Computational Analysis of Binding of the GBD Domain of WASP to Different Binding Partners

Maria K. Janowska,[a,b,c] Ruben Zubac,[c,d,e] and Bojan Zagrovic[c,d,e,*]

[a]*Department of Biochemistry, Biophysics and Biotechnology, Jagiellonian University,30-387 Krakow, Poland*
[b]*Department of Chemistry & Chemical Biology, Rutgers University, Piscataway, NJ 08854, USA*
[c]*Mediterranean Institute for Life Sciences, Split, HR-21000, Croatia*
[d]*Department of Physics, University of Split, Split, HR-21000, Croatia*
[e]*Department of Structural and Computational Biology, Max F. Perutz Laboratories, University of Vienna, Campus Vienna Biocenter 5, Vienna, AT-1030, Austria*

Abstract. The GTP-ase binding domain (GBD) of the signaling protein Wiskott-Aldrich Syndrome Protein (WASP) is intrinsically disordered and mutations in it have been linked with Wiskott-Aldrich Syndrome (WAS), an X-linked disorder characterized by thrombocytopenia, eczema and recurrent infections. Here, we use molecular dynamics simulations and the semi-empirical GROMOS 45A3 force field to study interaction of the GBD domain of WASP with a fragment of the protein $EspF_U$ as well as with the VCA domain of WASP (auto-inhibited state). $EspF_U$ is secreted and used by enterohaemorrhagic *Escherichia coli* to hijack eukaryotic cytoskeletal machinery, and it does so by competitively disrupting the auto-inhibitory interaction between GBD and VCA domains of WASP. In addition, naturally occurring mutations in the VCA domain cause different variants of WAS. Our simulations confirm that the $EspF_U$ domain binds the GBD domain similarly to the VCA domain, which explains why these two binding partners are competitive binders of the GBD domain. Furthermore, we propose a possible mechanism to explain the higher affinity of $EspF_U$ for the GBD domain. Finally, we show that the mutations in the VCA domain responsible for Wiskott-Aldrich syndrome can cause formation of β-sheets in the VCA domain. This effect, combined with the mutation-induced rearrangement of the salt bridge network, consequently disables tight binding between GBD and VCA domains. Overall, our results provide a microscopic, dynamic picture behind the two main ways through which the interactions involving the GBD domain of WASP participate in different disease processes.(doi: 10.5562/cca1806)

Keywords: Wiskott-Aldrich syndrome protein, GBD domain, VCA domain, $EspF_U$, molecular dynamics, GROMOS 45A3 force field

INTRODUCTION

Intrinsically disordered proteins (IDPs) or protein fragments are structurally unfolded in their native state. These proteins, also known as natively unfolded or intrinsically unstructured, are characterized by a lack of stable secondary and tertiary structure when unbound in solution.[1,2] This major characteristic of IDPs has attracted significant attention of the research community as it challenges the canonical paradigm of structural biology that ordered tertiary structure is the principal determinant of protein function. Remarkably, IDPs are able to carry out specific functions without well-defined secondary or tertiary structure.[3,4] What is more, IDPs or intrinsically disordered fragments are very common in

nature. For example, in eukaryotic cells, 30 % of all proteins are mostly unfolded, 50 % have long unfolded fragments, and 70 % of all signaling proteins are unfolded under native conditions.[5–7] Functionally, IDPs are involved in a number of different cellular processes including molecular recognition, signal transduction and protein modification,[1,4,6,7] and have been implicated in different diseases such as cancer, diabetes and amyloid formation.[8,9] Importantly, many IDPs fold into well-defined structures upon binding to their targets,[10] although folding does not always occur.[11] Finally, a number of IDPs have multiple binding partners, suggesting that novel physico-chemical mechanisms may be involved, compared with the typically less promiscuous, ordered proteins. For example, it has been shown that

---

† This article belongs to the Special Issue *Chemistry of Living Systems* devoted to the intersection of chemistry with life.
* Author to whom correspondence should be addressed. (E-mail: bojan.zagrovic@univie.ac.at)
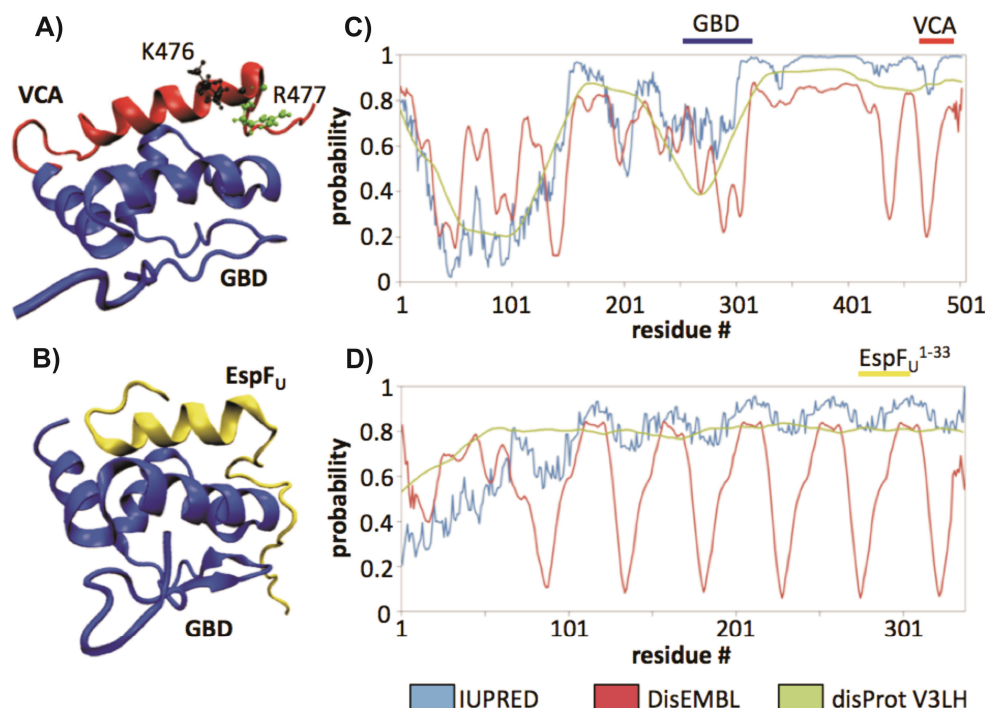
Figure 1. GBD domain complexes and disorder-prediction profiles of their components. A) GBD domain in complex with the VCA domain (WASP auto-inhibited form), B) GBD domain in complex with EspF$_U$. In both A and B, we show the original NMR structures (1st model). C) Disorder prediction by IUPRED, disProt VL3H and DisEMBL for WASP, including GBD and VCA domains, and D) for the complete EspF$_U$. Segments that were simulated in the present study are indicated on top of the C and D panels.

