Student Conference Proceedings
Vol 2, No 1, 2026, Article ID 2680, 4 Pages

Student Conference Lübeck

Proceedings Article

Molecular dynamics simulations of growth factor independence 1 (GFI1)

Erik Gromodka [a,b,*] · Jan Vorwerk [c] · Hauke Paulsen [b]

[a] Study program Biophysics, Universität zu Lübeck, Lübeck, Germany
[b] Institute of Physics, Universität zu Lübeck, Lübeck, Germany
[c] Department of Hematology and Oncology, University Hospital of Schleswig-Holstein, University Cancer Center Schleswig-Holstein, Universität zu Lübeck, Lübeck, Germany
[*] Corresponding author, email: erik.gromodka@student.uni-luebeck.de; hauke.paulsen@uni-luebeck.de

Received 17 February 2026; Accepted 01 September 2026; Published online 08 September 2026

Abstract

Growth factor independence 1 (GFI1) is a DNA-binding transcription factor and a key regulator of hematopoiesis. GFI1 is comprised of three distinct sections: (i) the C-terminal region containing six zinc-finger domains, (ii) an N-terminal region of 20 amino acids termed SNAG, and (iii) a central region connecting the other two. As large parts of the GFI1 protein seem to be intrinsically disordered, the present study tries to explore its configurational variety by employing molecular dynamics (MD) simulations. From the obtained MD trajectories the most significant configurations are extracted using a cluster algorithm and the Bayesian information criterion (BIC). It turns out that no dominant structure could be identified, which emphasizes the expected intrinsic structural disorder of the protein. During the four microsecond simulation no significant difference was found between two different variants of GFI1.

I. Introduction

Acting as a transcriptional repressor, growth factor independence 1 (GFI1) has a molecular mass between 47 and 55 kDa [1]. It is present in different cell types and plays an important role in many biological processes including, for example, defense against pathogens and hematopoiesis. A variant of this protein, characterized by an exchange of serine (S) to asparagine (N) at position 36, has a significantly increased prevalence among patients with acute myeloid leukemia (ALM) [2]. GFI1 is comprised of three distinct sections: (i) the C-terminal region containing six zinc-finger domains, (ii) an N-terminal region of 20 amino acids termed SNAG, and (iii) a central region connecting the other two. While an X-ray structure for the C-terminal region is known, the central region seems to be intrinsically disordered and the SNAG region can at best be described as conditionally ordered.

The present study employs molecular dynamics (MD) simulations for both variants (GFI1-36S and GFI1-36N) for the purpose of exploring the configurational variety of the GFI1 variants and elucidating the effect of the mutation at position 36. From the obtained MD trajectories the most significant configurations are extracted using cluster algorithms.

II. Methods

A starting structure for a complex of GFI1-36S with six zinc ions and a DNA segment with 12 nucleotides has been generated using AlphaFold3 [3]. An analogous approach was used in the preparation of a preliminary structure for a complex of GFI-36N and DNA (GFI1-36N). These complexes were positioned at the center of cubic simulation boxes with periodic boundary conditions

10.18416/SCP.2026.2680

and a side length of 15 nm. Subsequently, 105700 water molecules, 316 Na^+ ions and 305 Cl^- ions were introduced to achieve a neutral system with a sodium concentration of 100 mmol/L. To equilibrate this system, an energy minimization with fixed protein and DNA coordinates was performed, followed by a short MD simulation in the *NVE* ensemble. A weak potential was employed to impede the diffusion of zinc ions. The equilibration of the system was continued with an MD simulation in the *NVT* ensemble at a temperature of 310 K followed by an MD simulation in the *NpT* ensemble at a pressure of 1 bar, while the coordinates of the protein and the DNA segment were still constrained. The equilibrated system served as the starting point for 5 µs and 4 µs simulations for GFI1-36S and GFI1-36N, respectively, with unconstrained protein coordinates and constrained DNA coordinates to generate the trajectory to be analyzed.

