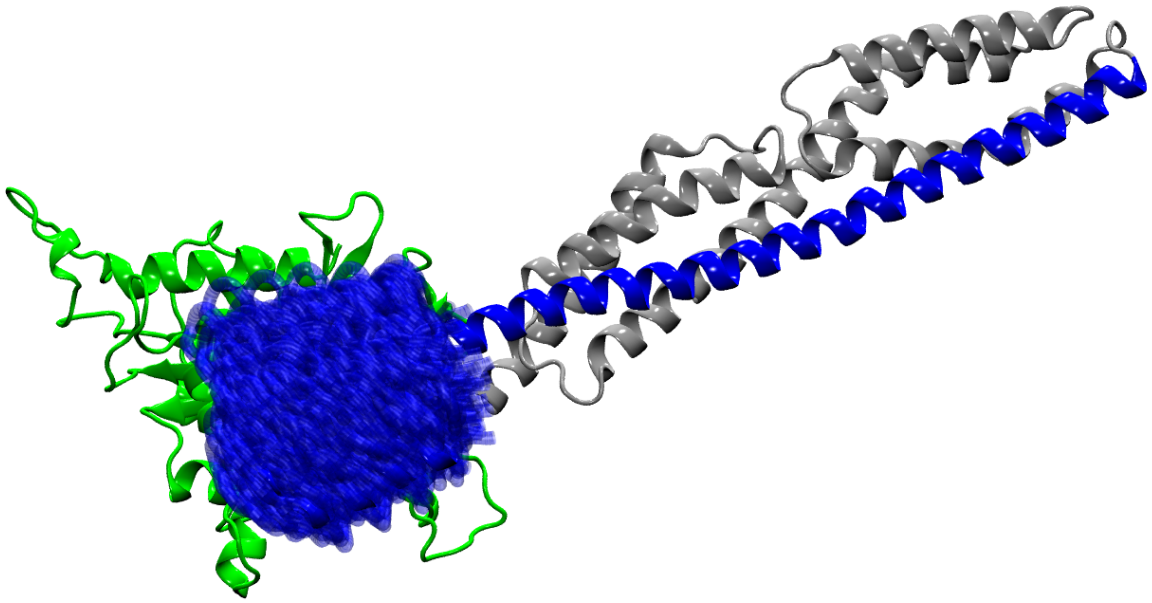


Figure 16: The figure visualizes the conformational space of the terminal part of the E domain as sampled in the cMD simulation of nucleotide-free mGBP2. To this end, the terminal E domain conformations from the middle structure of each cluster are overlaid. The E domain is shown in blue, the M domain in silver and the G domain in green.

recent biochemical models for this protein: It is assumed that the nucleotide binding activates hGBP1 and enables it only during hydrolysis to tetramerize or hetero-oligomerize via unveiling a new interaction surface in the E domain.[3,8,120] The finding of a minor state for $apo$-hGBP1 would already imply the possibility of a further interaction site in the $apo$-state. The question then arises, how the host would regulate the protein activity in case of the minor state predicted by hpFRET.

With regard to these findings MD simulations should study the FRET-labeled proteins in comparison to the native state to determine possible impacts of the FRET labels on the protein structure. In summary, the cMD simulation of $apo$-mGBP2 samples the corresponding major state observed in hpFRET experiments, which confirms the correct molecular modeling of mGBP2, but does not find the experimentally described minor state.

### 4.1.6 Interim Conclusion

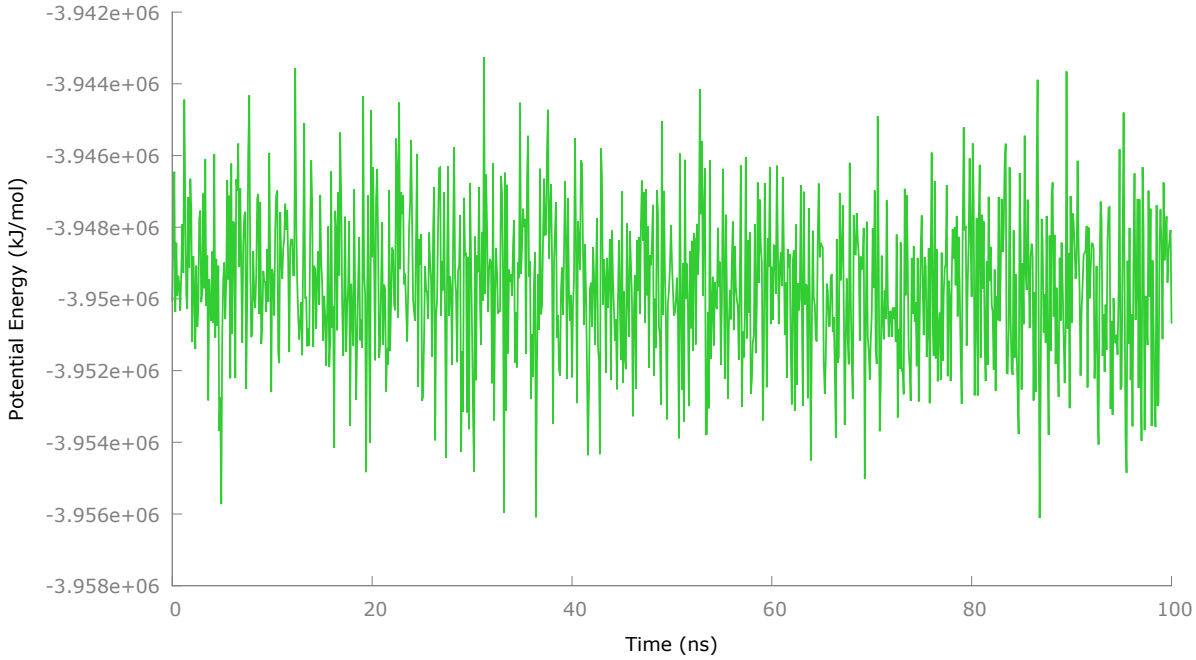
From the results for the $apo$-state modeling of mGBP2 from the ENM, as well as from the cMD and fMD simulations it could be shown, that the models are in good agreement with each other. Though a difference was found for the conformational spaces sampled by the cMD and fMD simulations, which might indicate an incomplete sampling. A further phenomenological study of the mGBP2 dynamics and the differences between the $apo$- and $holo$-state seems promising, as the cMD simulation was shown to correctly describe the corresponding major state of $apo$-hGBP1 from hpFRET experiments.
In order to now focus on the thorough examination of the differences between the $apo$- and $holo$-state and the comparison to experimental data, further analyses concentrate on the results from fMD simulations for the two mGBP2 states.

## 4.2 Comparing fMD results for the $apo$- and $holo$-mGBP2

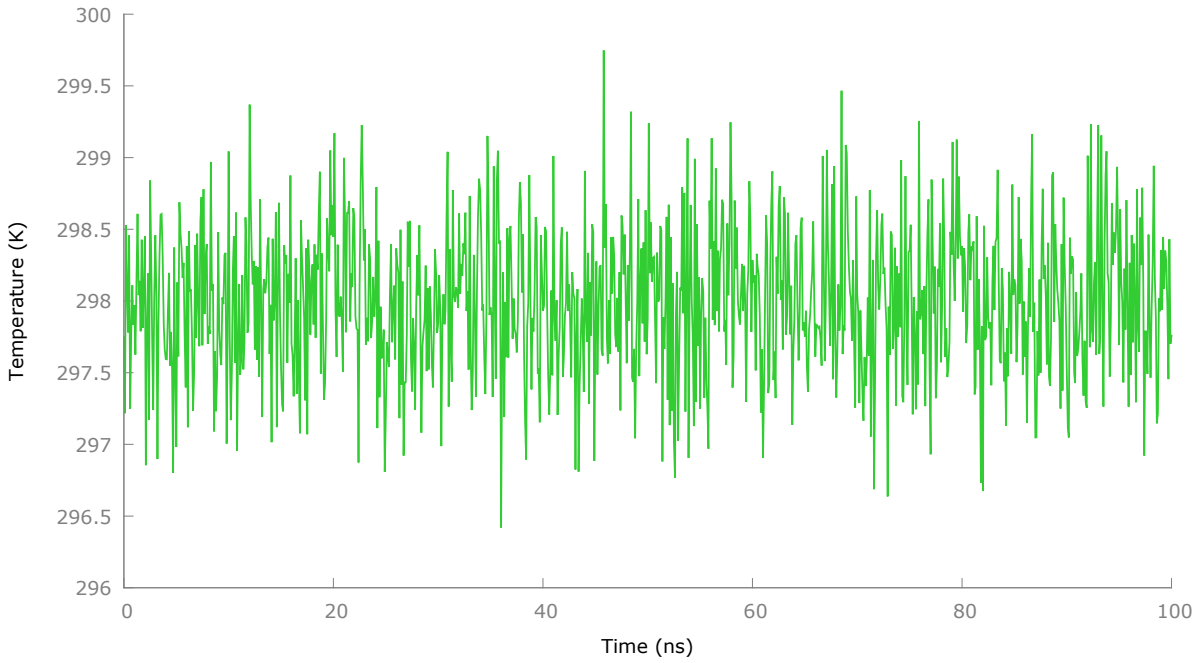
### 4.2.1 fMD simulations of $apo$-mGBP2 show stable binding of GTP

As for the simulations of $apo$-mGBP2, before going into analytical details the fMD simulation of $holo$-mGBP2 is to be examined regarding its ability to stably sample the nucleotide-bound state. This in particular includes the active site and binding of GTP.

Potential energy and temperature evolution indicate a stable simulation and successful thermostatting of the $holo$-state (see figure 17). Moreover, the RMSD plot (figure 18) portrays a similar development as found for the $apo$-state simulations, indicating a constant migration away from the initial structure without any disruptions of the protein.




(a) Potential energy.




(b) Temperature.

Figure 17: The potential energy of $holo$-mGBP2 in kJ/mol (a) and the temperature of the system in K (b) are plotted against the time in ns.

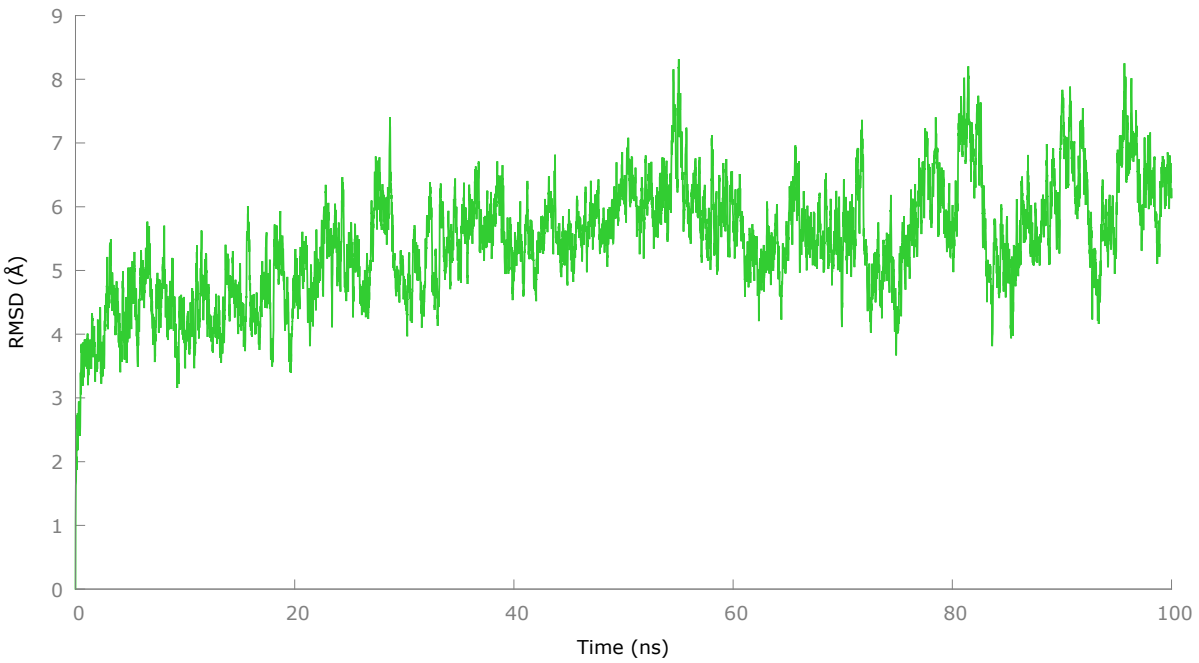
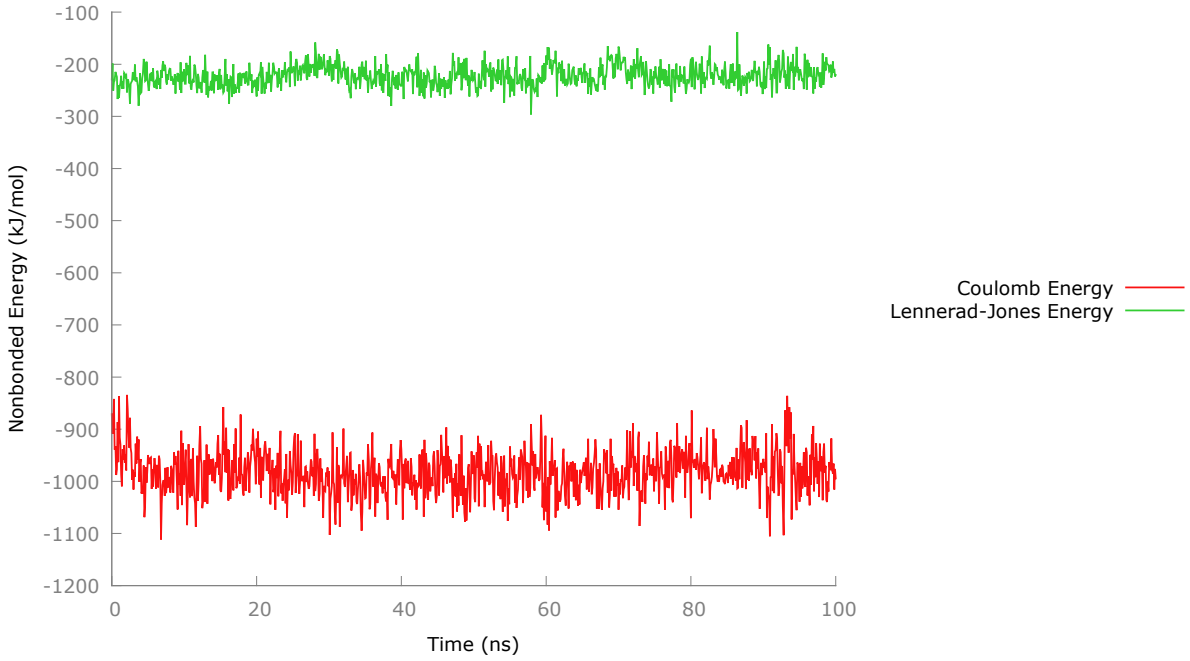
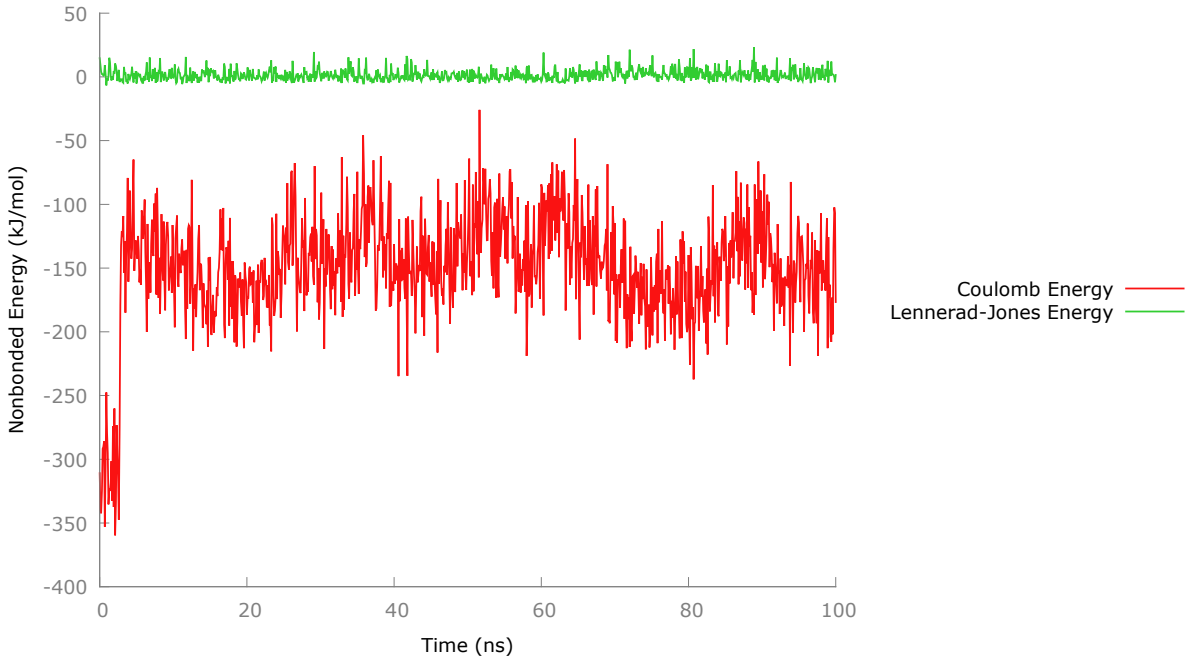