IDPs can fold into different conformations while binding to different binding partners.[4,12]

The GTP-ase binding domain (GBD domain) of the Wiskott-Aldrich Syndrome Protein (WASP) is a representative member of the IDP class. According to different biophysical characterizations, including circular dichroism and NMR,[13] the GBD domain of WASP is largely disordered. WASP is a signaling protein which carries signals from membrane receptors to proteins of the actin cytoskeleton complex.[14] Mutations in the WASP gene cause a rare X-linked disorder, Wiskott-Aldrich Syndrome (WAS), characterized by different symptoms depending on the type of mutation in the WASP gene. These include thrombocytopenia, eczema and recurrent infections caused by immunodeficiency.[15,16] The fragments of WASP that are known to be unfolded in the native, unbound state are the above-mentioned GBD domain and the VCA domain (where the latter abbreviation stands for Verprolin homology, Central hydrophobic and Acidic regions)[17] (Figure 1A). The GBD domain (residues 230 to 310 of WASP) can bind to the GTP-ase Cdc42, but also to the bacterial infection protein EspF$_U$.[13,18−21] Furthermore, the GBD domain also binds to the VCA domain of the same WASP molecule (residues 430 to 502),[13] and this process leads to auto-inhibition of WASP. Importantly,

while binding to Cdc42, the GBD domain folds to a different structural and functional state as compared to the complex with the VCA domain or with EspF$_U$. In the latter two complexes, the GBD domain folds in a similar way (Figure 1).[4,13,17,18]

WASP is inactive when the GBD domain is bound to the VCA domain. When the auto-inhibited structure is destabilized, the VCA domain of WASP is released and can interact with the Arp2/3 actin-nucleating complex, while the GBD domain binds to Cdc42 in the activated state.[13,18−21] EspF$_U$ is a protein that enables enterohaemorrhagic *E. coli* (EHEC) to take control over the actin cytoskeleton of the host by bringing under control members of the WASP family.[18−20] The mechanism of activation of WASP by EspF$_U$ is simple and direct: it binds to the GBD/VCA binding interface and displaces the VCA domain from the GBD domain in a competitive fashion.[18−20] More specifically, it has been shown that a single repeat of the EspF$_U$ protein binds WASP GBD with a dissociation constant $K_D = 18$ nM,[19] which is about 100-fold stronger binding than that of GBD and VCA.[13,19] The release of the auto-inhibitory GBD/VCA interaction, in turn, leads to the stimulation of the actin nucleation factor, the Arp2/3 complex,[21] which subsequently leads to the formation of actin pedestal, an actin-rich protrusion in the membrane under-

neath the site where EHEC attach to host cells. In effect, the structural mimicry and competitive displacement by EspF$_U$ of the VCA domain from the GBD domain of WASP is the cornerstone underlying EHEC infectivity on the molecular level.

The main goal of the present work is to understand the binding behavior of the GBD domain and to find differences and similarities between EspF$_U$ and the VCA domain, which form similar complexes with the GBD domain. In particular, we focus on elucidating structural and thermodynamic rationales for the greater affinity of EspF$_U$ for the GBD domain compared with the VCA domain.[13,18–20] Second, we analyze the mutations in the VCA domain responsible for WAS, and elucidate the microscopic background of the destabilizing effect they have on the GBD/VCA complex. These mutations in the VCA domain occur in nature and cause the typical form of WAS (K476E)[13,22] and the XLT form of WAS (R477K).[13,23] Overall, our ambition is to provide a microscopic, dynamic picture behind these two main ways, one mediated by an externally introduced protein and the other one by inherited mutations, through which the interactions involving the GBD domain of WASP participate in different disease processes.

METHODS

Molecular dynamics (MD) simulations were performed starting from two PDB structures containing the GBD domain: 1) GBD domain in complex with VCA domain, PDB code 1ej5,[13] and 2) GBD domain in complex with a fragment of the bacterial protein EspF$_U$, PDB code 2k42.[18] The 1ej5 structure is the auto-inhibited form of WASP. It contains the GBD domain (residues 242 to 310) connected by a short 6-residue linker with the cofilin fragment of the VCA domain of WASP (residues 461 to 492) to which it is bound. The 2k42 structure contains the GBD domain (residues 242 to 310) bound to a short fragment of the bacterial protein EspF$_U$ (residues 268 to 300, corresponding to residues 1−33 in this study), each preceded by a 3-residue linker. Analysis of the disordered regions of WASP and EspF$_U$ was performed using three different predictors: DisEMBL[24] (loops/coils definition), IUPRED[25] (long disorder mode) and disProt VL3H.[26] The molecular dynamics (MD) simulations were performed using Gromacs 3.3 software.[27] The following systems were simulated: 1) entire GBD complexes, *i.e.* the two structures from the PDB database mentioned above; 2) fragments of different complexes: GBD domain, VCA domain and the EspF$_U$ fragment alone, and 3) GBD domain in complex with a mutated VCA domain (K476E, R477K). Simulations of GBD and VCA domains from 1ej5 in isolation were performed with no linker included, while the simula-

tions of GBD and EspF$_U$ from 2k42 were performed with the preceding 3-residue linkers included in both domains. These linkers were excluded in the analysis, except for the calculation of solvent accessible surface area in the context of solvent entropy estimation, where inclusion of linkers was unavoidable. The two mutations in the VCA domain were introduced either simultaneously or each one separately using PDB2PQR software.[28,29]