The MD simulations have been performed using the GROMACS 2025.3 [4] software suite together with the CHARMM36 force field and the TIP3P (transferable intermolecular potential with 3 points) water model. Electrostatic interactions were treated using the PME method [5] with a grid spacing of 0.12 nm and a direct space cut-off of 1.0 nm. The same cut-off was used for the van der Waals interactions. The leap frog algorithm was chosen as the numerical integrator. To allow a time step of 2 fs length, all bonds involving hydrogen atoms were constrained using standard constraint algorithms implemented in GROMACS [6]. Where appropriate, the v-rescale thermostat [7] and the Parrinello-Rahman barostat [8] were used to couple the systems to external temperature and pressure.

To provide an approximate picture of the conformational variety of GFI1, as obtained from the MD trajectory, representative conformations are determined using a classification algorithm [9]. This algorithm divides the set of n structures comprising the trajectory into k subsets i with n_i members, known as 'clusters'. The number of clusters, k, depends on a given cutoff, R_c. To avoid overfitting, this cutoff is chosen with the help of the Bayesian information criterion, given by

$$B = k \ln n - 2 \ln L \,, \tag{1}$$

where L is the likelihood function

$$L = C \prod_{i=1}^{k} \exp\left(-n_i \overline{R}_i^2 / 2\sigma^2\right), \tag{2}$$

with a normalization constant C, the averaged RMSD value $\overline{R}_i$ of any structure in cluster i with respect to all the other $n_i - 1$ structures within the same cluster and the averaged RMSD value σ of any structure with respect to all the other $n-1$ structures of the complete trajectory. Disregarding a constant, the Bayesian information criterion can then be written as

$$B = k \ln n + \sum_{i=1}^{k} \frac{n_i \overline{R}_i^2}{\sigma^2} \,. \tag{3}$$

The cutoff R_c that controls the clustering of the trajectory and hence the values for k, n_i and $\overline{R}_i$ is finally chosen in a way as to minimize B. In addition, the radius of gyration, R_g, the solvent-accessible surface area (SASA), A_s, and the maximum distance d of any two protein atoms are calculated every 200 ps.

III. Results

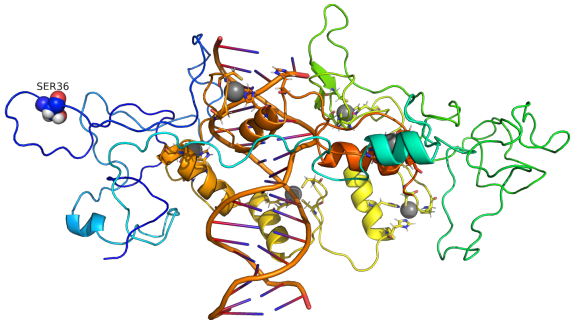

Figure 1: Simulated Structure of GFI1-36S. The cartoon representation is used for protein and DNA with a color spectrum from blue (N terminus) to red (C terminus). Zinc ions are depicted as grey spheres and histidin residues are represented by sticks. Residue 36S is highlighted by spheres.

Figure 1 shows a representative snapshot of the simulated structure of GFI1-36S in complex with a short DNA fragment.

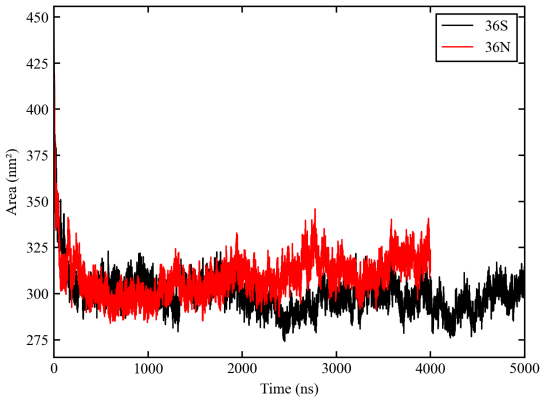

Figure 2: SASA of the GFI1 variants 36S (black) and 36N (red) calculated every 200 ps serving as a measure of protein compactness.