Figure 18: The RMSD in Å is plotted against the time in ns for the fMD simulation of nucleotide-bound mGBP2.

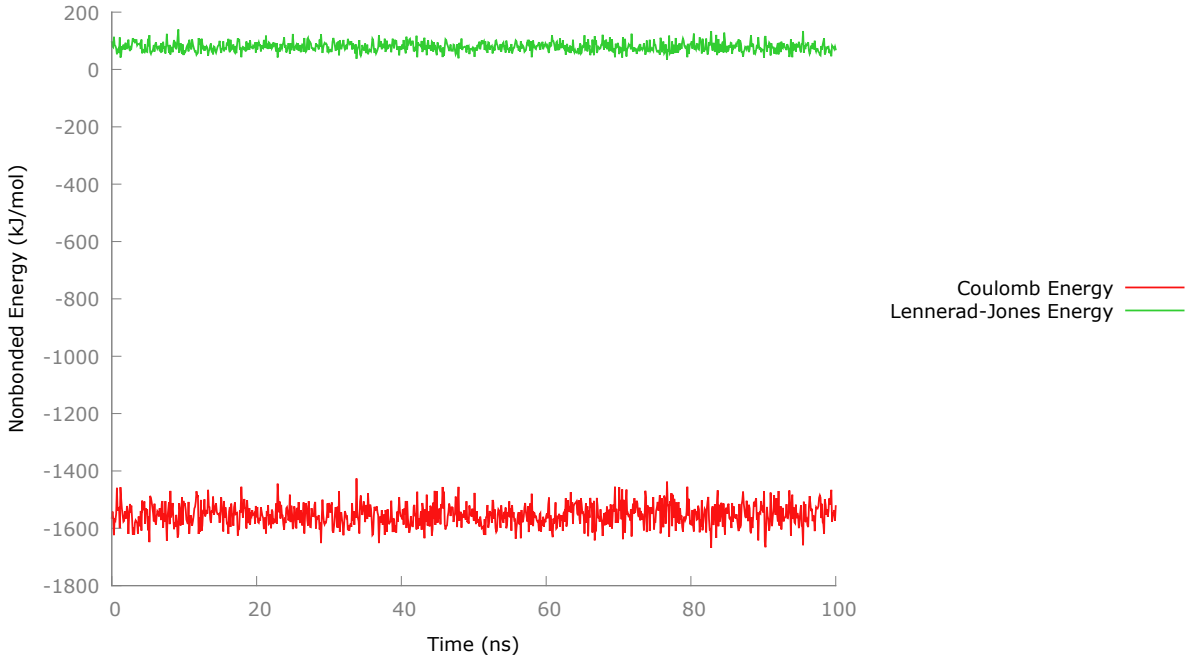
To check whether the nucleotide and co-factor are stably bound during the fMD simulation, the interaction energies between mGBP2 and GTP, mGBP2 and $Mg^{2+}$, and GTP and $Mg^{2+}$ are evaluated. Figure 19 shows these energies. The total interaction energies are all negative and show only little fluctuations indicating a stable binding of GTP and $Mg^{2+}$. The Coulomb interactions between GTP and $Mg^{2+}$ are by far the greatest, resulting from the fact that they are very close to each other without big fluctuations of the distance between them during the simulation as shown in figure 20(a). For the protein –$Mg^{2+}$ interactions (figure 19(b)) an interesting rise in energy from -350 kJ/mol to -100 kJ/mol between 2.6 ns and 3.2 ns was found. By examining the trajectory this can be explained by a movement of the THR75 side-chain, causing its $O_\gamma$-atom moving away from $Mg^{2+}$ to a new stable configuration.




(a) Non-bonded interaction energies between protein and GTP.



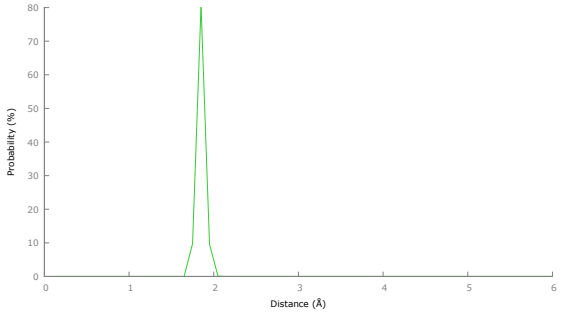

(b) Non-bonded interaction energies between protein and $Mg^{2+}$.



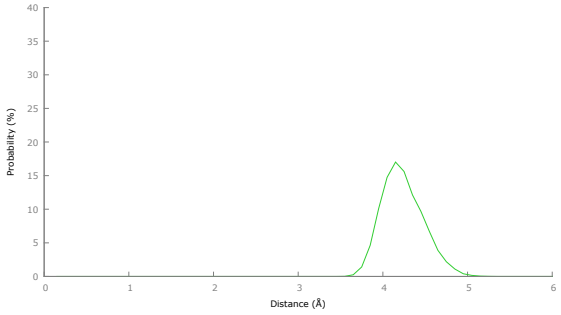

(c) Non-bonded interaction energies between GTP and $Mg^{2+}$.

Figure 19: Non-bonded interaction energies in kJ/mol are plotted against the time in ns for selected interactions of $holo$-mGBP2 during the fMD simulation. Non-bonded interactions consist of Coulomb interactions (red) and Lennard-Jones interactions (green).

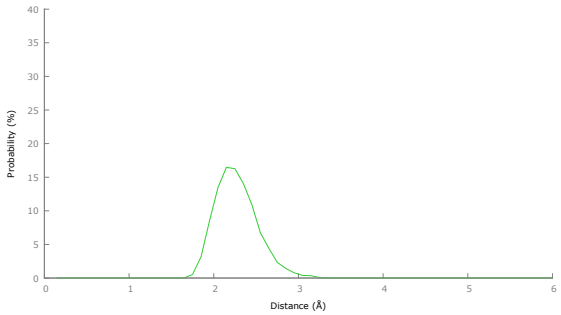
Finally distance distributions for atom pairs of GTP, $Mg^{2+}$ and mGBP2 are calculated from the fMD simulation of $holo$-mGBP2. They all show narrow distributions with the maximum being smaller than 5 Å. This approves the close and strong binding of nucleotide and co-factor in the active site of $holo$-mGBP2.



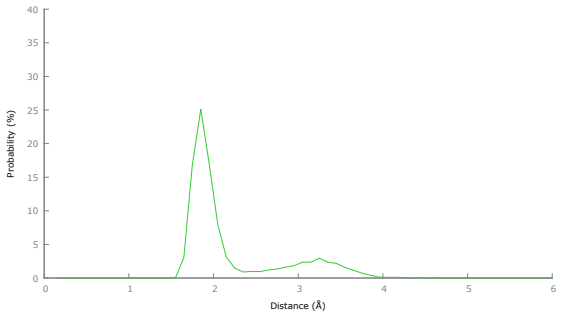

(a) Distance distribution between $Mg^{2+}$ and $O_\gamma$-GTP.




(b) Distance distribution between $Mg^{2+}$ and $H_N$-ARG48.




(c) Distance distribution between H-ARG48 and $O_\gamma$-GTP.




(d) Distance distribution between $O_\delta$-ASP182 and $H_N1$-GTP

Figure 20: The figures show the distributions for selected distances in $holo$-mGBP2 during the fMD simulation. Distances are measured in Å with a bin width of 0.1 Å, while the occurrence is given in %.

In summary, the fMD simulation is able to correctly characterize GTP binding of mGBP2 showing a stable binding pocket with constantly bound GTP and $Mg^{2+}$.

### 4.2.2 GTP binding stabilizes the G Domain of mGBP2

A very interesting question is, how GTP binding affects the structure and dynamics of mGBP2. Possible differences between the two states may lead to new insights in biological functions of mGBP2 and how the host regulates the functioning of mGBP2 via GTP binding and hydrolysis. Thus, the aim of the following investigations is to gain first answers via comparing the fMD simulations of the $apo$- and $holo$-state of mGBP2.

Differences in the flexibility and essential dynamics provide an understanding on the macroscopic changes induced by GTP binding. Figure 21 shows the absolute B-factors of $apo$- and $holo$-mGBP2. The eye-catching difference arising from nucleotide binding is the increase in rigidity of the whole protein, in particular of the G domain. The flexible parts are shifted from the G domain of mGBP2 to the M and E domain, especially to the $\alpha$13-helix, which is the most flexible motif now. The guanine cap, which is the most flexible part in the $apo$-state turns out to be rigid in the $holo$-state. Moreover, the switch I and II motifs

stiffened remarkably. This finding supports their biochemically expected importance in the hydrolysis of GTP[3] as a stable conformational arrangement of the active site is eminently important for function. On the other hand, the loop between the M and E domain stays very mobile. Interestingly, the loop around residue 188 (not having been of interest in experiments yet[3,8,14]) remains mobile, too, and represents the most flexible part of the $holo$ G domain.

The impacts of GTP binding on the guanine cap and the $\alpha$13-helix are moreover manifested in changes of the essential dynamics of mGBP2. Visualization of the principal components of $apo$- and $holo$-mGBP2 seem to be rather similar at first. The dominant collective motions of the $holo$-state also predict a twisting and kinking motion as described in section 4.1.2 for the $apo$-state (see figure 11). However, the eigenvector overlaps of the first two principal components are rather low (37% for EV1 and 12% for EV2) suggesting that there must be differences. The flexible part of the M and beginning E domain move similarly in $apo$- and $holo$-mGBP2, which already showed preserved flexibility in the B-factor analysis. The differences result from the part, which is responsible for the countermovement: While it is the guanine cap in the $apo$-state, in the $holo$-state the $\alpha$13-helix mainly acts as the counterpoint and, to a lesser extent, also the loop built of residues 155 to 170. This implies that the $holo$ E domain is most involved in the collective motions. These results are very well in line with the previous findings from B-factor analysis.

Reduced flexibility should go together with a reduced conformational space being discovered by the $holo$-state. Cluster analyses of $apo$- and $holo$-mGBP2 during the fMD simulations confirm this assumption. The differences between the two states are visualized in figure 22 by overlaying the middle structure of all found clusters for both fMD simulations and contrasting the already discussed highly influenced structural parts of mGBP2. They all demonstrate the same: GTP binding causes the G domain to become more rigid.

### 4.2.3 GTP binding causes great conformational changes in mGBP2

To further detail the structural changes of residues, which are greatly affected by GTP binding, their dominant conformations during the fMD simulations of $apo$- and $holo$-mGBP2 are overlaid. The dominant conformations are defined by the middle structure of the biggest cluster of each simulation, which accounts for 1/4 of the structures in the $apo$- and even for 1/3 in the $holo$-state.

The results are illustrated in figure 23. Accordingly GTP binding leads to the following major conformational changes:

- The guanine cap flips over to close the lower part of the binding pocket.
- The switch I region closes the binding pocket from the right side.