The GROMOS96 45A3[30] force field and the SPC water model[31] were used for the simulations. All systems, except the VCA domain alone, were equilibrated using a protocol involving position restraints placed on all atoms and then gradually decreasing the strength of position restraints, while gradually increasing the temperature in 5 steps. Each of these steps was a 20 ps trajectory, simulated in the NVT ensemble. The 5 steps were conducted at the following temperatures (in order): 100, 150, 200, 250, 300 K and with the following respective position restraints (kJ mol^{-1} nm^{-2}, in all three principal directions): 25000, 20000, 15000, 10000, 5000. This equilibration procedure was carried out with a 2 fs time step, reaction field electrostatics, Berendsen temperature coupling[32] with $\tau_{coupling}$ = 0.1 ps, a van der Waals cutoff of 1.4 nm and SHAKE[33] constraints on all bonds. This was followed by 20 ps long equilibration simulations in the NPT ensemble at 300 K, with pressure kept at 1 atm using Berendsen pressure coupling[32] with $\tau_{coupling}$ = 0.1 ps and κ = 4.5 10^{-5} bar^{-1} and other settings unchanged. Production simulations were extended for 50 ns thereafter for all of the simulated systems using the same conditions. The VCA domain alone was equilibrated using a 2-stage simulated annealing procedure. In the first stage, the system was equilibrated at 400 K, with the protein position-restrained and the water allowed to move freely. In the second stage, the system was cooled to 5 K and then heated to 300 K by simulated annealing. A single sequence for annealing points was used, with two annealing reference/control points for each temperature group (protein and solvent) at 0 ps and 90 ps. In both stages, we used time-steps of 2 fs with 50000 steps, reaction field electrostatics and Berendsen temperature coupling.[32] This was followed by a 50 ns production run using the above settings. For all of structural analysis, coordinate frames were output every 80 ps. A single trajectory was simulated for each of the above systems except for the 2k42 structure. Namely, to test reproducibility, we also simulated four additional trajectories of the 2k42 structure with GBD bound to EspF$_U$ (for a total of five independent 50 ns trajectories), initiated using different velocity assignments.

Conformational entropy of all proteins in the present study was calculated using the quasi-harmonic approach employing all-atom, mass-weighted Cartesian coordinate covariance matrices.[34,35,36] In all cases, trans-

lational-rotational fitting was performed using all-atoms and standard GROMACS routines.[27] For entropy analysis, coordinate frames were output every 50 fs.

RESULTS

The GBD domain of WASP is known experimentally to be unfolded when unbound in solution and to fold upon binding.[13] Moreover, it is known that it folds to similar conformations when binding to the VCA domain and bacterial protein EspF$_U$[18–20] (Figure 1A and 1B. Overall, the backbone RMSD between the two NMR structures of the GBD domain is only 0.095 nm (residues 247−310).[18] Most notably, both ligands bind to the GBD domain at the same location, interacting with an extensive hydrophobic pocket formed by residues emanating from helices α 1−4 of GBD (Figure 1A and 1B). Importantly, in the original NMR structures the main point of contact between GBD and its ligands is through the hydrophobic patch. For example, the extensive disordered C-terminal tail of the EspF$_U$ domain does not directly interact with the GBD domain in the experimental NMR structure, despite the presence of a highly hydrophobic tryptophan residue at the fragment's C-terminus. As discussed below, our simulations shed a somewhat different light on this.

The GBD domain is an IDP *i.e.* it destabilizes (unfolds) once its binding partners are removed. Its IDP nature is apparent from a bioinformatic analysis of sequence disorder propensity, albeit to a lesser extent compared with some other regions of WASP (*e.g.* region 130−230 or 300−502). We have used three different, widely used algorithms for disorder prediction to study its disorder properties, with somewhat different results (Figure 1C and 1D). In particular, the IUPRED predictor,[24] which is based on estimating the capacity of residues to establish sufficient intramolecular interactions, assigns significant disorder probability to the entire GBD sequence (disorder probability ≈ 0.6−0.7). On the other hand, the DisEMBL predictor,[25] which uses a combination of different methods to predict disorder including DSSP-based loop assignment and crystallographic B-factor information, assigns noticeably lower disorder propensity to the GBD domain (0.4−0.5). A similar situation is also seen in the case of the disProt VL3H predictor,[26] a neural network-based algorithm. Overall, it is possible that this borderline disorder propensity of the GBD domain is related to the fact that this domain folds upon binding *i.e.* can exist in both folded and unfolded state depending on its molecular environment. Finally, both partners of the GBD domain, VCA and EspF$_U$, appear significantly intrinsically disordered when studied using IUPRED and DisEMBL algorithms (Figure 1C and 1D), while disProt VL3H produces a more checkered profile, albeit with significant stretches corresponding to complete disorder.
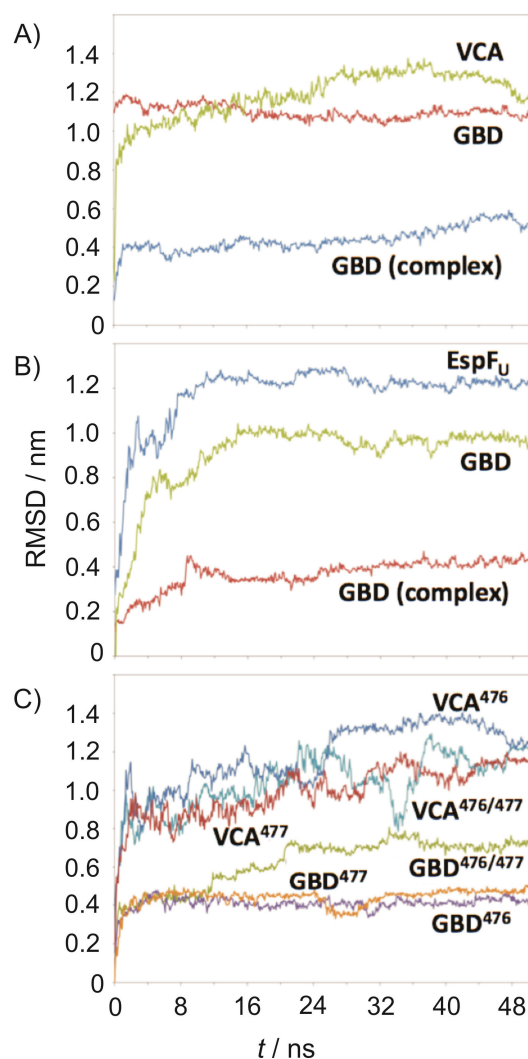
Figure 2. Backbone atom-positional RMSD from starting structures for different simulated systems. A) GBD/VCA complex: GBD domain simulated alone (red), GBD domain when in complex with VCA domain (blue), VCA domain alone (green); B) GBD/EspF$_U$: GBD domain simulated alone (green), GBD domain alone simulated in complex with EspF$_U$ (red) and EspF$_U$ domain alone (blue); C) GBD from a complex with K476E mutation (purple), GBD from a complex with R477K mutation (orange), GBD from a complex with both K476E and R477K mutations (green), VCA with K476E mutation (dark blue), VCA with R477K mutation (red), VCA with both K476E and R477K mutations (light blue). All simulations and calculations in C) were performed in the context of the GBD/VCA complex.