Figure 2 shows an early relaxation phase in which the SASA decreases significantly, indicating increasing compactness. This initial short phase is followed by a later adjustment, with partially overlapping values, before the variants reach a stable but differently positioned plateau phase. In the later simulation window (1500-4000 ns), the mean SASA of GFI1-36N ($\langle \text{SASA} \rangle = 311.85 \pm 8.30\,\text{nm}^2$)

10.18416/SCP.2026.2680

is higher than that of GFI1-36S ($\langle$SASA$\rangle = 298.14 \pm 7.54$ nm^2). The fluctuation amplitudes of the SASA are comparable for both variants, but differ in the mean level. Although it is not certain that the results obtained are statistically significant, this could indicate that GFI1-36S prefers more compact conformations whereas GFI1-36N is less compact on average.

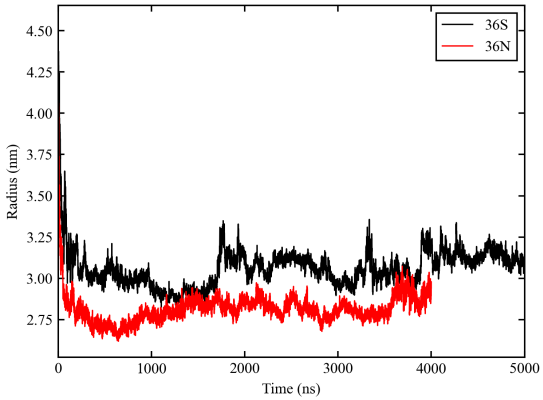

Figure 3: Rg of the GFI1 variants 36S (black) and 36N (red) calculated every 200 ps describing the global spatial extent of the protein.

For further analysis of global compactness, the radius of gyration (Rg) for both GFI1 variants was examined over time (see Fig. 3). Similar to the SASA a pronounced decrease in Rg is observed for both systems, with a stronger reduction for GFI1-36N than for GFI1-36S. Overall, GFI1-36N adopts more compact conformations over time than GFI1-36S. On average, the radius of gyration is $\langle$Rg$\rangle = 2.83 \pm 0.06$ nm and due to the comparatively low fluctuations, Rg remains within a narrow range of values. In contrast, GFI1-36S shows a higher mean Rg level of $\langle$Rg$\rangle = 3.06 \pm 0.08$ nm with sometimes significantly stronger temporal fluctuations. Expansions with Rg values above 3.2 nm occur repeatedly, followed by a reduction of the radius of gyration. The system alternates between conformational states of differing compactness. No clearly defined long-lasting plateau phase can be identified for GFI1-36S. The mean global expansion of GFI1-36S is thus approximately 0.23 nm higher overall compared to GFI1-36N.

To further characterize the conformational dynamics, the temporal evolution of the maximal intramolecular distances for the two GFI1 variants was analysed (see Fig. 4). GFI1-36S displays consistently larger mean maximum distances than GFI1-36N. After the initial relaxation and adaptation phase, the mean distance for GFI1-36S is $\langle$d$\rangle = 10.59 \pm 0.49$ nm, whereas a lower value of $\langle$d$\rangle = 9.29 \pm 0.40$ nm is observed for GFI1-36N. In addition, GFI1-36S exhibits significantly stronger temporal fluctuations than GFI1-36N. Overall, the intramolecular distance of GFI1-36S is increased by about 1.3 nm on average compared to GFI1-36N. This indicates larger spatial separations of the structural elements accompanied