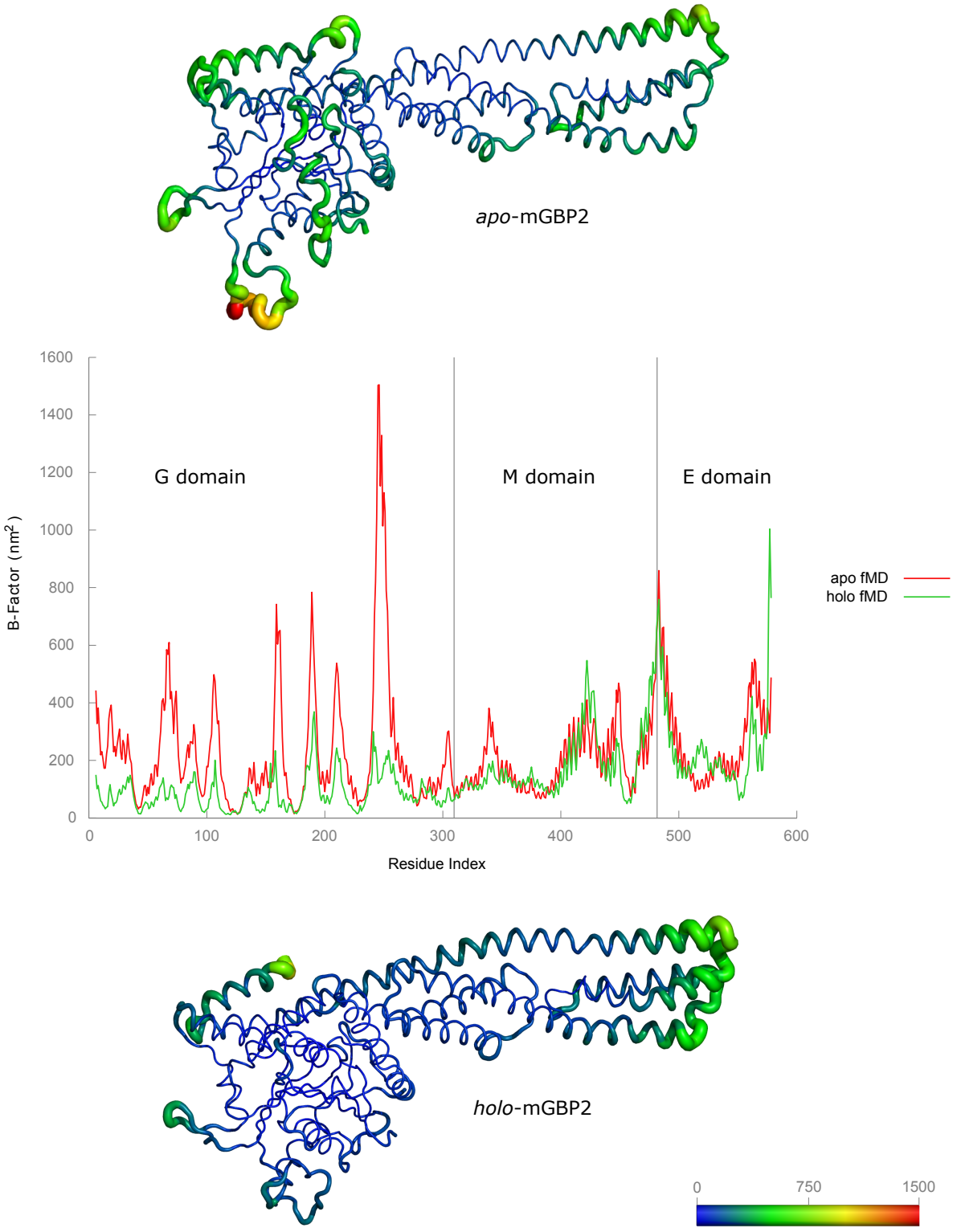
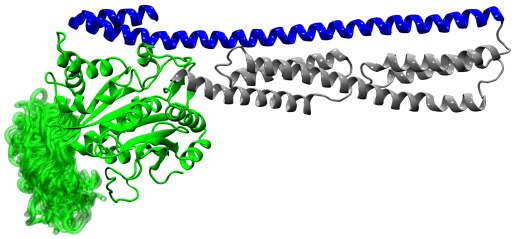
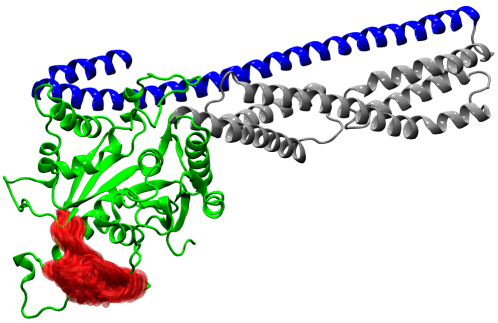


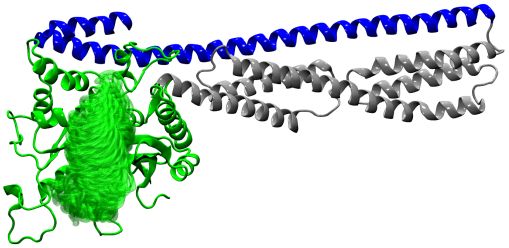

Figure 21: Absolute B-factors in nm$^2$ are plotted against the residue indices. Data from the fMD simulation of $apo$-mGBP2 is shown in red, the corresponding data of $holo$-mGBP2 in green. In the upper and lower parts of the figure B-factors are visualized in the mGBP2 structure. As underlying structure the middle structure of the biggest cluster of each simulation was chosen. Tube thickness and colors reflect the corresponding B-factors reaching from thin/blue (=0 nm$^2$, stiff) to thick/red (=1500 nm$^2$, flexible).



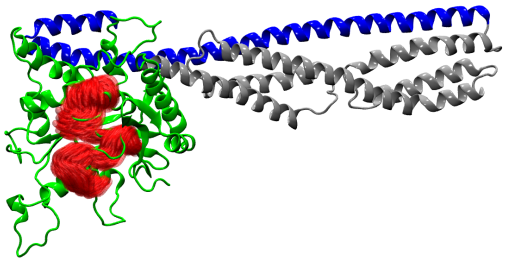
(a) Guanine cap of $apo$-mGBP2.



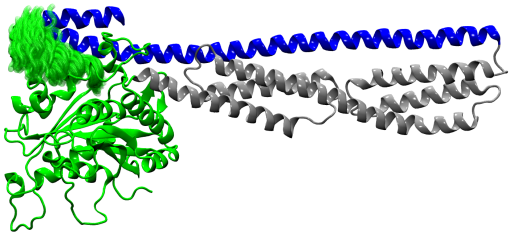
(b) Guanine cap of $holo$-mGBP2.



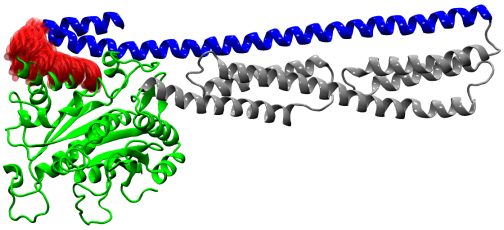
(c) Switch I and II of $apo$-mGBP2.



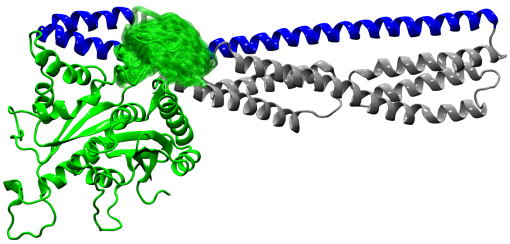
(d) Switch I and II of $holo$-mGBP2.



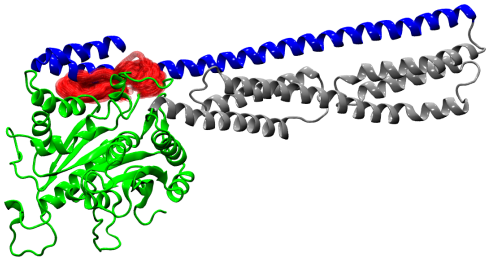
(e) $\alpha$4'-region of $apo$-mGBP2.



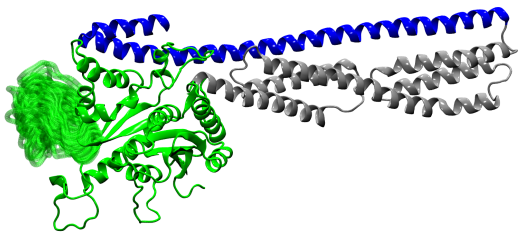
(f) $\alpha$4'-region of $holo$-mGBP2.



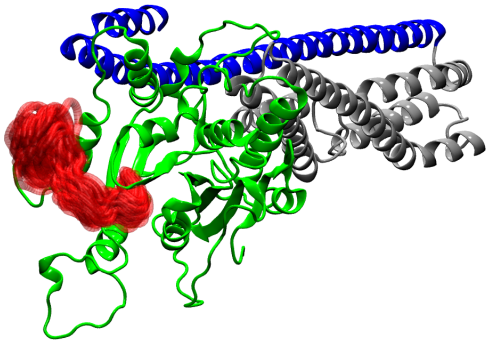
(g) Loop covering amino acids 155 to 170 of $apo$-mGBP2.



(h) Loop covering amino acids 155 to 170 of $holo$-mGBP2.



(i) Loop covering amino acids 180 to 195 of $apo$-mGBP2.

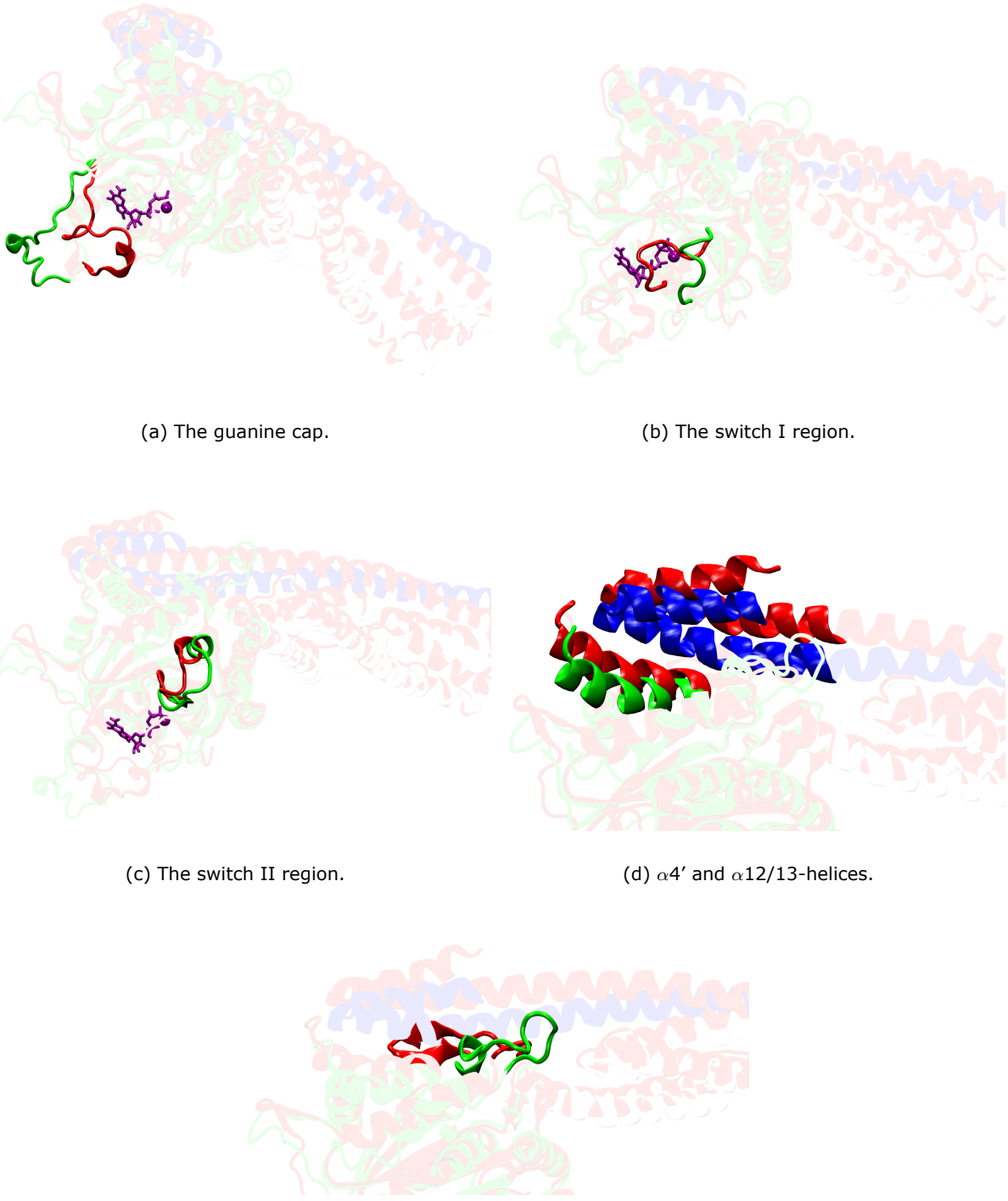


(j) Loop covering amino acids 180 to 195 of $holo$-mGBP2.

Figure 22: The conformational space of the G domain of $apo$- and $holo$-mGBP2 explored in fMD simulations. The middle structures of selected regions resulting from cluster analyses are overlaid onto the structure of the $apo$-mGBP2 model. For the $apo$-state the G domain is shown in green, the M domain in silver and the E domain in blue. The $holo$-state is represented in red.

- The switch II region shifts up on the left side.
- The $\alpha$4'-helix, the terminal part of the $\alpha$12-helix and the $\alpha$13-helix follow this movement.
- The loop consisting of amino acids 155 to 170 swings in and forms a new small $\beta$-sheet.

The immense structural change of the guanine cap upon nucleotide binding was already reported for hGBP1[30], while the observed characteristic of the switch I region shielding the binding pocket from the solvent has been identified in the $holo$-state of mGBP2[3]. This confirms the good agreement between fMD simulations and experiments.
The findings for the loop consisting of amino acids 155 to 170 once more indicate its importance as it shows extensive conformational changes when turning from $apo$- to $holo$-mGBP2. Figure 23(e) nicely depicts its solvent exposed, upright position in front of the terminal part of the $\alpha$13-helix in the $apo$-state. This might be a possible binding pocket in monomeric mGBP2 to host the hydrophobic membrane anchor attached at the C-terminus if mGBP2 is isoprenylated.
From the motion of the $\alpha$13-helix it seems as if it tries to swing out of the conformation prevalent in the $apo$-state of mGBP2. Displaying all conformational states found by clustering at once (figure 24) endorses this hypothesis. To finally validate the observation distance distributions between selected pairs of atoms forming salt bridges in the $apo$-state (for visualization see figure 2) are calculated for the fMD simulations of the $apo$- and $holo$-state. The results for the various salt bridges are shown in figure 25. They clearly confirm the start of the $\alpha$13-helix movement away from the $\alpha$12-helix. This is manifested by the salt bridge contacts: Except for the salt bridge between $O_{E1}$-GLU566 and $N_Z$-LYS226,



(a) The guanine cap.