What are the main structural features of the GBD domain in the unbound form? To address this question, we have simulated two structures of the isolated GBD, one taken from the complex with the VCA domain and the second from the complex with EspF$_U$. The GBD domains alone were simulated starting from their li-

gand-bound conformations, after removing the binding partners. Analysis of the simulated trajectories shows that the GBD domains in the unbound form largely lose their secondary and tertiary structure, and the degree of change in the two folded GBD domains alone is similar: the average RMSD over the last 30 ns of the trajectory is approximately 1 nm for both GBD folded structures (Figure 2) (RMSD of GBD from complex with VCA = 1.08 ± 0.01 nm, RMSD of GBD from complex with $EspF_U$= 0.93 ± 0.02 nm). Furthermore, during simulation of the isolated GBD domain, the molecule lost a large fraction of its α-helical secondary structure content. Namely, it changed from 42.7 ± 2.8 % and 41.4 ± 2.6 % of all residues in the bound form to 24.8 ± 3.3 % and 20.1 ± 2.4 % in the unbound form, for VCA and $EspF_U$ complexes, respectively. These results confirm that the GBD domain without a binding partner is significantly destabilized, which is in agreement with prior biophysical studies[13] and the results of sequence-based predictions discussed above. Notably, in both cases, structural destabilization of the GBD domain upon ligand removal occurs without any major change in the molecule's radius of gyration, R_g (1.18 ± 0.02 nm versus 1.18 ± 0.02 nm for GBD from VCA and $EspF_U$ complexes, respectively) or solvent-accessible surface area, SASA (50.12 ± 1.69 nm^2, 49.97 ± 1.69 nm^2 for GBD from VCA and $EspF_U$ complexes, respectively). These values should be compared with the values from experimental structures of 1.14 nm (R_g of GBD from VCA complex), 1.17 nm (R_g of GBD from $EspF_U$ complex), and 51.11 nm^2 (SASA of GBD from VCA complex) and 50.59 nm^2 (SASA of GBD from $EspF_U$ complex). In other words, although it loses a significant part of its secondary and tertiary structure, GBD domain in isolation retains its overall compactness.

In contrast, the GBD domain in complex with its binding partners is significantly more stable than when alone. For example, the GBD domain in complex with VCA exhibits an average backbone RMSD from the starting structure of 0.47 ± 0.03 nm over the last 30 ns in the two trajectories combined together (Figure 2A). The same trend is seen during the simulation of the GBD-$EspF_U$ complex ($<RMSD>^{20-50\ ns}$ = 0.4 ± 0.02 nm, compared with the starting structure) (Figure 2B). Interestingly, when the two WAS-inducing mutations are introduced into the GBD+VCA structure simultaneously, the whole complex destabilizes in several tens of nanoseconds, as shown by the RMSD of mutated complexes which is almost twice as big as the RMSD of the complex without mutation (*e.g.* RMSD of GBD + VCA for the double mutant is 1.46 ± 0.02 nm, RMSD of GBD + VCA is 0.79 ± 0.03 nm, for the double mutant both with respect to the starting structure). Importantly, when it comes to both secondary and tertiary structure, these two mutations result in a similar level of structural dis-

integration of the GBD domain as when its ligands are removed (Figure 2C). However, when they are introduced individually, the GBD domain retains its structural stability throughout the simulated trajectories (Figure 2C).

Our simulations also allowed us to address a methodological question. Namely, as discussed above, there are a number of different sequence-based algorithms for predicting protein disorder. On the other hand, MD simulations give a direct view of protein disorder. How do the two approaches compare? We have carried out an analysis of the degree of disorder in the simulated unbound and bound states of GBD, VCA and $EspF_U$ to compare them with the results of sequence-based prediction algorithms (Figures 1 and S1, see Supplementary Materials for methodological details). Interestingly, there is significant qualitative agreement about the general degree of disorder on a site-resolved basis, especially in comparison with DisEMBL predictions, and to a lesser extent with IUPRED predictions (Figures 1 and S1). On the other hand, the results of the disPROT VL3H predictor deviate significantly from the MD results. In some ways, the difference between MD-predicted disorder and sequence-based predictions is of a similar order as that between predictions using different sequence-based algorithms themselves (Figure 1).

What is the mode of binding of $EspF_U$ to the GBD domain? According to Cheng *et al.*,[18] this protein binds to the GBD domain in a similar way to the VCA domain. Moreover, these researchers also found that residues 16 to 20 of $EspF_U$ form additional interactions with the GBD domain (as compared to the VCA domain), whereas residues 21 to 33 are reported to be largely unfolded and not interacting with the GBD domain (Figure 3A). Contrary to this, our simulations show that residues 21 to 33 of $EspF_U$ interact directly with the GBD domain (Figure 3B, C and 4). In particular, the tail region of $EspF_U$ appears to adopt a dynamic, U-shaped, extended conformation, encompassing almost half of the GBD domain. This structure is stabilized by a direct interaction between W33 in $EspF_U$ and different residues located on the diametrically opposite side of the GBD domain compared to the helix-binding site (Figures 3 and 4). Out of five independent simulations of this complex, four have resulted in $EspF_U$'s W33 being bound to GBD, with only one simulation resulting in an unbound W33 after 50 ns. For example, in one of the simulations W33 binds to the highly hydrophobic pocket formed by residues V254, L267 and L300 in GBD (Figure 3B and C). A similar, hydrophobic interaction between W33 and GBD's V264 is seen in another simulation, while in the remaining two the interactions are more of an electrostatic nature (*e.g.* with H253 and R308, Figure 4A). These additional interactions provide
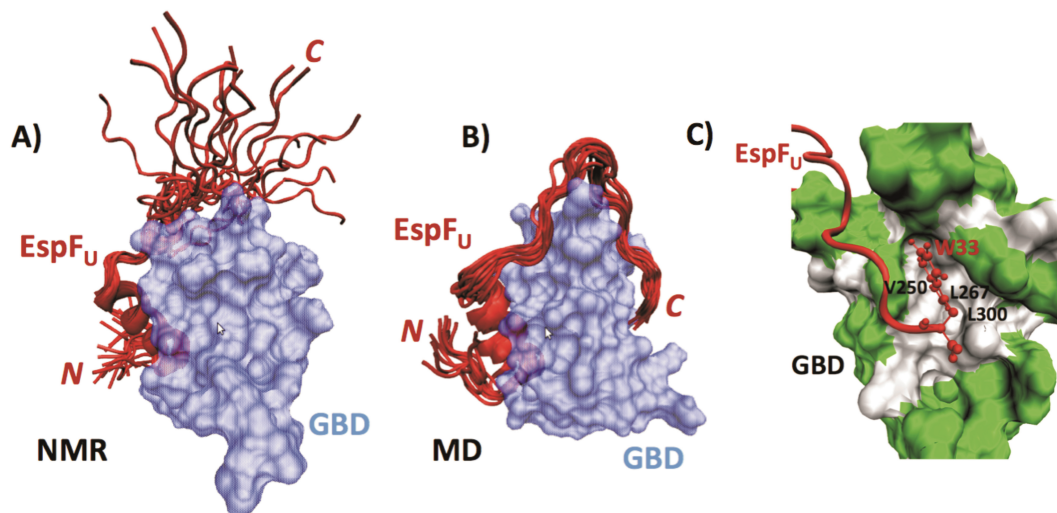
Figure 3. Binding of EspF$_U$ to GBD. Structures of the GBD/EspF$_U$ complexes from: A) NMR experiment (bundle of 20 lowest energy structures), B) MD simulation (bundle of 20 structures between 20 and 50 ns, spaced every 1.5 ns). C) close-up of the hydrophobic binding pocket for W33 seen in our simulations. Hydrophobic residues are colored white, while all other are green.