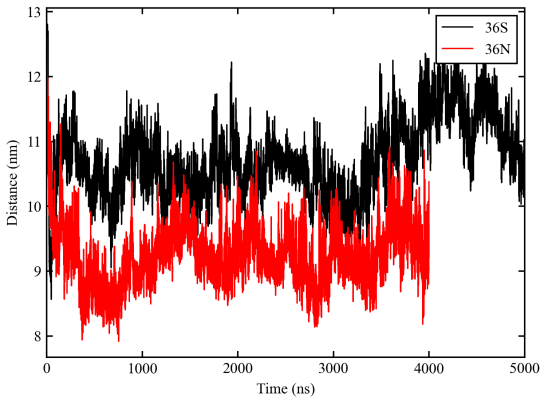

Figure 4: Temporal evolution of the maximum intramolecular distance between any two protein atoms for the GFI1 variants 36S (black) and 36N (red) calculated every 200 ps serving as a complementary measure of the spatial extent of the protein.

by increased conformational variability. Figure 5 shows

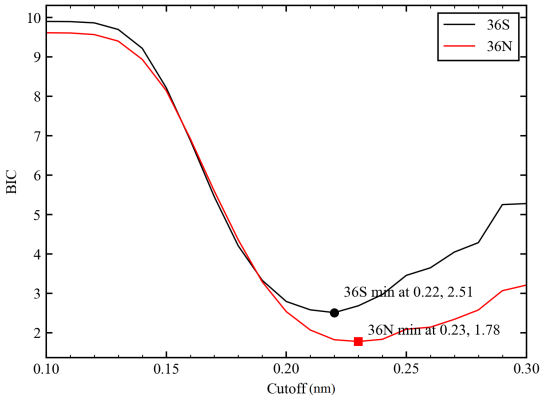

Figure 5: Normalized BIC as a function of the clustering cutoff R_c for the GFI1 variants 36S (black) and 36N (red). The BIC was used to determine the optimal clustering resolution by balancing model complexity and goodness of fit.

the Bayesian Information Criterion B (see also Table 1) as a function of R_c . For both GFI1 variants, B displays a U-shaped profile with a well-defined minimum, indicating the optimal cutoff. For GFI1-36S, the minimum is observed at a cutoff of 0.22 nm, while GFI1-36N shows a similar optimal cutoff at 0.23 nm. These cutoff values were used for the cluster analysis, the results of which are summarized in Table 1. Although the optimal cutoffs lie in a comparable range, the BIC curves differ between the variants: for GFI1-36S, the BIC increases more steeply at larger cutoffs than for GFI1-36N.

IV. Conclusion

In this study, the conformational behaviour of the wild-type GFI1-36S and the point mutation GFI1-36N was investigated using molecular dynamics simulations. Analysis of the simulated trajectories shows that neither variant forms a single stable conformation, but is best de-

10.18416/SCP.2026.2680

Table 1: Results of the cluster analysis for the GFI1 variants 36S and 36N based on the molecular dynamics trajectories. Shown are the total number of analyzed structures, the number of identified clusters, and the optimal clustering cutoff R_c determined using the Bayesian Information Criterion (see Fig. 5).

Mol.	# struc.	# clust.	cutoff (nm)
GFI1-36S	20001	764	0.22
GFI1-36N	15001	610	0.23

scribed by dynamic ensembles of different conformations. The absence of a dominant structure is a characteristic feature of intrinsically disordered proteins. Such an ensemble-based organization has been systematically described, particularly for regulatory proteins such as transcription factors [10].

Both variants exhibit pronounced structural flexibility, as shown by the time-dependent behaviour of global structural quantities such as solvent-accessible surface area and the radius of gyration. This analysis is complemented by the evaluation of intramolecular distances. Differences in spatial extent and in the frequency of highly expanded conformational states between the two variants become apparent. Such global observables are well suited for characterizing the conformational ensembles of intrinsically disordered proteins. Individual structures are of limited significance for this class of proteins [11].

Although the structural organization of the two variants is similar, differences in the temporal dynamics and distribution of structural states in the ensemble can be observed. For example, the GFI1-36N mutant shows more compact states with moderate fluctuations on average, while the wild-type GFI1-36S undergoes several transitions between more expanded and more compact conformations. The observed differences can be interpreted as different frequency distributions of the conformational states represented in the ensemble [11].