(b) The switch I region.

(c) The switch II region.

(d) $\alpha$4' and $\alpha$12/13-helices.

(e) The loop consisting of residues 155 to 170, which becomes a $\beta$-sheet in $holo$-mGBP2.

Figure 23: The conformational changes upon nucleotide binding are visualized for different structural motifs. The motifs of the $apo$-state are shown in green and the $holo$-state is colored in red. Conformations were obtained from cluster analysis and the middle structure of the highest populated cluster are shown for both $apo$- and $holo$-mGBP2. The entire structure of $apo$-mGBP2 is shown in the background in transparent green for the G domain, silver for the middle and blue for the E domain. The $holo$-state is colored in transparent red. The nucleotide GTP and the $Mg^{2+}$ co-factor are colored in magenta.

for which the contact increases from 17% to 36% during the fMD simulation time of the $apo$- to $holo$-state, respectively, all other contacts drop by more than 60%. The greatest difference is seen for the salt bridge between $O_{E1}$-GLU561 and $N_{H1}$-ARG225: A salt bridge being present during almost the entire simulation of $apo$-mGBP2, it is only observed in 5% of the time for $holo$-mGBP2. The postulated $\alpha$13-helix motion would furthermore move the membrane anchor of isoprenylated mGBP2 away from the G domain, possibly forming a new dimer interaction surface and enabling membrane attachment. This would make the observation plausible (figure 23(e)) that the loop around amino acids 155 to 170 moves toward the G domain upon nucleotide binding if this is the binding pocket for the membrane anchor in monomeric $apo$-mGBP2.

The presented findings are in good agreement with what was reported by Vöpel et al. for the hGBP1 dimer.[16]

Interestingly, the loop around amino acid 188 shows no major changes upon nucleotide binding, what perfectly matches previous findings making this structural motif special. Its exposure to the solvent renders it a possible interaction surface to other proteins/co-factors which is maintained during GTP binding, indicating its possible importance for dimerization, which should be studied in the future.

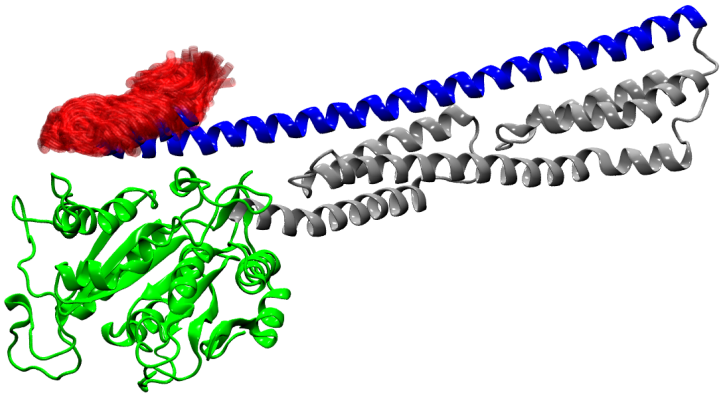
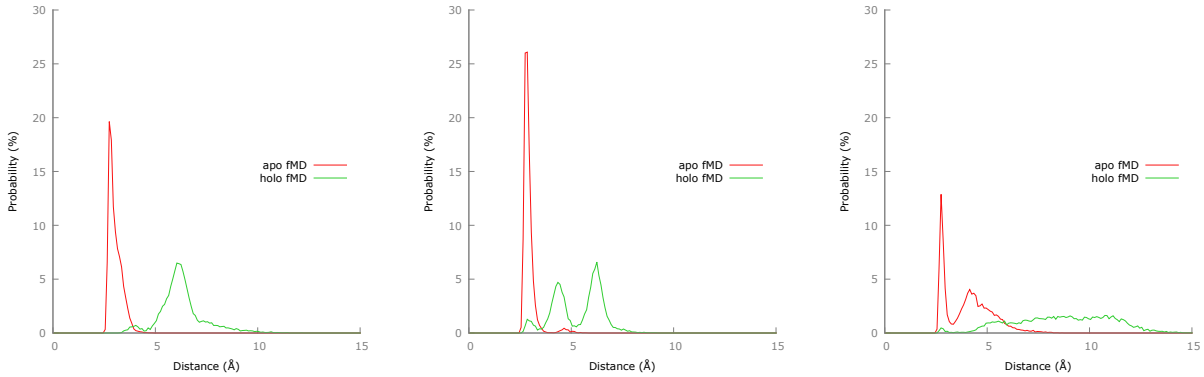


Figure 24: The conformational space of the $\alpha$13-helix in the GTP-bound state of mGBP2 is shown in red by overlaying all middle structures from cluster analysis at once. They are superimposed onto the initial structure of $apo$-mGBP2. The G domain is colored in green, the M domain in silver and the E domain in blue.

Displaying the changes in the active site all at once (figure 26) allows the assumption that all seen conformational changes during GTP binding cooperatively prepare the active site for GTP hydrolysis and produce a force also altering distant parts of mGBP2 to activate the protein for its biological function. This could happen in the following steps resembling the mechanism of allosteric regulation:

1. GTP and the $Mg^{2+}$ co-factor enter the active site.
2. The guanine cap and switch I region close to shield the active site from surrounding solvent.




(a) Distance between $O_{E1}$-GLU561 and $N_{H1}$-ARG225.

(b) Distance between $O_{E1}$-GLU554 and $N_{H2}$-ARG225.

(c) Distance between $O_{E2}$-GLU573 and $N_Z$-LYS226.

Figure 25: Distance distributions for three salt bridges postulated between the G domain and the terminal E domain for the $apo$-state are shown. $apo$-state data is given in red, $holo$ data in green.

3. The movement of the switch I region causes the P-loop and switch II region (all three together connected via $\beta$-sheets) to slightly shift to the upper left forming a circular hole on top of the $\gamma$-phosphate of GTP and $Mg^{2+}$, possibly allowing the hydrolysis products to leave the active site.

4. The resulting switch II motion is translated to the $\alpha$4'-helix pushing it up and finally causing the $\alpha$13-helix to spin off via breakage of the salt bridges between the G and E domain.

The sketched process is visualized in figure 26 indicating the possible dynamics by orange arrows.

To verify this hypothesis one might set up an extensive QM/MM study to be able to follow binding and hydrolysis of GTP and resulting conformational effects. In order to reduce the computational effort a first approach could concentrate on the G domain, as for hGBP1 it was experimentally shown, that the G domain is stable itself – it was crystalized[17] – and able to hydrolyze GTP on its own[17,121]. Moreover one could try to follow GTP effects via FRET studies, e.g., one might be able to investigate whether one can see a sequential change of distances in the G domain by labeling the respective structural motifs.

To sum up the binding of GTP not only causes differences in the flexibility of mGBP2 but moreover extensive changes in the conformational appearance are the result possibly leading to a biochemically active state.

### 4.2.4 Hydrogen Bonds stabilize $holo$ G domain and ensure solid binding of GTP

To understand how GTP is bound in the active site and which kind of interactions might explain the higher rigidity of the G domain in the $holo$-state, the presence of H-bonds in

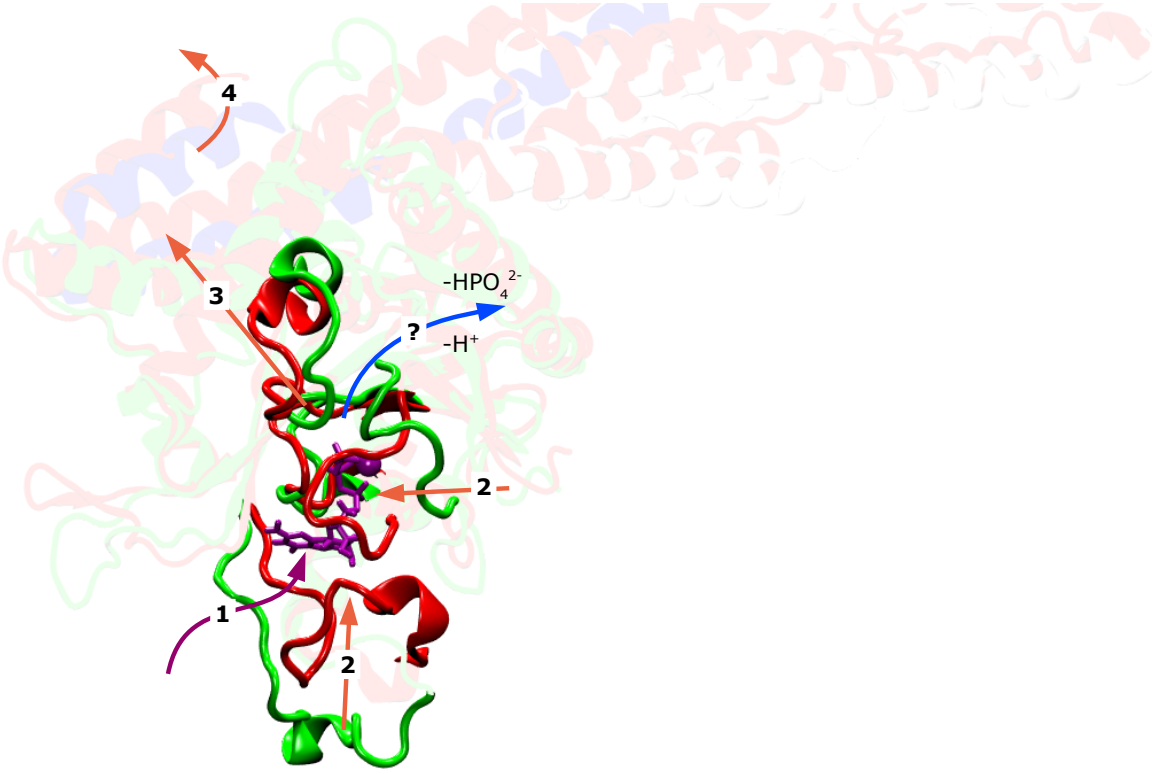



Figure 26: The figure sketches the conformational changes upon nucleotide binding where the directions of resulting dynamics are indicated by orange arrows. The GTP binding site in the $apo$-state is shown in green and the $holo$-state is colored in red. Conformations were obtained from cluster analysis and the middle structure of the highest populated cluster of both states are overlaid. The entire structure of $apo$-mGBP2 is shown in the background in transparent green for the G domain, silver for the middle and blue for the E domain. The $holo$-state is colored in transparent red. The nucleotide GTP and the $Mg^{2+}$ co-factor are colored in magenta.

the fMD simulations is evaluated as explained in section 3.4.7.

The observation of conformational changes of the active site following GTP binding (section 4.2.3) invoked further analysis of the interactions between the two switch regions, between the P-loop and the switch regions, as well as between the guanine cap and the switch I region, aiming to explain the stiffening of the G domain.
Figure 27 exemplarily shows the results for the H-bond analysis between the P-loop and the G4 motif. For further visualization selected views on the active site will be given.

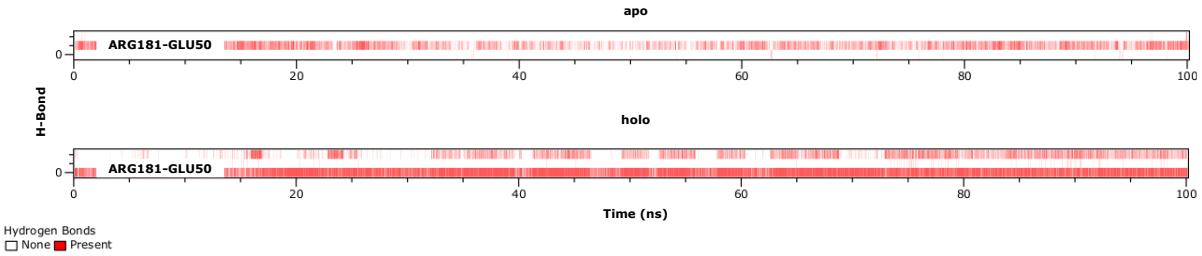



Figure 27: The results of the H-bond analysis between the P-loop and the G4 motif are shown. The upper part reflects the $apo$-state, the lower part the $holo$-state. The strongest H-bonds are found between GLU50 and ARG181.