a potential explanation for EspF$_U$'s higher binding affinity, as discussed below. In terms of kinetics, the interactions between W33 in EspF$_U$ and different binding sites on GBD are typically established after approximately 10 ns and remain present for the rest of the trajectories (Figure 4B). We have also analyzed other potential interactions between the tail region of EspF$_U$ and GBD, but there were no other major conspicuous interactions seen. For example, potential salt-bridge-forming residues R21 and E25 in EspF$_U$ remain exposed to the solvent and do not form any interactions with the GBD domain.

Despite their relatively stable structure compared to that of either GBD or VCA or EspF$_U$ alone, the complexes of GBD with the two ligands are still dynamic, as

indicated by the relatively sizable RMSD from the starting structures (Figure 2). This suggests that conformational entropy may be a relevant factor determining the relative affinities of VCA and EspF$_U$ for the GBD domain. We have evaluated conformational entropy differences between EspF$_U$ and VCA complexes with the GBD domain, and the results are intriguing (Figure 4). First, the conformational entropy contribution to the free energy ($-T\Delta S_{QH}$) of just the GBD domain in complex with EspF$_U$ is approximately 250 kJ/mol more positive than that of GBD in complex with VCA for the period of 20−50 ns in our simulations, as evaluated by the quasi-harmonic approach at T = 300 K (Figure 5, *GBD*). Similarly, the conformational entropy part of free energy of just the VCA domain in complex with GBD is
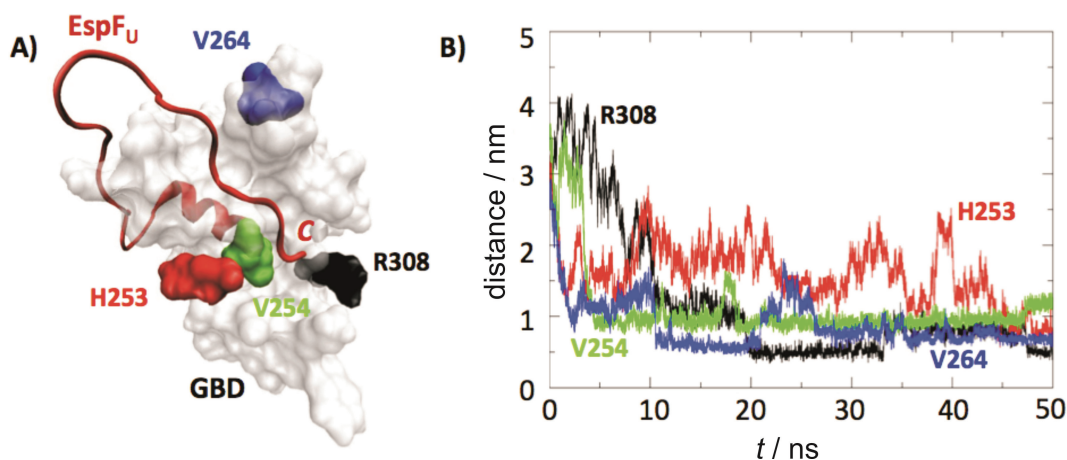
Figure 4. EspF$_U$ binds to different binding locations on GBD. A) locations of all of the detected binding locations on GBD for EspF$_U$'s C-terminal W33 residue. B) time course of binding of W33 to different locations on GBD.
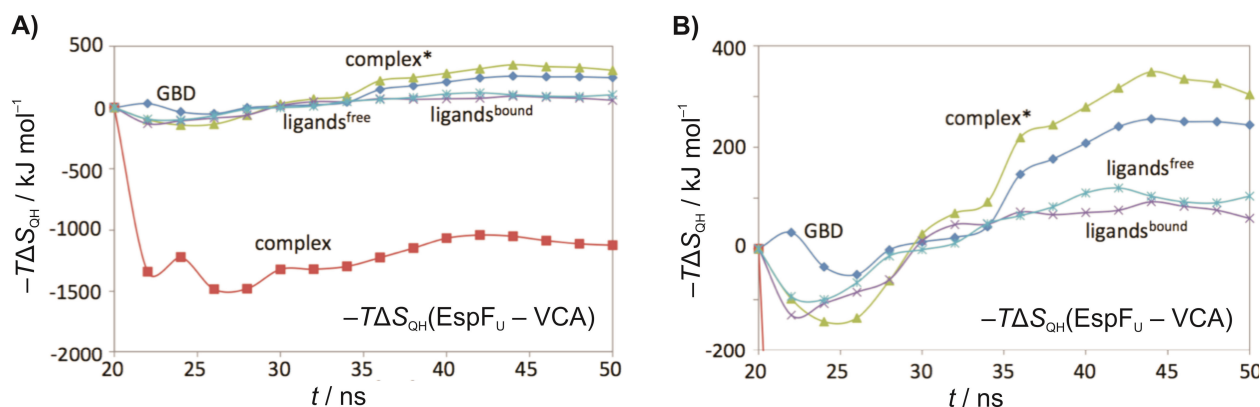
Figure 5. Analysis of the contribution of quasi-harmonic conformational entropy to free energy of GBD complexes. In all of the graphs in A), we plot cumulative values of $-T\Delta S_{QH}$, as a function of total time and normalized by the number of atoms, for the difference between $EspF_U$ and VCA-bound complexes, including the difference for the ligands (both bound and free), GBD domains, as well as for complete complexes, calculated either by performing quasi-harmonic conformational entropy analysis on individual components of the complex, and then adding the contributions up (*complex**), or on the whole complex (*complex*). B) close-up of A).