The cluster analysis performed in conjunction with the Bayesian Information Criterion provides a discrete description of the conformational landscape of the two variants. In order to adequately describe the observed conformations, it is necessary to divide the trajectories into several structurally distinguishable states. This observation confirms that an ensemble-based approach is necessary to capture the dynamics of both variants. Ensemble-based differences of this kind are typical for intrinsically disordered proteins and can be captured by statistical evaluation of extensive trajectories [12].

The point mutation at position 36 does not cause a fundamental structural reorganization of the protein. Instead, it influences the relative frequency of the conformational states within an ensemble and thus the global dynamics of the system. Such small systematic changes in ensemble composition represent a fundamental mechanism by which intrinsically disordered proteins can regulate their interaction properties [10].

Based on the current data, it is too early to draw significant conclusions about functional properties derived from the observed ensemble shifts. Future experimental and theoretical studies could help to clarify potential functional implications.

Acknowledgments

The work has been carried out at the Institute of Physics, Universität zu Lübeck.

Author's statement

Conflict of interest: Authors state no conflict of interest.

References

[1] T. Möröy. The zinc finger transcription factor Growth factor independence 1 (Gfi1). *The International Journal of Biochemistry & Cell Biology*, 37(3):541–546, 2005, doi:10.1016/j.biocel.2004.08.011.

[2] D. Frank, P. K. Patnana, J. Vorwerk, L. Mao, L. M. Gopal, *et al.* Germ line variant GFI1-36N affects DNA repair and sensitizes AML cells to DNA damage and repair therapy. *Blood*, 142(25):2175–2191, 2023, doi:10.1182/blood.2022015752.

[3] J. Abramson, J. Adler, J. Dunger, R. Evans, T. Green, *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 630(8016):493–500, 2024, doi:10.1038/s41586-024-07487-w.

[4] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1-2:19–25, 2015, doi:10.1016/j.softx.2015.06.001.

[5] T. Darden, D. York, and L. Pedersen. Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems. *The Journal of Chemical Physics*, 98(12):10089, 1993, doi:10.1063/1.464397.

[6] B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije. LINCS: A linear constraint solver for molecular simulations. *J. Comp. Chem.*, 18:1463–1472, 1997.

[7] G. Bussi, D. Donadio, and M. Parrinello. Canonical sampling through velocity rescaling. *J. Chem. Phys.*, 126:014101, 2007, doi:10.1063/1.2408420.

[8] M. Parrinello and A. Rahman. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics*, 52(12):7182–7190, 1981, doi:10.1063/1.328693.

[9] X. Daura, K. Gademann, B. Jaun, D. Seebach, W. F. Van Gunsteren, and A. E. Mark. Peptide Folding: When Simulation Meets Experiment. *Angewandte Chemie International Edition*, 38(1-2):236–240, 1999, doi:10.1002/(SICI)1521-3773(19990115)38:1/2<236::AID-ANIE236>3.0.CO;2-M.

[10] P. E. Wright and H. J. Dyson. Intrinsically disordered proteins in cellular signalling and regulation. *Nature Reviews Molecular Cell Biology*, 16(1):18–29, 2015, doi:10.1038/nrm3920.

[11] U. R. Shrestha, P. Juneja, Q. Zhang, V. Gurumoorthy, J. M. Borreguero, *et al.* Generation of the configurational ensemble of an intrinsically disordered protein from unbiased molecular dynamics simulation. *Proceedings of the National Academy of Sciences*, 116(41):20446–20452, 2019, doi:10.1073/pnas.1907251116.

[12] U. R. Shrestha, J. C. Smith, and L. Petridis. Full structural ensembles of intrinsically disordered proteins from unbiased molecular dynamics simulations. *Communications Biology*, 4(243), 2021, doi:10.1038/s42003-021-01759-1.