The results of the H-bond analysis for the $apo$- and $holo$-state clearly reflect the findings from the B-factor and conformation analyses. The guanine cap and switch I region are stabilized by an H-bond between SER69 and SER246 (compare figure 28), while a corresponding H-bond is not present in the $apo$-state. The results for the switch I and II region resemble the reduced flexibility of the G domain in the $holo$-state. Whereas in $apo$-mGBP2 H-bonds highly fluctuate between several residues, the number of H-bonds found in the $holo$-state is significantly reduced but far more stable. In the first case the strongest bond is found between LYS76 and ASP112 and in the latter between LYS76 and GLU99.
The P-loop interactions appear to play a central role. Being located between the switch I and II region and the G4 motif it seems to be the central motif for the G domain stabilization. Whereas in the $apo$-state just a few highly fluctuating H-bonds occur – among others the H-Bbond between LYS51 and THR98, which stabilizes the P-loop in the first third of the trajectory and which was also found for hGBP1[80] – in nucleotide-bound mGBP2 strong H-bonds are present between the four motifs making the P-loop the core element. It holds together the two switch regions and the G4 motif, stabilizing the position of the switch I motif and the closure of the guanine cap in $holo$-mGBP2. This conclusion is supported by the fact that for the P-loop no great flexibility was found and hardly any conformational changes between the $apo$- and the $holo$-state could be observed in earlier investigations. This conformational conservation is likely to be indispensible for the formation of the active site when binding GTP. The key residues for these H-bonds seem to be ARG48 and THR70 for the P-loop–switch I interaction (compare figure 29(a)), SER52 and ASP97 for the P-loop–switch II interaction (figure 29(b)) and GLY50 and ARG181 for the P-loop–G4 inter-

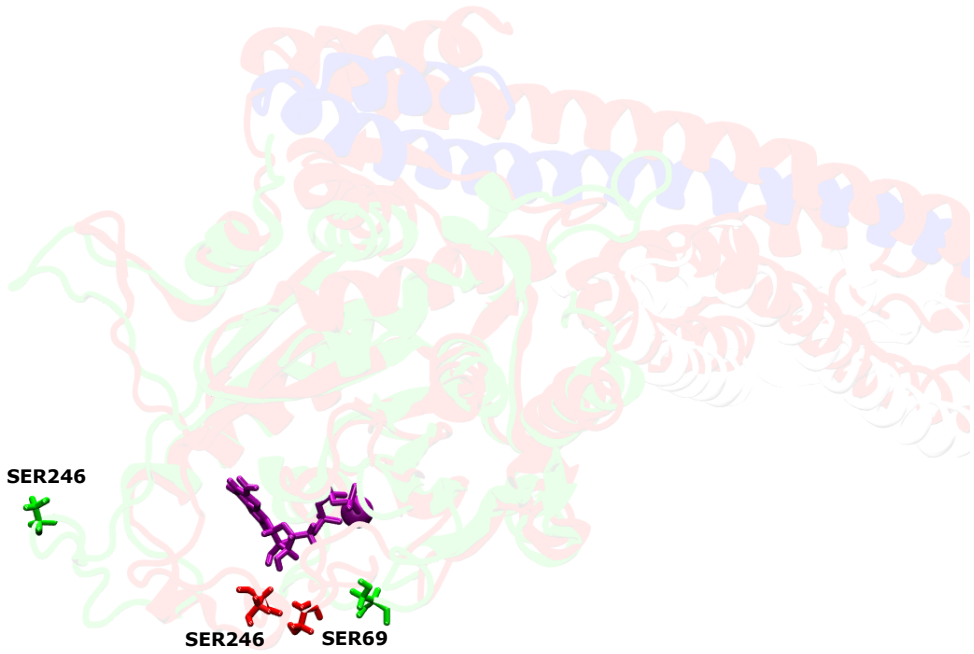
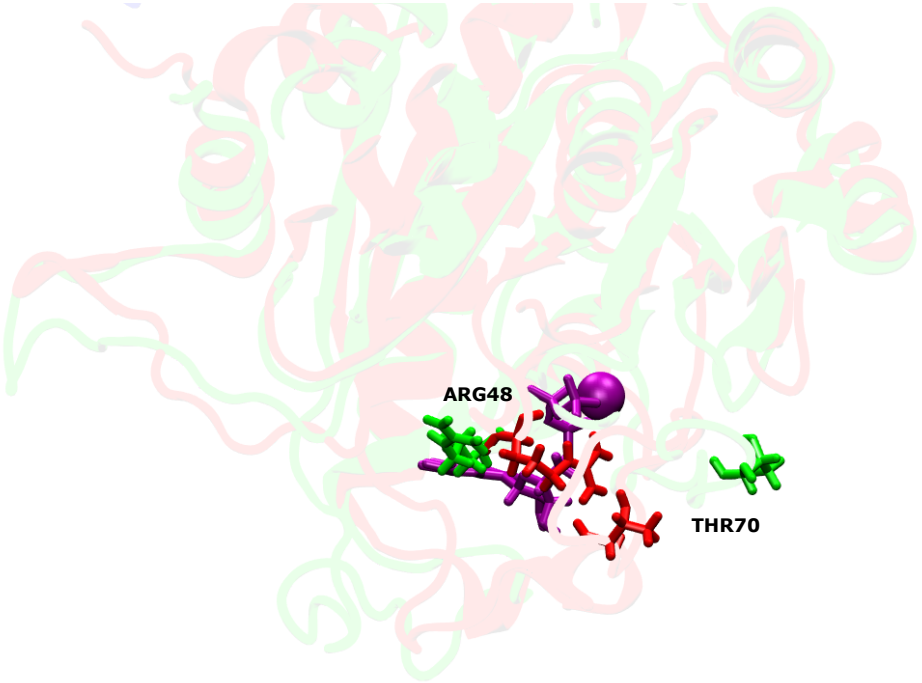


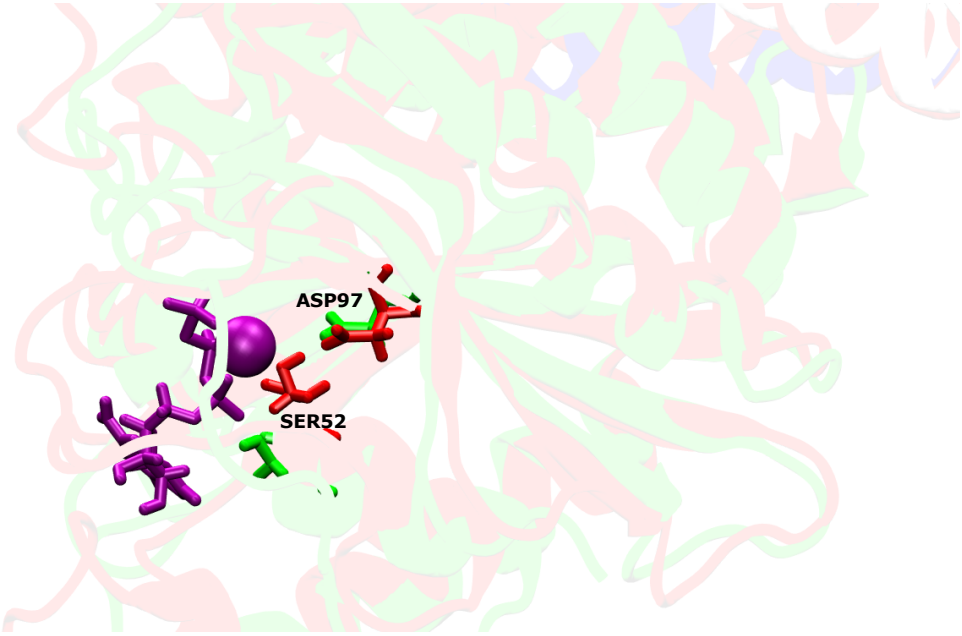

Figure 28: The H-bond between SER69 and SER246 in the $holo$-state is visualized. The named residues are shown in red for the $holo$-state and in green for the $apo$-state. In the background the structure of the entire mGBP2 is displayed with red coloring for the nucleotide-bound protein, in green the G domain of $apo$-mGBP2, in silver the M domain and in blue the E domain. GTP and $Mg^{2+}$ are visualized in purple.

action. This corresponds well to Kravets' findings. The interplay between LYS51 (P-Loop) and GLU99 (switch II), which is observed for Ras proteins[122,123], was not found for mGBP2 as these two residues move away from each other upon nucleotide binding. Regarding the P-loop–G4 interactions, Kravets hypothesized that two G4 residues contact the P-loop[3]. As shown above this can only be confirmed partially here as only one G4 residue forms an H-bond with the P-loop. Yet experimentally unknown, simulations indicate that this important G4 residue is the already mentioned ARG181.

Similar to their importance in stabilization and stiffening of the G domain, H-bonds are also pivotal for the solid binding of GTP. The active site with residues found to be important for GTP binding is visualized in figure 30.




(a) H-bond between ARG48 and THR70.




(b) H-bond between SER52 and ASP97.

Figure 29: Visualization of H-bonds, where the residues involved are shown in red for the $holo$-state and in green for the $apo$-state. In the background the structure of the entire mGBP2 is displayed with red coloring for the nucleotide-bound protein, green for the G domain of $apo$-mGBP2, silver for the M domain and blue for the E domain. GTP and $Mg^{2+}$ are visualized in purple.

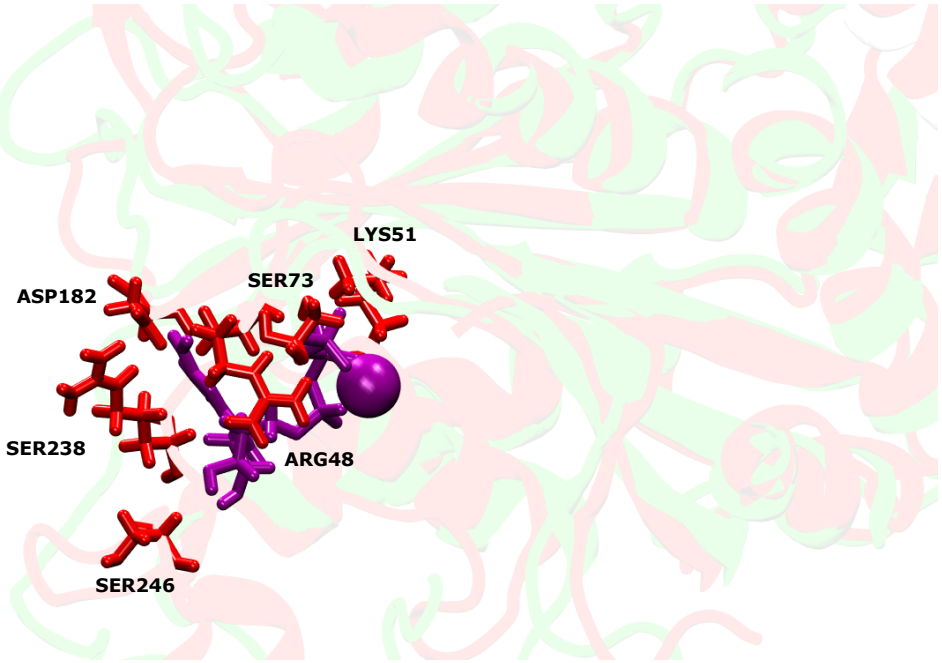



Figure 30: Visualization of the H-bonds between GTP and residues SER246, ARG238, ASP182, SER73, ARG48 and LYS51. These residues are shown in red. In the background the structure of the entire mGBP2 is displayed with red coloring for the nucleotide-bound protein, green for the G domain of $apo$-mGBP2, silver for the M domain and blue for the E domain. GTP and $Mg^{2+}$ are visualized in purple.

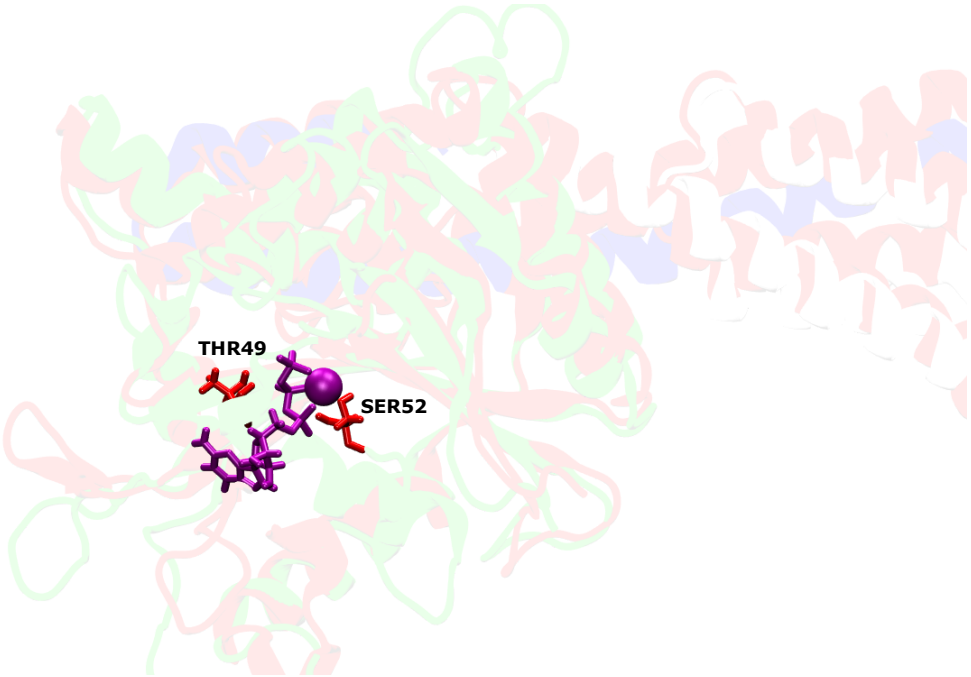



Figure 31: Visualization of the coordination of $Mg^{2+}$ by SER52. Residues are shown in red. In the background the structure of the entire mGBP2 is displayed with red coloring for the nucleotide-bound protein, green for the G domain of $apo$-mGBP2, silver for the M domain and blue for the E domain. GTP and $Mg^{2+}$ are visualized in purple.

It is seen that the switch II region has a negligible impact on GTP binding with only one weak H-bond between GLY100 and the nucleotide $\gamma$-phosphate. This confirms earlier conclusions drawn from experimental work[3]. The guanine cap mainly binds to GTP via SER246. The biochemically proposed contact between ARG238 and GTP is found to be less important[3]. ASP182 binds GTP via two H-bonds as seen for hGBP1[124]. Predominantly responsible for the stable binding of GTP are the switch I region via SER73 and the P-loop. The latter forms 26 different H-bonds with the $\alpha$, $\beta$ and $\gamma$-phosphate of GTP. Thereof 8 H-bonds are prevalently present in more than 95% of the simulation time. In particular ARG48 and LYS51 account for these interactions. Kravets also found LYS51 to play an essential role in nucleotide binding, multimerization and catalytic activity[3]. A divergence is seen for ARG48 as Kravets did not find a notable importance of this residue for GTP binding. The $Mg^{2+}$ co-factor is mainly bound by the P-loop residue SER52 and the $\beta$ and $\gamma$-phosphate of GTP confirming the hypothesis of Kravets[3]. The same was already known for Ras proteins[30,31,125]. For THR49 no such importance was found (see figure 31).

In summary, the findings predict a pivotal role of H-bonds for the stiffening of the G domain and stable GTP binding.Residues ARG48 and LYS51 are exceptionally important for nucleotide binding whereas the P-loop seems to be the key motif for the structure of the active site, enabling the conformational adaption upon GTP binding and the enlarged rigidity of the $holo$-state G domain.