approximately 100 kJ/mol more favorable than that of the $EspF_U$ fragment over the same stretch (Figure 5, *ligands bound*). Interestingly, the difference is very similar for the unbound ligands as well (*ligands free*). On the other hand, if one analyzes complete complexes, it appears that the conformational entropy part of the free energy change is about 1500 kJ/mol more favorable for the GBD/$EspF_U$ than for the GBD/VCA complex at 300 K (Figure 5, *complex*). Curiously, this is only so if one treats the entire complex as one unit when it comes to rotation/translation fitting in the course of quasi-harmonic analysis. If, on the other hand, one calculates the conformational entropy of each complex as a sum of individual contributions coming from GBD and the ligand in question, one sees the opposite behavior: in that case, the VCA complex becomes more favorable by approximately 300 kJ/mol (Figure 5, *complex**). These results, taken together, suggest that the relative movement of the GBD domain and $EspF_U$ in the bound state, rather than their intrinsic dynamics, may be responsible for increasing the conformational entropy content of this complex. Finally, we have also evaluated the solvent entropy contribution to the binding free energy change for the two complexes. Assuming that the solvent entropy change is proportional to the change in solvent accessible surface area of the two complexes and their constituting fragments upon binding, and using the formalism of Freire *et al.*,[37,38] our simulations suggest that the solvent entropy contribution to the binding free energy change is about 28.4 kJ/mol more favorable at $T = 300$ K for the GBD/$EspF_U$ as compared to the GBD/VCA complex, which is significantly less in absolute magnitude than conformational entropy.
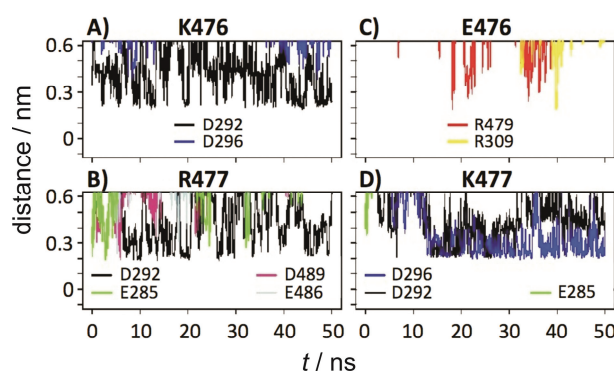
Binding of $EspF_U$ is an example of how modulation of GBD and VCA interactions can lead to a disease process and, in this case, modulation is caused by an externally introduced factor. However, inherited mutations can also affect these interactions and directly lead to abnormal physiological activity. As discussed above, mutations in the VCA domain (K476E and R477K) are known to cause different forms of WAS.[13,22,23] These mutations occur at the C-terminus of the VCA domain,[18] and we have introduced them in our simulations either simultaneously or one at a time. As a consequence of introducing the mutations simultaneously, the whole complex significantly destabilizes (Figure 2C). The RMSD for the entire structure reaches 1 nm (Figure 2A), even though RMSD for individual α-helical fragments in the GBD domain is small (<0.2 nm in all cases). After introducing the mutations, we also observe formation of a β-sheet in the VCA domain in the course of the simulation trajectory. Secondary structure time-series DSSP plots demonstrate the appearance of this additional secondary structure already within the first few nanoseconds of simulation (Figure 6). Importantly, GBD domain's helices maintain their structure, as indicated by a sizable fraction of α-helical residues over the last 30 ns of the trajectory (33.8 ± 1.9 %) but do start to move away from each other.

Finally, and this we believe is the main cause of structural disintegration, the doubly mutated GBD+VCA complex also exhibits a major rearrangement of the salt-bridge network keeping the complex together. Namely, in the native state, there is a strong interaction of K476 and R477 with the 292−296 segment of the GBD domain (K476 with D292 and D296, and R477 with D292 only) (Figure 7A and 7B). Moreo-
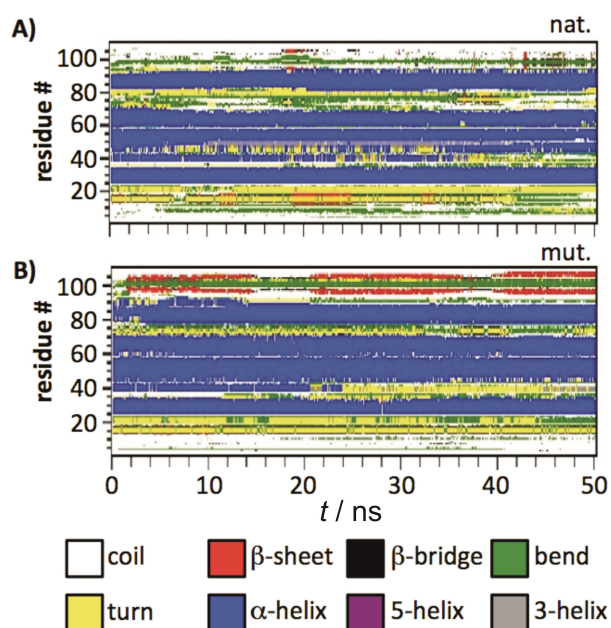
Figure 6. Secondary structure analysis of GBD/VCA complexes. A) GBD domain in complex with the VCA domain, and B) GBD domain in complex with the mutated VCA domain. The formation of an additional β-sheet at the N-terminus of the VCA domain occurs in the mutated complex.