# 5 Conclusion and Outlook

The first aim of this work was to assess the capabilities of different molecular modeling approaches (ENM, as well as cMD and fMD simulations) in describing the conformations and dynamics of the mGBP2 protein.
It could be shown that all three models are in good agreement with each other, predicting similar protein flexibility and essential dynamics. The conformational flooding technique using ENM data as reaction coordinates was successfully implemented and does not suffer from the ''tip effect''[119]. Yet it was observed that the choice of parameters is not straightforward and extensive testing was necessary. Finally computational costs remained at an affordable level.
Comparison of the cMD simulation of $apo$-mGBP2 with results from hpFRET experiments for hGBP1 revealed the correct sampling of the experimentally predicted major state. However, the minor conformation of the protein was not found. This can be due to two possible reasons: Either the cMD simulation was not converged – which is likely regarding the size of the simulation system – or the necessary mutation and labeling of amino acids for FRET measurements changed the protein structure more than anticipated. The mentioned sampling problem will have to be further addressed in future work, e.g., by changing the fMD simulations from constant to the adaptive flooding method, by means of Hamiltonian REMD (H-REMD) or a combination of H-REMD and the conformational flooding method applied in this work, and finally by incorporating FRET restraints in MD simulations to see whether the experimental predictions can be reproduced. Moreover, to investigate possible structural influence of FRET fluorophors on mGBP2, MD simulations should be performed integrating residue mutations and the respective dyes.

A detailed analysis of the effects on the protein caused by nucleotide binding unraveled tremendous changes in conformations and protein flexibility, as well as the observation of new dynamics. Results from B-factor analysis, principal component analysis, clustering and conformational space analysis, as well as salt bridge and hydrogen bond examination perfectly complement each other, drawing a detailed picture of the induced changes upon GTP binding. These findings are moreover in good agreement with the experimentally confirmed importance of the characteristic G domain motifs conserved from the Ras structure[3,8,14,120].
In addition, two interesting loops that have not been described yet are found: The loop consisting of amino acids 155 to 170 might account for the shielding of the isoprenylated C-terminus, while the loop around amino acid 188 may possibly be important in the dimer interaction surface of mGBP2.
To clarify the suggested role of the loop of residues 155 to 170 additional simulations resolving the C-terminal CAAX sequence including isoprenylation should be done. Thereby, influences on the motions of the $\alpha$12/13-helix should be reviewed to complete the knowledge about the mGBP2 monomer.

Based on the results of this work a model resembling an allosteric regulation mechanism was presented. It incorporates all observations to explain emerging forces that transfer flexibility and dynamics from the G domain to the $\alpha$13-helix as recently reported for the hGBP1 dimer by Vöpel et al.[16]
To verify the proposed model it would preliminary be helpful if the crystal structure of $holo$-mGBP2 was solved to check for the correctness of the applied structural model of mGBP2 as the full-length protein was only obtained by merging the $holo$-state model of the G domain together with the M and E domain of the $apo$-mGBP2 model. Furthermore, it should be discussed whether the use of a dummy model[126--128] to describe the $Mg^{2+}$ co-factor and a polarizable force field can provide deeper insights. A promising yet challenging approach would be a combined QM/MM study to enable the description of GTP binding and hydrolysis to instantly follow the resulting structural changes and induced dynamics.

In summary, the present work validates the suitability of molecular modeling approaches to describe the murine Guanylate Binding Protein 2 to complement insights from experimental studies and to identify new points of interest.

# 6 Appendix

## 6.1 Appendix 1 – Simulation protocols

### 6.1.1 Energy Minimization

$Apo$-state energy minimizations were done using the GROMACS software with the following *mdp* file specifications:

```
integrator  = steep
emtol       = 500.0
emstep      = 0.001
nsteps      = 10000

nstxout     = 10

; Neighbor searching
cutoff-scheme           = Verlet
nstlist                 = 20
ns-type                 = grid
verlet-buffer-tolerance = 0.005
pbc                     = xyz

; Electrostatics
coulombtype    = PME
pme-order      = 4
fourierspacing = 0.10
rcoulomb       = 1.2

; VdW
rvdw           = 1.2
DispCorr       = EnerPres

constraints = none
```

The following *mdp* file specifications were used for energy minimization of the $holo$-state:

```
integrator  = steep
emtol       = 500.0
emstep      = 0.01
nsteps      = 20000

nstxout     = 10

; Neighbor searching
cutoff-scheme           = Verlet
nstlist                 = 20
ns-type                 = grid
verlet-buffer-tolerance = 0.005
pbc                     = xyz

; Electrostatics
coulombtype    = PME
pme-order      = 4
```

```
fourierspacing = 0.10
rcoulomb       = 1.2

; VdW
rvdw           = 1.2
DispCorr       = EnerPres

constraints = none
```

### 6.1.2 Equilibration

*Apo*-state equilibration runs were done using the GROMACS software with the following *mdp* file specifications:

```
; Preprocessing
define        = -DPOSRES

; Run control
integrator    = md
dt            = 0.002
tinit         = 0
nsteps        = 50000
comm-mode     = Linear
nstcomm       = 100

; Output control
nstxout            = 5000
nstvout            = 5000
nstfout            = 5000
nstxout-compressed = 500
nstlog             = 100
nstcalcenergy      = 100
nstenergy          = 100

; Neighbor searching
cutoff-scheme = Verlet
nstlist       = 20
ns-type       = grid
pbc           = xyz

; Electrostatics
coulombtype    = PME
pme-order      = 4
fourierspacing = 0.10
rcoulomb       = 1.2

; VdW
rvdw          = 1.2
DispCorr      = EnerPres

; Temperature coupling
tcoupl          = v-rescale
tc-grps         = Protein Non-Protein
tau-t           = 0.1  0.1
ref-t           = 298  298
```

```
nsttcouple      = 10

; Pressure coupling
pcoupl           = berendsen
tau-p            = 1.0
pcoupltype       = isotropic
compressibility  = 4.5e-5
ref-p            = 1.0
refcoord-scaling = all
nstpcouple       = 10

; Velocity generation
gen-vel       = yes
gen-temp      = 298

; Bonds
continuation         = no
constraints          = all-bonds

constraint-algorithm = LINCS
lincs-order          = 6
lincs-iter           = 1
lincs-warnangle      = 30
```

The following *mdp* file specifications were used for the first equilibration of the $holo$-state:

```
; Preprocessing
define        = -DPOSRES_PROTEIN -DPOSRES_MG -DPOSRES_GTP

; Run control
integrator    = md
dt            = 0.002
tinit         = 0
nsteps        = 50000
comm-mode     = Linear
nstcomm       = 100

; Output control
nstxout            = 5000
nstvout            = 5000
nstfout            = 5000
nstxout-compressed = 500
nstlog             = 100
nstcalcenergy      = 100
nstenergy          = 100

; Neighbor searching
cutoff-scheme = Verlet
nstlist       = 20
ns-type       = grid
pbc           = xyz

; Electrostatics
coulombtype    = PME
pme-order      = 4
fourierspacing = 0.10
```

```
rcoulomb       = 1.2

; VdW
rvdw          = 1.2
DispCorr      = EnerPres

; Temperature coupling
tcoupl          = v-rescale
tc-grps         = Protein_MG_GTP Water_and_ions
tau-t           = 0.1  0.1
ref-t           = 298  298
nsttcouple      = 10

; Pressure coupling
pcoupl           = berendsen
tau-p            = 1.0
pcoupltype       = isotropic
compressibility  = 4.5e-5
ref-p            = 1.0
refcoord-scaling = all
nstpcouple       = 10

; Velocity generation
gen-vel       = yes
gen-temp      = 298

; Bonds
continuation         = no
constraints          = all-bonds

constraint-algorithm = LINCS
lincs-order          = 6
lincs-iter           = 1
lincs-warnangle      = 30
```

The following *mdp* file specifications were used for the second equilibration of the $holo$-state:

```
; Preprocessing
define        = -DPOSRES_PROTEIN

; Run control
integrator    = md
dt            = 0.002
tinit         = 0
nsteps        = 50000
comm-mode     = Linear
nstcomm       = 100

; Output control
nstxout            = 5000
nstvout            = 5000
nstfout            = 5000
nstxout-compressed = 500
nstlog             = 100
nstcalcenergy      = 100
```

```
nstenergy          = 100

; Neighbor searching
cutoff-scheme = Verlet
nstlist       = 20
ns-type       = grid
pbc           = xyz

; Electrostatics
coulombtype    = PME
pme-order      = 4
fourierspacing = 0.10
rcoulomb       = 1.2

; VdW
rvdw          = 1.2
DispCorr      = EnerPres

; Temperature coupling
tcoupl          = v-rescale
tc-grps         = Protein_MG_GTP Water_and_ions
tau-t           = 0.1  0.1
ref-t           = 298  298
nsttcouple      = 10

; Pressure coupling
pcoupl           = berendsen
tau-p            = 1.0
pcoupltype       = isotropic
compressibility  = 4.5e-5
ref-p            = 1.0
refcoord-scaling = all
nstpcouple       = 10

; Velocity generation
gen-vel       = no

; Bonds
continuation         = yes
constraints          = all-bonds

constraint-algorithm = LINCS
lincs-order          = 6
lincs-iter           = 1
lincs-warnangle      = 30
```

### 6.1.3 MD production runs

*Apo*-state MD production runs were done using the GROMACS software with the following *mdp* file specifications:

```
; Run control
integrator    = md
dt            = 0.002
tinit         = 0
```

```
nsteps        = 50000000
comm-mode     = Linear
nstcomm       = 100

; Output control
nstxout            = 5000
nstvout            = 5000
nstfout            = 5000
;nstxout-compressed = 500
nstlog             = 100
nstcalcenergy      = 100
nstenergy          = 100

; Neighbor searching
cutoff-scheme = Verlet
nstlist       = 20
ns-type       = grid
pbc           = xyz

; Electrostatics
coulombtype    = PME
pme-order      = 4
fourierspacing = 0.10
rcoulomb       = 1.2

; VdW
rvdw          = 1.2
DispCorr      = EnerPres

; Temperature coupling
tcoupl          = Nose-Hoover
nhchainlength   = 1
tc-grps         = Protein Non-Protein
tau-t           = 2.0  2.0
ref-t           = 298  298
nsttcouple      = 10

; Pressure coupling
pcoupl           = Parrinello-Rahman
tau-p            = 2.0
pcoupltype       = isotropic
compressibility  = 4.5e-5
ref-p            = 1.0
refcoord-scaling = all
nstpcouple       = 10

; Velocity generation
gen-vel       = no

; Bonds
continuation         = yes
constraints          = all-bonds

constraint-algorithm = LINCS
lincs-order          = 6
lincs-iter           = 1
lincs-warnangle      = 30
```

The following *mdp* file specifications were used for MD production runs of the $holo$-state:

```
; Run control
integrator    = md
dt            = 0.002
tinit         = 0
nsteps        = 50000000
comm-mode     = Linear
nstcomm       = 100

; Output control
nstxout            = 5000
nstvout            = 5000
nstfout            = 5000
nstxout-compressed = 5000
nstlog             = 100
nstcalcenergy      = 100
nstenergy          = 100
energygrps         = Protein GTP

; Neighbor searching
cutoff-scheme = Verlet
nstlist       = 20
ns-type       = grid
pbc           = xyz

; Electrostatics
coulombtype    = PME
pme-order      = 4
fourierspacing = 0.10
rcoulomb       = 1.2

; VdW
rvdw          = 1.2
DispCorr      = EnerPres

; Temperature coupling
tcoupl          = Nose-Hoover
nhchainlength   = 1
tc-grps         = Protein_MG_GTP Water_and_ions
tau-t           = 2.0  2.0
ref-t           = 298  298
nsttcouple      = 10

; Pressure coupling
pcoupl           = Parrinello-Rahman
tau-p            = 2.0
pcoupltype       = isotropic
compressibility  = 4.5e-5
ref-p            = 1.0
refcoord-scaling = all
nstpcouple       = 10

; Velocity generation
gen-vel       = no
```

```
; Bonds
continuation         = yes
constraints          = all-bonds

constraint-algorithm = LINCS
lincs-order          = 6
lincs-iter           = 1
lincs-warnangle      = 30
```

### 6.1.4 Flooding input file

Below the parameter definition part of an *edi* input file for fMD simulations is given:

```
#MAGIC
 670
#NINI
 293924
#FITMAS
 0
#ANALYSIS_MAS
 0
#OUTFRQ
 5000
#MAXLEN
 0
#SLOPECRIT
 0.000000
#PRESTEPS
 0
#DELTA_F0
 150.000000
#INIT_DELTA_F
 0.000000
#TAU
 0.000000
#EFL_NULL
 10.000000
#ALPHA2
 1.000000
#KT
 2.483323
#HARMONIC
 0
#CONST_FORCE_FLOODING
 0
```

## 6.2 Appendix 2 – Sequence Alignment of hGBP1 and mGBP2

The following figure 32 shows the sequence alignment of hGBP1 and mGBP2 using the EMBOSS stretcher tool[129] for global pairwise alignment available at *http://www.ebi.ac.uk/Tools/psa/emboss_stretcher/*. Default settings were applied.