Figure 7. Salt-bridge analysis of GBD/VCA complexes. Time-course of salt-bridge pairings: of A) native K476, B) native R477, C) mutated E476 and D) mutated K477.

ver, R477 also interacts with E285 and K288 (Figure 7C). The interactions with D292 and E285[13] have been reported to be very important for the GBD/VCA binding. The mutation K476E changes the positively charged residue into a negatively charged one and, expectedly, results in the loss of salt bridges at position 476 with the GBD domain. Consequently, what we observe is a shift of salt-bridge-forming „responsibility" in this region of the structure from the native K476/R477 pair to the mutated K477, joined by R479 whose salt-bridging network also significantly changes upon mutations of K476 and R477 (not shown). Most notably, R479, which before mutations formed salt bridges inside the VCA domain only, now interacts with two different GBD residues (D292 and E295). In addition, it also takes over the salt bridge with E476 (which interacted with R477 before the mutation). However, its interactions with D485 and E488 are now lost. Strong salt-bridge interactions are preserved with the D292−D296 segment of GBD (with an additional salt bridge forming with E295 and with the salt bridge to D296 appearing to be more stable and stronger) (Figure 7). However, salt bridges with E285 and K288 of the GBD domain appear to be lost. Moreover, K477 now fails to form salt bridges with E486 and D489 that R477 interacted with before the mutation. In short, the double mutation preserves strong interactions of VCA with the 292−296 segment of GBD, results in the loss of interaction with 285−288 of GBD and most importantly, leads

to a loss of salt bridges to D485, E486, E488 and D489 which we propose is responsible for the destabilization of the VCA domain. Finally, we postulate that the abrupt loss of salt bridge interactions within the 485−489 fragment favors the formation of a β-sheet, as discussed above, further destabilizing the native conformation of the VCA domain. Interestingly, individual mutations in the VCA domain do not lead to any major reorganization or unfolding of the GBD domain, but they do induce unfolding of the VCA domain similar to the one discussed here (Figure 2C). Finally, structural rearrangement of the GBD/VCA complex in the presence of mutations also leads to a change in its dynamic behavior. As the most important consequence, the conformational entropy of the complex decreases in all three cases, indicative of a reduction in the molecule's flexibility. More precisely, relative to the conformational entropy of the native complex (26.7 kJ mol^{-1} K^{-1}), the conformational entropy of the K476E mutant is 1.1 kJ mol^{-1} K^{-1} lower, that of the R477K mutant 1.7 kJ mol^{-1} K^{-1} lower, and that of the double mutant 1.3 kJ mol^{-1} K^{-1} lower. Using the conformational entropy, solvent entropy and enthalpy contributions, one can also estimate the free energy change of binding upon mutation. While the quantitative details are likely imprecise, using this approach, and assuming that the solvent-protein enthalpic contribution does not change, we obtain that the effect of each of the three mutations is, as expected, unfavorable compared to the wild type ($\Delta\Delta G_{mut}$ = 562.9 kJ/mol, 247.4 kJ/mol and 375.3 kJ/mol for K476E, R477K and the double mutant, respectively).

DISCUSSION

MD simulations are a powerful method for investigating the behavior of biomolecular systems on the atomistic level and with femtosecond temporal resolution[39]. Using different 50 ns MD simulations, we have shown that the EspF$_U$ fragment binds the GBD domain with additional

interactions of its C-terminal tail, which wraps around the GBD domain and could contribute to EspF$_U$'s higher binding affinity as a competitive binder to the VCA domain. Second, our simulations allowed us to dissect the contribution of conformational entropy to the difference between VCA and EspF$_U$ in binding to the GBD domain. Finally, our simulations have shown that the molecular basis of the destabilization of the GBD domain complex with the VCA domain consists primarily of the destabilization of salt bridges within the VCA domain.

Cheng *et al.*[18] have experimentally shown that the first 18 residues of the EspF$_U$ fragment bind to the GBD domain with a K_D of 4 ± 2 μM, while if the additional 15 residues of the C-terminal tail are included, the affinity rises to a K_D of 35 ± 4.9 nM. We would like to suggest that this additional stabilization of the complex is due to the strong interaction of W33 with different hydrophobic pockets on GBD, as discussed above, or other binding interactions of this residue with the GBD surface. The interaction of W33 with GBD was not detected by NMR,[18] but given the heterogeneity of its potential binding sites on GBD, as seen in our simulations, this is to be expected. Importantly, we did not detect any other major interactions (*e.g.* salt bridges) between the C-terminal tail of EspF$_U$ and the rest of GBD, suggesting that the above-mentioned hydrophobic and/or electrostatic interactions may be the key determinant of binding affinity. Further experimental work should shed more light on this possibility: in particular, our simulations predict that mutating W33 to a more hydrophilic amino acid would have a major effect in the direction of reducing the binding affinity of EspF$_U$. Similarly, we predict that structure-preserving mutations in the hydrophobic pockets on the GBD domain surrounding V254 and V264, whereby these pockets would be rendered more hydrophilic, would produce a similar effect.

Using quasi-harmonic approaches, we have calculated the conformational entropy of GBD and its complexes. Most importantly, we have found that when it comes to conformational entropy, binding of EspF$_U$ is more favorable compared to VCA, under the assumption that their conformational entropies in the unbound state are similar. Although our simulations were potentially not long enough to reach fully converged conformational entropy values, most notably in the case of disordered conformers, it is possible that the trends seen in our simulations would remain valid even with more extensive sampling. Additionally, quasi-harmonic analysis critically depends on precise roto-translational fitting of structures, and it is possible that in the case of more mobile complexes, such as GBD/EspF$_U$, this may be difficult to optimally perform. Limited sampling and fitting artifacts notwithstanding, this analysis suggests

that a part of the rationale for higher affinity of EspF$_U$ for GBD, and for its competitive displacement of the VCA domain, may reside in conformational entropy contributions. However, there are likely other rationales for this, both entropic and enthalpic in nature. For example, we would like to suggest that the fact that EspF$_U$ is actually expressed as a sequence of 5−6 highly similar repeats[19] (Figure 1D) increases to a major degree the affinity of EspF$_U$ for GBD. By concatenating 5−6 EspF$_U$ binding motifs, the overall affinity of the complete EspF$_U$ for its target is multiplicatively increased. Sallee *et al.*[19] have demonstrated that multiple repeats of EspF$_U$ are needed for powerful activation of actin polymerization, and have suggested that this is due to EspF$_U$ binding multiple WASP proteins. Of course, this mechanism is largely independent of our suggestion that multiple repeats increase affinity – it is possible that both mechanisms operate in parallel. Finally, it is also possible that the favorable conformational entropy of the EspF$_U$ complex may be counteracted by an opposite trend when it comes to EspF$_U$ or VCA alone in solution.

By analyzing and comparing in parallel two different molecular scenarios which lead to disease in the context of WASP, we hope to have provided a foundation which could contribute to finding new pharmacological treatments for Wiskott-Aldrich Syndrome. In addition, we hope the findings presented herein will help harness the tremendous potential of IDP molecules like GBD, EspF$_U$ or VCA to be used in different contexts ranging from enzyme design and nanotechnology to biotechnology.