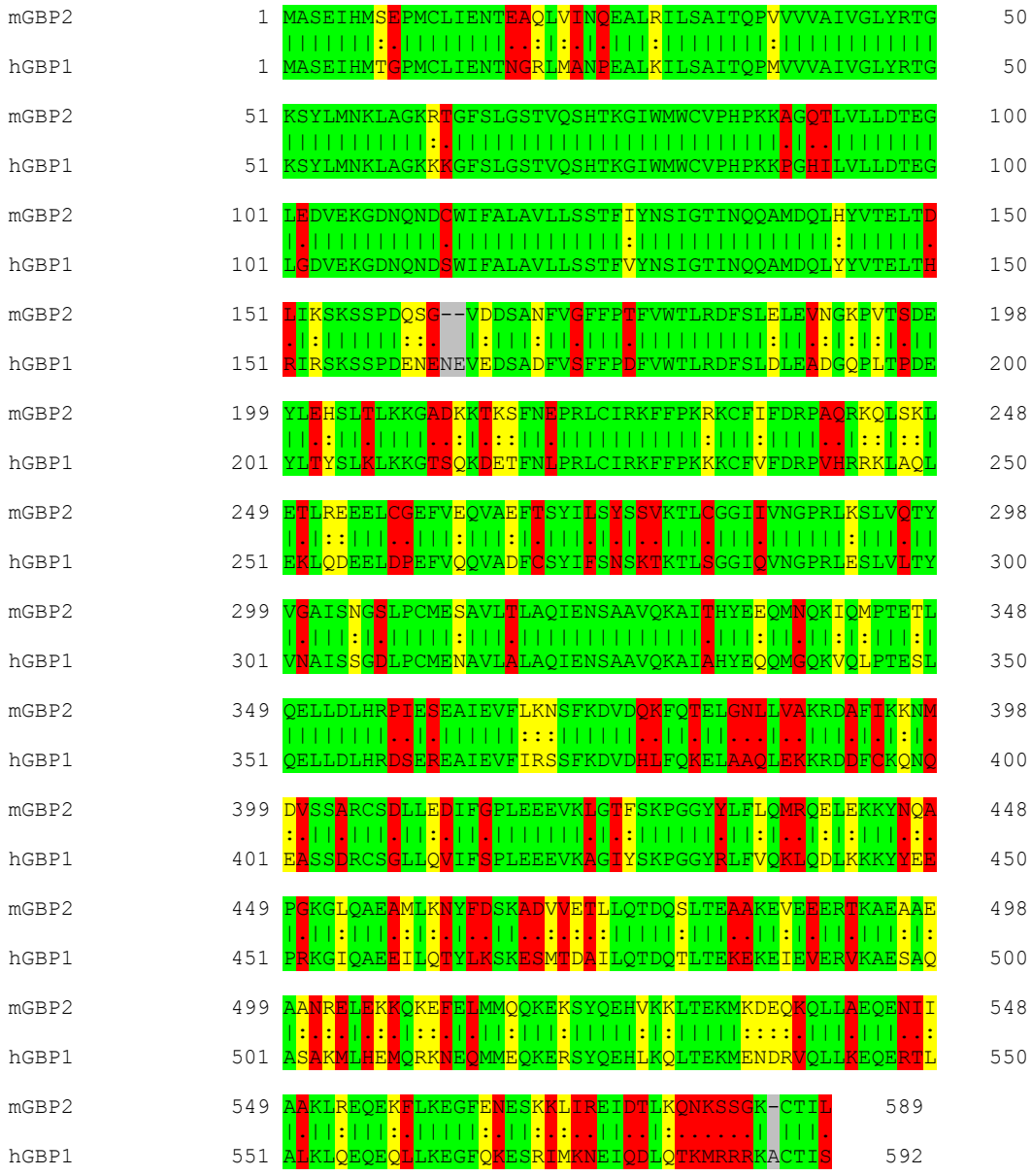



Figure 32: The sequence alignment of hGBP1 and mGBP2 using the EMBOSS stretcher tool for global pairwise alignment is shown. Green indicates identical sequence parts, yellow marks biochemically similar residues and red means no similarity. Gray parts show gaps between the sequences.

The sequence of the entire mGBP2 is at 68% identical and at 82% similar to hGBP1. The sequence identity between mGBP2 and hGBP1 only considering the G domain is 76% and the similarity is 86%.

# References


[1] G. J. K. Praefcke and H. T. McMahon. The dynamin superfamily: universal membrane tubulation and fission molecules? *Nature reviews Molecular cell biology*, 5:133--147, 2004.

[2] Y. S. Cheng, R. J. Colonno, and F. H. Yin. Interferon induction of fibroblast proteins with guanylate binding activity. *Journal of Biological Chemistry*, 258:7746--7750, 1983.

[3] E. Kravets. *Charakterisierung des murinen Guanylat-bindenden Proteins 2 (mGBP2)*. PhD thesis, Heinrich-Heine-Universität Düsseldorf, Düsseldorf, Germany, 2012.

[4] K. Faelber, Y. Posor, S. Gao, M. Held, Y. Roske, D. Schulze, V. Haucke, F. Noé, and O. Daumke. Crystal structure of nucleotide-free dynamin. *Nature*, 477:556--560, 2011.

[5] A. R. Shenoy, B.-H. Kim, H.-P. Choi, T. Matsuzawa, S. Tiwari, and J. D. MacMicking. Emerging themes in IFN-$\gamma$-induced macrophage immunity by the p47 and p65 GTPase families. *Immunobiology*, 212:771--784, 2008.

[6] A. Kresse, C. Konermann, D. Degrandi, C. Beuter-Gunia, J. Wuerthner, K. Pfeffer, and S. Beer. Analyses of murine gbp homology clusters based on in silico, in vitro and in vivo studies. *BMC genomics*, 9:158, 2008.

[7] D. Degrandi, C. Konermann, C. Beuter-Gunia, A. Kresse, J. Würthner, S. Kurig, S. Beer, and K. Pfeffer. Extensive characterization of IFN-induced GTPases mGBP1 to mGBP10 involved in host defense. *The Journal of Immunology*, 179:7729--7740, 2007.

[8] D. Degrandi, E. Kravets, C. Konermann, C. Beuter-Gunia, V. Klümpers, S. Lahme, E. Wischmann, A. K. Mausberg, S. Beer-Hammer, and K. Pfeffer. Murine guanylate binding protein 2 (mGBP2) controls toxoplasma gondii replication. *Proceedings of the National Academy of Sciences*, 110:294--299, 2013.

[9] B. J. Luft and J. S. Remington. Toxoplasmic encephalitis in AIDS. *Clinical Infectious Diseases*, 15:211--222, 1992.

[10] K. Kim and L. M. Weiss. Toxoplasma gondii: the next 100years. *Microbes and Infection*, 10:978--984, 2008.

[11] K. Kim and L. M. Weiss. Toxoplasma gondii: the model apicomplexan. *International Journal for Parasitology*, 34:423--432, 2004.

[12] G. J. K. Praefcke, S. Kloep, U. Benscheid, H. Lilie, B. Prakash, and C. Herrmann. Identification of residues in the human guanylate-binding protein 1 critical for nucleotide binding and cooperative GTP hydrolysis. *Journal of molecular biology*, 344:257--269, 2004.

[13] M. A. Olszewski, J. Gray, and D. J. Vestal. In silico genomic analysis of the human and murine guanylate-binding protein (gbp) gene clusters. *Journal of interferon & cytokine research*, 26:328--352, 2006.

[14] E. Kravets, D. Degrandi, S. Weidtkamp-Peters, B. Ries, C. Konermann, S. Felekyan, J. M. Dargazanli, G. J. K. Praefcke, C. A. M. Seidel, S. H. Smits L. Schmitt, and K. Pfeffer. The GTPase activity of murine guanylate-binding protein 2 (mGBP2) controls the intracellular localization and recruitment to the parasitophorous vacuole of toxoplasma gondii. *Journal of Biological Chemistry*, 287:27452--27466, 2012.

[15] C. A. M. Seidel. personal information, unpublished data.

[16] T. Vöpel, C. S. Hengstenberg, T. Peulen, Y. Ajaj, C. A. M. Seidel, C. Herrmann, and J. P. Klare. 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, 2014.

[17] 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*, 440:101--104, 2006.

[18] U. Boehm, L. Guethlein, T. Klamp, K. Ozbek, A. Schaub, A. Fütterer, K. Pfeffer, and J. C. Howard. Two families of GTPases dominate the complex cellular response to IFN-$\gamma$. *The Journal of Immunology*, 161:6715--6723, 1998.

[19] T. T. Nguyen, Y. Hu, D. P. Widney, R. A. Mar A, and J. B. Smith. Murine gbp-5, a new member of the murine guanylate-binding protein family, is coordinately regulated with other gbps in vivo and in vitro. *Journal of interferon & cytokine research*, 22:899--909, 2002.

[20] S. Martens and J. Howard. The interferon-inducible GTPases. *Annual Review of Cell and Developmental Biology*, 22:559--589, 2006.

[21] J. John, R. Sohmen, J. Feuerstein, R. Linke, A. Wittinghofer, and R. S. Goody. Kinetics of interaction of nucleotides with nucleotide-free h-ras p21. *Biochemistry*, 29:6058--6065, 1990.

[22] E. Guenzi, K. Töpolt, E. Cornali, C. Lubeseder-Martellato, A. Jörg, K. Matzen, C. Zietz, E. Kremmer, F. Nappi, M. Schwemmle, C. Hohenadl, G. Barillari, E. Tschachler, P. Monini, B. Ensoli, and M. Stürzl. The helical domain of gbp-1 mediates the inhibition of endothelial cell proliferation by inflammatory cytokines. *The EMBO journal*, 20:5568--5577, 2001.

[23] V. Y. Gorbacheva, D. Lindner, G. C. Sen, and D. J. Vestal. The interferon (IFN)-induced GTPase, mGBP2 role in IFN-$\gamma$-induced murine fibroblast proliferation. *Journal of Biological Chemistry*, 277:6080--6087, 2002.

[24] G. A. Taylor, C. G. Feng, and A. Sher. p47 GTPases: regulators of immunity to intracellular pathogens. *Nature Reviews Immunology*, 4:100--109, 2004.

[25] S. L. Anderson, J. M. Carton, J. Lou, L. Xing, and B. Y. Rubin. Interferon-induced guanylate binding protein-1 (gbp-1) mediates an antiviral effect against vesicular stomatitis virus and encephalomyocarditis virus. *Virology*, 256:8--14, 1999.

[26] C. C. Carter, V. Y. Gorbacheva, and D. J. Vestal. Inhibition of VSV and EMCV replication by the interferon-induced GTPase, mGBP2: differential requirement for wild-type GTP binding domain. *Archives of virology*, 150:1213--1220, 2005.

[27] A. F. Messmer-Blust, S. Balasubramanian, V. Y. Gorbacheva, J. A. Jeyaratnam, and D. J. Vestal. The interferon-$\gamma$-induced murine guanylate-binding protein-2 inhibits rac activation during cell spreading on fibronectin and after platelet-derived growth factor treatment: Role for phosphatidylinositol 3-kinase. *Molecular biology of the cell*, 21:2514--2528, 2010.

[28] K. Arnold, L. Bordoli, J. Kopp, and T. Schwede. The swiss-model workspace: a web-based environment for protein structure homology modelling. *Bioinformatics*, 22:195--201, 2006.

[29] M. Biasini, S. Bienert, A. Waterhouse, K. Arnold, G. Studer, T. Schmidt, F. Kiefer, T. G. Cassarino, M. Bertoni, L. Bordoli, and T. Schwede. Swiss-model: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic acids research*, page gku340, 2014.

[30] M. Wehner, S. Kunzelmann, and C. Herrmann. The guanine cap of human guanylate-binding protein 1 is responsible for dimerization and self-activation of GTP hydrolysis. *Febs Journal*, 279:203--210, 2012.

[31] T. Vöpel, A. Syguda, N. Britzen-Laurent, S. Kunzelmann, M.-B. Lüdemann, C. Dovengerds, M. Stürzl, and C. Herrmann. Mechanism of GTPase-activity-induced self-assembly of human guanylate binding protein 1. *Journal of molecular biology*, 400:63--70, 2010.

[32] N. Abdullah, M. Balakumari, and A. K. Sau. Dimerization and its role in GMP formation by human guanylate binding proteins. *Biophysical journal*, 99:2235--2244, 2010.

[33] M. M. Tirion. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Physical review letters*, 77:1905, 1996.

[34] S. Hayward and B. L. De Groot. *Normal modes and essential dynamics*, chapter 5, pages 89--106. Springer, 2008.

[35] K. Hinsen. Analysis of domain motions by approximate normal mode calculations. *Proteins Structure Function and Genetics*, 33:417--429, 1998.

[36] K. Hinsen, A.-J. Petrescu, S. Dellerue, M.-C. Bellissent-Funel, and G. R. Kneller. Harmonicity in slow protein dynamics. *Chemical Physics*, 261:25--37, 2000.

[37] T. Haliloglu, I. Bahar, and B. Erman. Gaussian dynamics of folded proteins. *Physical Review Letters*, 79:3090--3093, 1997.

[38] S. Mahajan and Y.-H. Sanejouand. On the relationship between low-frequency normal modes and the large-scale conformational changes of proteins. *Archives of biochemistry and biophysics*, 567:59--65, 2015.

[39] D. Frenkel and B. Smit. *Understanding Molecular Simulation: From Algorithms to Applications*. Academic Press, San Diego, USA, 2002.

[40] D. C. Rapaport. *The Art of Molecular Dynamics Simulation*. Cambridge University Press, New York, USA, 2004.

[41] R. W. Hockney. Potential calculation and some applications. *Methods in Computational Physics*, 9:135--211, 1970.

[42] W. F. van Gunsteren and H. J. C. Berendsen. Computer simulation of molecular dynamics: Methodology, applications, and perspectives in chemistry. *Angewandte Chemie International Edition in English*, 29:992--1023, 1990.

[43] J. W. Ponder and D. A. Case. Force fields for protein simulations. *Advances in protein chemistry*, 66:27--85, 2003.

[44] A. D. Mackerell. Empirical force fields for biological macromolecules: overview and issues. *Journal of computational chemistry*, 25:1584--1604, 2004.

[45] O. Guvench and A. D. MacKerell Jr. *Comparison of protein force fields for molecular dynamics simulations*, chapter 4, pages 63--88. Springer, 2008.

[46] K. Lindorff-Larsen, P. Maragakis, S. Piana, M. P. Eastwood, R. O. Dror, and D. E. Shaw. Systematic validation of protein force fields against experimental data. *PloS one*, 7:e32131, 2012.

[47] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. Charmm: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry*, 4:187--217, 1983.

[48] A. D. MacKerell, M. Feig, and C. L. Brooks. Improved treatment of the protein backbone in empirical force fields. *Journal of the American Chemical Society*, 126:698--699, 2004.

[49] B. R. Brooks et al. Charmm: The biomolecular simulation program. *Journal of Computational Chemistry*, 30:1545--1614, 2009.

[50] W. F. van Gunsteren and H. J. C. Berendsen. Groningen molecular simulation (gromos). *Library manual, Biomos, Groningen, The Netherlands*, pages 1--221, 1987.

[51] W. F. van Gunsteren, S. R. Billeter, A. A. Eising, P. H. Hünenberger, P. K. H. C. Krüger, A. E. Mark, W. R. P. Scott, and I. G. Tironi. *Biomolecular simulation: The $\{GROMOS96\}$ manual and user guide*. Hochschulverlag AG an der ETH Zürich, Zürich, Switzerland, 1996.

[52] J. Wang, P. Cieplak, and P. A. Kollman. How well does a restrained electrostatic potential (resp) model perform in calculating conformational energies of organic and biological molecules? *Journal of Computational Chemistry*, 21:1049--1074, 2000.

[53] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling. Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function, and Bioinformatics*, 65:712--725, 2006.

[54] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, and D. E. Shaw. Improved side-chain torsion potentials for the amber ff99sb protein force field. *Proteins: Structure, Function, and Bioinformatics*, 78:1950--1958, 2010.

[55] M. B. Kubitzki and B. L. De Groot. Molecular dynamics simulations using temperature-enhanced essential dynamics replica exchange. *Biophysical journal*, 92:4262--4270, 2007.

[56] R. C. Bernardi, M. C. R. Melo, and K. Schulten. Enhanced sampling techniques in molecular dynamics simulations of biological systems. *Biochimica et Biophysica Acta (BBA)-General Subjects*, 1850:872--877, 2015.

[57] S. A. Adcock and J. A. McCammon. Molecular dynamics: survey of methods for simulating the activity of proteins. *Chemical reviews*, 106:1589--1615, 2006.

[58] G. M. Crippen and H. A. Scheraga. Minimization of polypeptide energy, viii. application of the deflation technique to a dipeptide. *Proceedings of the National Academy of Sciences*, 64:42--49, 1969.

[59] G. M. Torrie and J. P. Valleau. Nonphysical sampling distributions in monte carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics*, 23:187--199, 1977.

[60] T. Huber, A. E. Torda, and W. F. Gunsteren. Local elevation: a method for improving the searching properties of molecular dynamics simulation. *Journal of computer-aided molecular design*, 8:695--708, 1994.

[61] R. K. Hart, R. V. Pappu, and J. W. Ponder. Exploring the similarities between potential smoothing and simulated annealing. *Journal of Computational Chemistry*, 21:531--552, 2000.

[62] H. Grubmüller. Predicting slow structural transitions in macromolecular systems: Conformational flooding. *Physical Review E*, 52:2893, 1995.

[63] B. G. Schulze, H. Grubmüller, and J. D. Evanseck. Functional significance of hierarchical tiers in carbonmonoxy myoglobin: conformational substates and transitions studied by conformational flooding simulations. *Journal of the American Chemical Society*, 122:8700--8711, 2000.

[64] O. F. Lange, L. V. Schäfer, and H. Grubmüller. Flooding in gromacs: Accelerated barrier crossings in molecular dynamics. *Journal of computational chemistry*, 27:1693--1702, 2006.

[65] A. F. Voter. Hyperdynamics: Accelerated molecular dynamics of infrequent events. *Physical Review Letters*, 78:3908, 1997.

[66] A. F. Voter. A method for accelerating the molecular dynamics simulation of infrequent events. *The Journal of chemical physics*, 106:4665--4677, 1997.

[67] D. Hamelberg, J. Mongan, and J. A. McCammon. Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. *The Journal of chemical physics*, 120:11919--11929, 2004.

[68] A. Laio and M. Parrinello. Escaping free-energy minima. *Proceedings of the National Academy of Sciences*, 99:12562--12566, 2002.

[69] S. Kirkpatrick, D. C. Gelatt, M. P. Vecchi, et al. Optimization by simulated annealing. *science*, 220:671--680, 1983.

[70] Y. Sugita and Y. Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chemical physics letters*, 314:141--151, 1999.

[71] A. Mitsutake, Y. Sugita, and Y. Okamoto. Generalized-ensemble algorithms for molecular simulations of biopolymers. *Peptide Science*, 60:96--123, 2001.

[72] J.-P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics*, 23:327--341, 1977.

[73] H. Bekker, H. J. C. Berendsen, E. J. Dijkstra, S. Achterop, R. van Drunen, D. van der Spoel, A. Sijbers, H. Keegstra, B. Reitsma, and M. K. R. Renardus. Gromacs: a parallel computer for molecular dynamics simulations. In *Physics Computing*, volume 92. World Scientific, 1993.

[74] H. J. C. Berendsen, D. van der Spoel, and R. van Drunen. Gromacs: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications*, 91:43--56, 1995.

[75] E. Lindahl, B. Hess, and D. van der Spoel. Gromacs 3.0: a package for molecular simulation and trajectory analysis. *Journal of Molecular Modeling*, 7:306--317, 2001.

[76] D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen. Gromacs: fast, flexible, and free. *Journal of computational chemistry*, 26:1701--1718, 2005.

[77] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl. Gromacs 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of chemical theory and computation*, 4:435--447, 2008.

[78] S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, et al. Gromacs 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, page btt055, 2013.

[79] A. E. García. Large-amplitude nonlinear motions in proteins. *Physical review letters*, 68:2696, 1992.

[80] 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*, 403:567--571, 2000.

[81] F. C. Bernstein, T. F. Koetzle, G. J.B. Williams, E. F. Meyer Jr., M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi. The protein data bank: A computer-based archival file for macromolecular structures. *Journal of Molecular Biology*, 112:535--542, 1977.

[82] W. Humphrey, A. Dalke, and K. Schulten. Vmd: Visual molecular dynamics. *Journal of Molecular Graphics*, 14:33--38, 1996.

[83] Maestro 9.7. Schrödinger Inc. New York NY 2014.

[84] Y. Liu, R. A. Kahn, and J. H. Prestegard. Dynamic structure of membrane-anchored ARF · GTP. *Nature structural & molecular biology*, 17:876--881, 2010.

[85] J. L. Banks, H. S. Beard, Y. Cao, A. E. Cho, W. Damm, R. Farid, A. K. Felts, T. A. Halgren, D. T. Mainz, J. R. Maple, et al. Integrated modeling program, applied chemical theory (impact). *Journal of computational chemistry*, 26:1752--1780, 2005.

[86] C. Gao and L. A. Eriksson. Impact of mutations on k-ras-p120gap interaction. *Computational Molecular Bioscience*, 3:9--17, 2013.

[87] K. L. Meagher, L. T. Redman, and H. A. Carlson. Development of polyphosphate parameters for use with the amber force field. *Journal of computational chemistry*, 24:1016--1025, 2003.

[88] C. I. Bayly, P. Cieplak, W. Cornell, and P. A. Kollman. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the resp model. *The Journal of Physical Chemistry*, 97:10269--10280, 1993.

[89] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox. Gaussian09 Revision D.01. Gaussian Inc. Wallingford CT 2009.

[90] G. Todde, S. Hovmöller, A. Laaksonen, and F. Mocci. Glucose oxidase from penicillium amagasakiense: Characterization of the transition state of its denaturation from molecular dynamics simulations. *Proteins: Structure, Function, and Bioinformatics*, 82:2353--2363, 2014.

[91] C. González-Bello, L. Tizón, E. Lence, J. M. Otero, M. J. van Raaij, M. Martinez-Guitian, A. Beceiro, P. Thompson, and A. R. Hawkins. Chemical modification of a dehydratase enzyme involved in bacterial virulence by an ammonium derivative: Evidence of its active site covalent adduct. *Journal of the American Chemical Society*, 137:9333--9343, 2015.

[92] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case. Development and testing of a general amber force field. *Journal of computational chemistry*, 25:1157--1174, 2004.

[93] J. Wang, W. Wang, P. A. Kollman, and D. A. Case. Automatic atom type and bond type perception in molecular mechanical calculations. *Journal of molecular graphics and modelling*, 25:247--260, 2006.

[94] C. E. A. F. Schafmeister, W. S. Ross, and V. Romanovski. Leap. *University of California, San Francisco*, 1995.

[95] D. A. Case, J. T. Berryman, R. M. Betz, D. S. Cerutti, T. E. Cheatham III, T. A. Darden, R. E. Duke, T. J. Giese, H. Gohlke, A. W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T. S. Lee, S. LeGrand, P. Li, T. Luchko, R. Luo, B. Madej, K. M. Merz, G. Monard, P. Needham, H. Nguyen, H. T. Nguyen, I. Omelyan, A. Onufriev, D. R. Roe, A. Roitberg, R. Salomon-Ferrer, C. L. Simmerling, W. Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X. Wu, D.M. York, and P.A. Kollman. Amber 2015. *University of California, San Francisco*, 2015.

[96] A. W. S. da Silva and W. F. Vranken. Acpype-antechamber python parser interface. *BMC research notes*, 5:367, 2012.

[97] T. L. Rodgers, D. Burnell, P. D. Townsend, E. Pohl, M. J. Cann, M. R. Wilson, and T. C. B. McLeish. $\Delta\Delta$PT: a comprehensive toolbox for the analysis of protein motion. *BMC bioinformatics*, 14:183, 2013.

[98] P. Durand, G. Trinquier, and Y.-H. Sanejouand. A new approach for determining low-frequency normal modes in macromolecules. *Biopolymers*, 34:759--771, 1994.

[99] K. Suhre and Y.-H. Sanejouand. Elnemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement. *Nucleic acids research*, 32:W610--W614, 2004.

[100] S. P. Tiwari, E. Fuglebakk, S. M. Hollup, L. Skjærven, T. Cragnolini, S. H. Grindhaug, K. M. Tekle, and N. Reuter. Webnm@ v2.0: Web server and services for comparing protein flexibility. *BMC bioinformatics*, 15:427, 2014.

[101] W. Zheng and S. Doniach. A comparative study of motor-protein motions by using a simple elastic-network model. *Proceedings of the National Academy of Sciences*, 100:13253--13258, 2003.

[102] A. Kitao and N. Gō. Conformational dynamics of polypeptides and proteins in the dihedral angle space and in the cartesian coordinate space: Normal mode analysis of deca-alanine. *Journal of computational chemistry*, 12:359--368, 1991.

[103] A. Kitao, S. Hayward, and N. Gō. Comparison of normal mode analyses on a small globular protein in dihedral angle space and cartesian coordinate space. *Biophysical chemistry*, 52:107--114, 1994.

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

[105] M. H. M. Olsson, C. R. Søndergaard, M. Rostkowski, and J. H. Jensen. Propka3: consistent treatment of internal and surface residues in empirical pka predictions. *Journal of Chemical Theory and Computation*, 7:525--537, 2011.

[106] C. R. Søndergaard, M. H. M. Olsson, M. Rostkowski, and J. H. Jensen. Improved treatment of ligands and coupling effects in empirical calculation and rationalization of pka values. *Journal of Chemical Theory and Computation*, 7:2284--2295, 2011.

[107] H. J. C. Berendsen, J. Pl. M. Postma, W. F. van Gunsteren, A. R. H. J. DiNola, and J. R. Haak. Molecular dynamics with coupling to an external bath. *The Journal of chemical physics*, 81:3684--3690, 1984.

[108] G. Bussi, D. Donadio, and M. Parrinello. Canonical sampling through velocity rescaling. *The Journal of chemical physics*, 126:014101, 2007.

[109] S. Nosé. A molecular dynamics method for simulations in the canonical ensemble. *Molecular physics*, 52:255--268, 1984.

[110] W. G. Hoover. Canonical dynamics: equilibrium phase-space distributions. *Physical Review A*, 31:1695, 1985.

[111] M. Parrinello and A. Rahman. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied physics*, 52:7182--7190, 1981.

[112] S. Nosé and M. L. Klein. Constant pressure molecular dynamics for molecular systems. *Molecular Physics*, 50:1055--1076, 1983.

[113] M. G. de S. Costa, P. R. Batista, P. M. Bisch, and D. Perahia. Exploring free energy landscapes of large conformational changes: Molecular dynamics with excited normal modes. *Journal of Chemical Theory and Computation*, 2015.

[114] J. Peng and Z. Zhang. Simulating large-scale conformational changes of proteins by accelerating collective motions obtained from principal component analysis. *Journal of Chemical Theory and Computation*, 10:3449--3458, 2014.

[115] D. van der Spoel, E. Lindahl, B. Hess, and the GROMACS development team. Gromacs user manual version 4.6.7, 2014. www.gromacs.org.

[116] J. M. Goodfellow. *Molecular dynamics: applications in molecular biology*, volume 16. CRC Press, 1990.

[117] B. T. M. Willis and A. W. Pryor. *Thermal vibrations in crystallography*, volume 1. Cambridge University Press Cambridge, 1975.

[118] S. Kalinin, T. Peulen, S. Sindbert, P. J. Rothwell, S. Berger, T. Restle, R. S. Goody, H. Gohlke, and C. A. M. Seidel. A toolkit and benchmark study for fret-restrained high-precision structural modeling. *Nature methods*, 9:1218--1225, 2012.

[119] M. Lu, B. Poon, and J. Ma. A new method for coarse-grained elastic normal-mode analysis. *Journal of chemical theory and computation*, 2:464--471, 2006.

[120] C. Dovengerds. *Membraninteraktionen von großen Dynamin-verwandten GTPasen*. PhD thesis, Ruhr-Universität Bochum, Bochum, Germany, 2013.

[121] M. Schwemmle and P. Staeheli. The interferon-induced 67-kda guanylate-binding protein (hGBP1) is a GTPase that converts GTP to GMP. *Journal of Biological Chemistry*, 269:11299--11305, 1994.

[122] J. Goldberg. Structural basis for activation of ARF GTPase: mechanisms of guanine nucleotide exchange and GTP-myristoyl switching. *Cell*, 95:237--248, 1998.

[123] I. R. Vetter and A. Wittinghofer. The guanine nucleotide-binding switch in three dimensions. *Science*, 294:1299--1304, 2001.

[124] B. Prakash, L. Renault, G. J. K. Praefcke, C. Herrmann, and A. Wittinghofer. Triphosphate structure of guanylate-binding protein 1 and implications for nucleotide binding and GTPase mechanism. *The EMBO Journal*, 19:4555--4564, 2000.

[125] N. Abdullah, B. Srinivasan, N. Modiano, P. Cresswell, and A. K. Sau. Role of individual domains and identification of internal gap in human guanylate binding protein-1. *Journal of Molecular Biology*, 386:690--703, 2009.

[126] P. Oelschlaeger, M. Klahn, W. A. Beard, S. H. Wilson, and A. Warshel. Magnesium-cationic dummy atom molecules enhance representation of dna polymerase $\beta$ in molecular dynamics simulations: Improved accuracy in studies of structural features and mutational effects. *Journal of molecular biology*, 366:687--701, 2007.

[127] F. Duarte, P. Bauer, A. Barrozo, B. A. Amrein, M. Purg, J. Åqvist, and S. C. L. Kamerlin. Force field independent metal parameters using a nonbonded dummy model. *The Journal of Physical Chemistry B*, 118:4351--4362, 2014.

[128] Q. Liao, S. C. L. Kamerlin, and B. Strodel. Development and application of a nonbonded $Cu^{2+}$ model that includes the jahn-teller effect. *The journal of physical chemistry letters*, 6:2657--2662, 2015.

[129] P. Rice, I. Longden, A. Bleasby, et al. Emboss: the european molecular biology open software suite. *Trends in genetics*, 16:276--277, 2000.