REFERENCES

1. A. F. Fink, *Curr. Opin. Struc. Biol.* **15** (2005) 35−41.
2. P. Radivojac, L. M. Iakoucheva, C. J. Oldfield, Z. Obradovic, and V. N. Uversky, *Biophys. J.* **92** (2007) 1439−1456.
3. A. Schlessinger, M. Punta, and B. Rost, *BMC Bioinformatics.* **23** (2007) 2376−2384.
4. H. J. Dyson and P. E. Wright, *Curr. Opin. Struct. Biol.* **12** (2002) 54−60.

5. J. J. Ward, S. J. Sodhi, L. J. McGuffin, B. F. Buxton, and D. T. Jones, *J. Mol. Biol.* **337** (2004) 635−645.
6. L. M. Iakoucheva, C. J. Brown, J. D. Lawson, Z. Obradovic, and A. K. Dunker, *J. Mol. Biol.* **323** (2002) 573−584.
7. A. K. Dunker, C. J. Brown, J. D. Lawson, L. M. Iakoucheva, and Z. Obradovic, *Biochemistry-US* **41** (2002) 6574−6582.
8. V. N. Uversky, C. J. Oldfield, and A. K. Dunker, *Ann. Rev. Biophys.* **37** (2008) 215−246.
9. V. N. Uversky, C. J. Oldfield, U. Midic, H. Xie, B. Xue, S. Vucetic, L. M. Iakoucheva, Z. Obradovic, and A. K. Dunker, *BCM Genomics* **10(Suppl I)** (2009) S7.
10. V. N. Uversky, *Protein Sci.* **11** (2002) 739−756.
11. A. B. Sigalov, A. V. Zhuravleva, and V. Y. Orekhov, *Biochimie* **89** (2007) 419−421.
12. L. M. Espinoza-Fonseca, *Biochem. Bioph. Res. Co.* **382** (2009) 479−482.
13. A. S. Kim, L. T. Kakalis, N. Abdul-Manan, G. A. Liu, and M. K. Rosen, *Nature* **404** (2000) 151−158.
14. A. J. Thrasher and S. O. Burns, *Nat. Rev. Immunol.* **10** (2010) 182−192.
15. H. D. Ochs and A. J. Thrasher, *J. Allergy Clin. Immunol.* **117** (2006) 725−738.
16. L. D. Nortarangelo, C. H. Miao, and H. D. Ochs, *Curr. Opin. Hematol.* **15(1)** (2008) 30−36.
17. N. Abdul-Manan, B. Aghazadeh, A. G. Liu, A. Majumdar, O. Ouerfelli, S. A. Siminovitch, and M. S. Rosen. *Nature* **399** (1999) 379−383.
18. H. Cheng, B. M. Skehan, K. G. Campellone, J. M. Leong, and M. K. Rosen, *Nature* **454** (2008) 1009−1014.
19. N. A. Sallee, G. M. Rivera J. E. Dueber, D. Vasilescu, R. Dyche Mullins, B. J. Mayer, and W. A. Lim, *Nature* **454** (2008) 1005−1008.
20. S. B. Padrick and M. K. Rosen, *Annu. Rev. Biochem.* **79** (2010) 707−735.
21. H. N. Higgs and T. D. Pollard, *Annu. Rev. Biochem.* **70** (2001) 649−676.
22. R. Kolluri, A. Chehabeldin, M. Peacocke, A-M. Lamhonwah, K. Teichert-Kuliszewska, S. M. Weissman, and K. A. Simonovitch, *Hum. Mol. Genet.* **4** (1995) 1119−1126.
23. J. M. J. Derry, J. A. Kerns, K. I. Weinberg, H. D. Ochs, V. Volpini, X. Estivill, A. P. Walker, and U. Francke, *Hum. Mol. Genet.* **4** (1995) 1127−1135.
24. Z. Dosztanyi, V. Csizmok, P. Tompa, and I. Simon, *Bioinformatics.* **21** (2005) 3433−3434.
25. R. Linding, L. J Jensen, F. Diella, P. Bork, T. J. Gibson, and R. B. Russell, *Structure* **11** (2003) 1453−1459.
26. Z. Obradovic, K. Peng, S. Vucetic, P. Radivojac, C. J. Brown, and A. K. Dunker, *Prot. Struc. Funct. Genet.* **53** (2003) 566−572.
27. D. van der Spoel, E. Lindahl, B. Hess, A. R. van Buuren, E. Apol, P. J. Meulenhoff, D. P. Tieleman, A. L. T. M. Sijbers, K. A. Feenstra, R. van Drunen, and H. J. C. Berendsen, Gromacs User Manual version 3.3 2005.
28. T. J. Dolinsky, P. Czodrowski, H. Li, J. E. Nielsen, J. H. Jensen, G. Klebe, and N. A. Baker NA, *Nucleic Acids Res.* **35** (2007) W522-5.
29. T. J. Dolinsky, J. E. Nielsen, J. A. McCammon, and N. A. Baker, *Nucleic Acids Res.* **32** (2004) W665−W667.
30. L. D. Schuler, X. Daura, and W. F. van Gunsteren, *J. Comput. Chem.* **22** (2001) 1205−1218.
31. H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Hermans, Intermolecular Forces Pullman B. (ed.). Reidel, Dordrecht, 1981, p. 331.
32. H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J.R. Haak, *J. Chem. Phys.* **81** (1984) 3684−3690.
33. J-P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen. *J. Comput. Phys.* **23** (1977) 327−341.
34. M. Karplus and J. N. Kushick, *Macromolecules* **14** (1981) 325−332.
35. R. M. Levy, M. Karplus, J. Kushick, and D. Perahia, *Macromolecules* **17** (1984) 1370−1374.
36. I. Andricioaei and M. Karplus. *J. Chem. Phys.* **115** (2001) 6289−6292.
37. J. A. D'Aquino, J. Gomez, V. J. Hilser, K. H. Lee, L. M. Amzel, and E. Freire. *Prot. Struc. Funct. Genet.* **25** (1996) 143−156.
38. V. J. Hilser, B. E. Garcia-Moreno, T. G. Oas, G. Kap, and S. T. Whitten. *Chem. Rev.* **106** (2006) 1545−1558.
39. W. F. van Gunsteren , D. Bakowies, R. Baron, I. Chandrasekhar, M. Christen, X. Daura, P. Gee, D. P. Geerke, A. Glattli, P. H. Hunenberger, M. A. Kastenholz, C. Oostenbrink, M. Schenk, D. Trzesniak, N. F. A. van der Vegt, and H. B. Yu, *Angew. Chem. – Inter. Ed.* **45** (2006) 4064−4092.